Understanding the Nature of Abrupt Decadal Shifts in a Changing Climate

James Henry Ricketts, BAppSci(RMIT), Grad Dip(Deakin), Grad Dip(RMIT), MAppSci(RMIT)

Victoria Institute of Strategic Economic Studies, Victoria University

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

June, 2019

Abstract

Planning for future climate risk tends to incorporate assumptions of smoothly accelerating climate change, around which unchanged variability constitutes the risk boundaries. This constitutes hypothesis *H1* – forced warming and natural variability evolve gradually and independently and the climate response is trend-like. Against this, there is evidence for *H2* – forced warming interacts with natural variability and the climate response includes abrupt steps. Earlier than expected breaching of risk bounds follows from H2.

New automation tools, and post-detection tests find and characterise step-like regime onsets in temperatures.

With these tools I show that step-like temperature regime shifts are detectable at all spatial scales at the land and ocean surface, and in the vertical temperature structure of the ocean. Based on published climate models shifts respond to warming by becoming more intense, wider-spread and more frequent. Regimes are regional, differ qualitatively between land and ocean, align with natural variability coincident with known bio-physical shifts. Two, circa 1976 and 1996, align with the Pacific Decadal Oscillation, involving rapid vertical ocean restructuring. One, 1968 in the Southern Hemisphere ocean does not, and 1986 in the Northern Hemisphere reflects atmospheric reorganisation.

H2 is strongly supported by the findings. Step-like warming dominates trends, increasingly so at finer scale.

Student Declaration

Doctor of Philosophy Declaration

I, James Henry Ricketts, declare that the PhD thesis entitled “Understanding the Nature of Abrupt Decadal Shifts in a Changing Climate” is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work”.

Signature [redacted] Date: 28 June 2019

Table of Contents

List of Figures

List of Tables

Dedication

Dedicated to Dr Penelope Helen (Penny) Whetton,

Climatologist, Artist and a Friend of the World.

5 January 1958 – 11 September 2019

A gentle and perceptive research leader, and an intellectual great, Penny advised and supported me in this work as she had done for many others. She once advised me to go feral. This thesis is the result.

Acknowledgements

Funding

J. H. Ricketts was the holder of a Victoria University postgraduate research scholarship.

Additional income resulted from ongoing casual support work for the Grazing Lands Systems Group within the Queensland Department of Science, and Information Technology, now known as Department of Environment and Science. Thanks are owed to Mr Ken Day, and Dr Ramona Dalla Poza, and Dr Jacqi Willcocks for the continued working relationship.

I was welcomed to, and accommodated within, the now Victoria Institute of Strategic Economic Studies in Melbourne.

Supervision

My principal Supervisor is Professor Roger N. Jones.

Three people filled the role of co-supervisor at different times: Dr Roger Bodman, Dr Masha Fridman, and Professor Peter Sheehan. My thanks to all for support when I flagged, also conversations both general and specific.

Support

My partner Lee has been very supportive. During this time, Mrs Lorna Palmer, and Mrs Jan Thompson left us all. I would also like to acknowledge support from Claire Ricketts, Emily, Andrew, Ruby, and Lily Compson, Allan Ricketts, Gretchen and Mark Polido, and Rachel Ricketts.

Thanks are also due to Dr Ian Smith, formerly of CSIRO, and Professor Geoffrey Head of the Baker Heart & Diabetes Institute, who recommended me.

The late Dr Penny Whetton reviewed several draft chapters. I wish she could have seen the result.

Celeste Young kept me on track with coffee and occasional talks over lunchs.

Many wonderful people work within QDES, CSIRO, various universities and the Bureau of Meteorology.

Examiners

Finally, thanks are due to the examiners who kindly dealt with numerous typos, made several helpful suggestions, and pointed me at some fruitful areas of ongoing research.

Data sources

Koninklijk Nederlands Meteorologisch Instituut (KNMI) make data available via the KNMI Climate Explorer and this was a valuable resource.

The Australian Bureau of Meteorology (BoM) made available to me their Global Climate Model Holdings on the National Computational Infrastructure (NCI).

Other data has been sourced from;

Met Office Hadley Centre,

National Aeronautics and Space Administration,

Goddard Institute for Space Studies,

National Oceanographic Data Center and United States National Climatic Data Center,

Berkeley Earth,

Cowtan and Way,

CMIP5 archives are made available by the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM).

Chapter 1: Introduction

> *"Human influence has been detected in warming of the atmosphere and the ocean, in changes in the global water cycle, in reductions in snow and ice, and in global mean sea-level rise; and it is extremely likely to have been the dominant cause of the observed warming since the mid-20th century. In recent decades, changes in climate have caused impacts on natural and human systems on all continents and across the oceans. Impacts are due to observed climate change, irrespective of its cause, indicating the sensitivity of natural and human systems to changing climate" (Pachauri and Meyer, 2014).*

Overview

This thesis addresses an area of some controversy – the hypothesis that under greenhouse gas-induced radiative forcing, climate changes in a step-like manner. The controversy arises because it is almost universally accepted that the forced component of climate change, especially global mean surface temperature (GMST), is trend-like. For example, the IPCC has stated that, "*On decadal to interdecadal timescales and under continually increasing effective radiative forcing (ERF), the forced component of the GMST trend responds to the ERF trend relatively rapidly and almost linearly (medium confidence)*." (Stocker et al., 2013a, Box TS.3, p62). Juxtaposed to this there is an emerging view that climate change proceeds non-linearly.

Jones and Ricketts (2017b) (JR2017), in addressing this controversy, proposes two main hypotheses concerning the relationship between external drives on climate, and internally produced natural variability of climate, as expressed at decadal time scales: (H1) forced warming and natural variability proceed gradually and independently, with the response to forced warming best represented as trend-like, and natural variability sometime obscuring the relationship; (H2) forced warming and natural variability interact.

The hypothesis that warming is step-like has been fringe for many years. Step-like warming has been widely accepted for palaeoclimate but not under greenhouse gas-induced radiative forcing influencing current and future climate. This thesis builds on the work of Jones who used the Maronna-Yohai bivariate test (Maronna and Yohai, 1978) (henceforth MYBT) to detect inhomogeneities in observed data while building a high-quality data set for southeastern Australia. He found the test was detecting climatically-driven shifts in climate data, principally rainfall (Jones, 1995, Jones et al., 2001). Further work tracing the impact of the millennium drought in south-eastern Australia began by investigating a shift in 1997 involving rainfall and temperature and whether it was a repeat of the previous shift in south-western Western Australia in the late 1960s and early 1970s. That was an instance where rain-bearing frontal systems bringing winter and spring rains to the region had shifted south, coincident with a shift to warmer conditions. A paper describing shifts in observed temperature within the region that also applied inverse linear methods to the detection of nonlinear anthropogenic warming was published based on this work (Jones, 2012) (J2012). Similar patterns of warming were found in regional simulations from climate models. This work was later expanded to detection of shifts at regional to global scales and the investigation of what that may mean economically (Jones et al., 2013).

During this time, the MYBT was settled on as the most reliable, sensitive and flexible test available for detecting step changes in climate data, being on a par with the Alexanderssen test, but more flexible with its use of reference data. Rodionov's STARS test, based on a modified t-test was also tested but found to be slightly less reliable with dates, and required some tuning, which the MYBT did not. The detection of multiple step changes in climate data was not automated, and involved a degree of researcher choice.

During the thesis we (Jones and I) published a paper severely testing the alternative hypotheses of step-like and trend-like warming (JR2017), supported by short paper outlining a multi-step analysis method developed for the project (Ricketts, 2015a). The latter was accompanied by a paper on a novel comparative method (Ricketts, 2015b), and later, preliminary work that has become Chapter 4 was presented for discussion (Ricketts and Jones, 2017). Finally, we have submitted a paper describing a self-regulating heat engine in the tropical Pacific that governs the process of step-like change in climate (Jones and Ricketts, 2019) (JR2019).

Terms and scope

In this theses, unless otherwise stated, trends and trend analysis refers to ordinary least squares (OLS) linear trend analysis. Steps are abrupt change in the mean of a time series. Shifts are described as measured abrupt change at a particular time, taking into account the possibility of other phenomena in time series.

Within the literature trends within time-series may be more complex, but are generally smoothly varying, or at least without discontinuities. They may be estimated by a variety of methods other than OLS.

In the literature "decadal variability" is generally treated as smoothly varying quasi oscillatory behaviour, albeit for many measures of variability the indices derived to describe them may be assigned discrete phases and times of phase change. For example the Pacific Decadal Oscillation (PDO) is a complex ocean system still being actively researched, but the PDO-index used to track it is derived as a time-series that tends to more or less smoothly alternate between runs of positive and negative numbers (positive and negative phase), (Overland et al., 2008, section 2).

When the IPCC uses the term "abrupt climate change" it is defined on geological time-scales thus, "Abrupt climate change is defined in this IPCC Fifth Assessment Report (AR5) as a large-scale change in the climate system that takes place over a few decades or less, persists (or is anticipated to persist) for at least a few decades and causes substantial disruptions in human and natural systems." (ibid. pp 70) .Such a definition is consistent with "tipping points" as discussed in the Literature Review. When they refer to "decadal variability" it is described purely in terms of changes of trend thus, "Despite the robust multi-decadal warming, there exists substantial interannual to decadal variability in the rate of warming, with several periods exhibiting weaker trends (including the warming hiatus since 1998)" (ibid. pp 36).

Greenhouse gas (GHG) induced radiative forcing is often referred to as anthropogenic global warming (AGW), and a component of a more general phenomenon of human induced climate change.

Major issues

The problem of abrupt or non-linear climate change matters for a number of reasons. In JR2017, H1 and H2 represent very different relationships between forced warming and natural variability, including decadal variability. Under H1, for the purposes of understanding and projecting climate change, natural variability is treated as independent of forced warming, and for statistical parameterization, it is treated as noise or as a nuisance parameter. Under H2, natural variability is not separate from the rest of climate and interacts with forced warming; abrupt changes are expected in natural variability, they are not noise to be ignored or adjusted for.

1. If it is not noise then abrupt change is a signal of regime changes in natural variability, and hence detection of climate shifts is important to other science as well.
2. It matters because the predominant driver of climate change, GHG, shows trending behaviour and so non-linear response to this implies heat buffering with periodic release which would be expected to lead to clustering of effects.
3. Thus it matters to risk managers and planners because it changes the risk boundaries near term (Jones et al., 2013).

Because GHG emissions are a principal factor in forced warming, and because they are rising nearly steadily and GMST is also rising nearly steadily, the predominant mode of analysis, especially in climate change risk analysis is projection of warming trends onto various risk factors. Methods such as pattern scaling (Mitchell et al., 1999, Mitchell, 2003), for example project future global GMST onto local historical records with strictly linear relationships to estimate future local conditions (Jones and Page, 2001, Ricketts and Page, 2007, Ricketts, 2009, Ricketts et al., 2013).

Carter (2006) published an op-ed which included the spurious claim that warming had ceased after 1998 (hence the term "warming hiatus" above). The Garnaut Report (Garnaut, 2008, Box 401, p 79) for example specifically investigated whether a statistical break in the GMST trend occurred in the late 1990s, concluding no. This focus on trend at the expense of other metrics then discourages investigation of changes that occurred at the same time, for example sudden Siberian cooling with unexpected expansion of evergreens (He et al., 2017), or North American drought (Delworth et al., 2015).

Despite a number of recent papers supporting the position of "no hiatus" (Cahill et al., 2015, Lewandowsky et al., 2015, Rajaratnam et al., 2015, Risbey et al., 2018), papers taking a contrary position that the climate changed around that time, abound (Sillmann et al., 2014, Trenberth et al., 2014, Lee et al., 2015, Zang et al., 2018). Much of the conflict revolves around the differing role statistics is assigned in these investigations.

Analysis of discontinuities in climate records to delineate both climate and bio-physical regimes has a long history, often citing changes in many systems at once. For instance multiple changes associated with the PDO (Minobe, 1997, Newman et al., 2016), and very wide-spread changes in the later 1980s (Reid et al., 2015). Recently papers have been published which extract a stair-case like pattern of change (Bartsev et al., 2017, Belolipetsky, 2014, Belolipetsky et al., 2015, Saltykov et al., 2017). The levels of agreement on timing of change, across disciplines, once non-linear change is explored, means that abrupt shifts simply cannot be dismissed as noise.

Research question

Due to the complex nature of the climate system, statistical inference is the major form of reasoning used when analysing and interpreting climate data. This has led to theoretical conclusions about the Earth system being made on the basis of statistical tests of time-series passing a given *p*-value, and this forms the basis of the climate in terms of trends (North et al., 1995, Santer et al., 2011, Hasselmann, 2002, Rahmstorf et al., 2017). It was major motivation for the application of severe testing (Mayo and Spanos, 2006, Mayo and Spanos, 2011) to the climate analysis in JR2017. The philosophical basis behind severe testing is that tests providing statistical confirmation cannot be used as evidence to support a theory unless they also discount their rivals through methods that provide a high level of falsification. A great deal of climate science is confirmatory only, in that a hypothesis is confirmed as plausible within a specific experimental and/or conceptual framework. (RN Jones, Personal Communication).

JR2017 outlines a synthesis of these ideas into a theoretical-mechanistic/statistical-inductive framework, in which amongst other things, confirmation on the basis of *p*-values first requires joint probative criteria in support and opposition to be adequately addressed.

Additionally, reasoning with statistical tests requires that the tests and data be adequately matched (Mayo and Spanos, 2004), and that statistical model families be selected with care (Spanos, 2010).

Although JR2017 provided a general conceptual framework, the questions about the nature of step-like changes, trigger mechanisms and sustaining processes, spatio-temporal extents, and regional distribution remained.

Aims and objectives

A major goal for this project was to construct an objective rule-based multi-step system to remove the potential for human bias from the results – now known as the “multi-step bivariate test” (MSBV). The resulting system is robust, flexible and fast enough to allow analysis of spatial global climate model data. The algorithm is progressive, and incorporates repeated testing and refinement of previous results. Breakpoint methods are not widely considered to provide reliable evidence for nonlinear change, and have been the subject of criticism. Extensive sensitivity testing of the method was performed to ensure that the test was suited to task and used within limits.

A second major goal was to ensure that the MSBV change-point results had high probative value. Climate data is intrinsically complex, incorporating the effects of multiple feedbacks, as well as forced warming. Most statistical tests used in climatology, including tests of trends and the MYBT, are framed under simplified assumptions. A full chapter is devoted to post-detection testing of the change-points and the data in which they are detected, inspired by misspecification testing (Mayo and Spanos, 2004). This ensures that the data are not random-walks, the change-points are not false positives, and indications of false negatives are screened for.

Another was to measure the contributions of trends and shifts to overall warming. If warming interacts with natural variability and that in turn is indicated by shifts then the relationship between warming and patterns of shifts can be explored. Are regimes global phenomena that act everywhere, or large scale phenomena that act over large regions – perhaps continents or ocean basins, or are they still more localised? If they are more localised, do the locations, durations, and intensities of these events change under forced warming or are they independent of it? These questions are at the heart of considerable contention. Much of Jones's earlier work had concentrated on evidence for local to regional changes in climate. Jones et al. (2013) applied his methods to global data and found similar step-like changes. The PDO, now recognised by characteristic changes in the Northern Pacific Sea Surface Temperatures, was first detected as fluctuations in salmon production and found to correlate with abrupt changes in Northern climate (Mantua et al., 1997). Other researchers found similar abrupt changes in climate parameters, and at a variety of scales, but even so, little is available that applies the same methodology from global to fine scale to enable global scale phenomena to be mapped to local phenomena.

Two chapters build on the prior chapters and augment analyses of data reported in JR2107. They present analyses, mostly of surface temperatures, from global, down to what is referred to as grid scale (similar to the scales at which global climate models work). They concentrate mostly on the so-called historical or industrial period (after 1850 or 1880 depending on the data source). Global climate models are used as the only source of information prior to industrialisation and into the future.

To address the question of interaction being expressed as a relationship between warming and patterns of shifts under as wide a variety of warming conditions as possible climate model surface temperature data was used in conjunction with observations. The thesis first concentrates on coarse scale averages of temperature with analysis of global averages and averages of the six major bands or zones of the Earth. The frequency of shifts in each zone was collated with the warming conditions from a hypothetical pre-industrial Earth (models only), through four stages of increased industrialisation (observations and models), to the end of the 21st Century under mild and extreme warming (models only). It then concludes with a breakdown of the observed zones into 45 degree sectors, to compare the influence of data averaging on the measured contributions of step and trend.

JR2017 had also found steps in tropospheric temperatures that mirrored those of the surface. The question naturally arises as to whether the vertical structure of the ocean changes concurrently. Since the previous results imply that shifts occur regionally, but the warming appears to either expand coherent regimes or create new ones, the thesis thus looks at observed data at finer grid scale, and extends the analysis to ocean temperatures at two depths of interest. A single GCM with multiple realisations is chosen for comparison with observations, and to assess differing model evolutions given identical model and forcing.

Because the same data are averaged at a variety of spatial scales it was clear that at finer scales steps predominated more over trends. This result can only be obtained if steps in fact exist and are not artefacts of detection, and the systems producing steps are regional rather than global.

All results are consistent with land changes following ocean based ones.

Research approach

Thus the research is structured so as to further extend J2012, JR2017, and related work.

1. Adapt the step-detection methods already established by Jones for the detection of single step changes in climate data, to multiple step changes. Following this, to add empirical sensitivity testing to allow for improved confidence limits, especially those parameters which do not seem to have been stressed in previous publications.
2. Accepting that all statistical tests rest on sets of ruling assumptions about the data, so then to assemble a battery of statistical tests for the purpose of ensuring that the assumptions of the MSBV are accounted for and elements of the data not consistent with these assumptions are interpreted correctly.
3. Apply the new statistical tests to key datasets reported upon in JR2017, so as to extend some of its key findings, and to thus ground the later research. Then, using

observations, and climate models under widely varying conditions, to examine the interaction of forced warming and regime changes represented as step-like shifts.

4. Key events and features (well documented transitions in major modes of natural variability) are sought at finer scale in observations, and particular spatial signatures are shown to correspond to these. These are then examined in a sample climate model. Patterns in the ocean vertical heat distributions are compared to those in the surface temperatures.
5. Further evidence of the importance of shifts is also sought by comparing the contribution of shifts to total warming, deriving these contributions from analyses performed at varying spatial scales.

Thesis structure

Chapter 2 contains a literature review focussed on the following chapters, and a section which draws on the literature to summarise some more philosophical statistical theory and to draw together a synthesis. In particular the foundations of the theoretic-statistical/mechanistic-inductive framework, are explained; as are the applicability of severe testing and misspecification (M-S) testing.

Chapter 3 describes the MSBV, its algorithmic structure, sensitivity testing, and its validation against synthetic data. It demonstrates three different applications drawn from later work in the thesis.

Chapter 4 presents a suite of tools, selected by consideration of M-S testing, and applied with additional supplemental information. As mentioned climate data is more complex than many statistical tests allow for. Thus a validation suite designed to assess the suitability of the data for analysis was produced. These supplement probative tests reported in JR2017. The individual tests are carefully combined with consideration of the assumptions of each test, to produce a classification scheme for data segments in which are detected single change-points. The latter proves useful in the following chapter.

Chapter 5 extends the surface temperature analysis of JR2017, and applies the validation tools. Climate model analysis is extended from global to zonal data, and from combined land/ocean averages to separate land and ocean data. Observations are also extended to 30^o by 45^o sectors inside each 30^o by 360^o zone. In these regimes appear to occur at roughly half ocean basin and continental scale. A wide-spread abrupt shift in global climate around the latter half of the 2011-2020 decade would be consistent with the model consensus. Changes in the zonality, frequency and intensity of shifts can be attributed to warming.

Chapter 6 examines gridded observed surface temperatures from the point of view of spatio-temporal patterns of step-like changes. Events identified by regional attribution in the Southern Hemisphere, circa 1968, and three major events documented in the literature circa 1976, 1986 and 1997 all correspond to regional patterns of step-like changes. The 1976 and 1997 events correspond spatially and temporally to changes in ocean temperatures at 100m and 700m, and align with the PDO. An event in circa 1924, in the East of the North Atlantic corresponds a feature found circa 1994 and both align with the AMO. In all cases the findings are consistent with the literature. One example GCM with multiple realisations and some similarity to observations during the 20th Century is chosen for examination but whilst similar features to various observed patterns are found they appear to have coordination to each other or to observations.

Chapter 7 is a summary discussion.

Chapter 2: Survey of literature

Introduction

As stated in the introductory chapter, two main hypotheses exist concerning the relationship between external drives on climate, and internally produced natural variability of, climate as expressed at decadal time scales. Jones and Ricketts (2017b) (JR2017) focusses on anthropogenic warming due to greenhouse gases, primarily CO2 mediated by other emissions as the external driver of interest. JR2017 proposes: (H1) forced warming and natural variability proceed gradually and independently, with the response to forced warming best represented as trend-like, and natural variability sometime obscuring the relationship; (H2) forced warming and natural variability interact. When analysed under the assumptions of H1, variation in trends may also be analysed using non-discontinuous segmented statistical models (Cahill et al., 2015).

JR2017 identifies two possibilities for H2. Patterns of response may project onto modes of climate variability as proposed by Corti et al. (1999), who found that the (then) recent Northern Hemispheric warming could be interpreted as projections of forced warming onto natural patterns of atmospheric variability, stipulating the findings were not evidence against anthropogenic forcing, thus hypothesising a principally one way interaction. Alternatively the relationship may be two-way as reported by Branstator and Selten (2009) who finds a small degree of feedback from these patterns to their response, which thus dynamically modifies those structures.

This review is based on the accepted science of climate change as the science pertains to the last century and half, through to the end of the current century. In particular this thesis accepts that anthropogenic causes, particularly greenhouse gas emissions, are responsible for most of the observed changes in mean global surface temperatures (GMST). The Intergovernmental Panel on Climate Change (IPCC), Summary for Policy Makers, Synthesis Report (Pachauri and Meyer, 2014, SPM 1.1) commences with the following words *"Warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia. The atmosphere and ocean have warmed, the amounts of snow and ice have diminished, and sea level has risen*". It continues almost immediately, "*Anthropogenic greenhouse gas emissions have increased since the pre-industrial era, driven largely by economic and population growth, and are now higher than ever. This has led to atmospheric concentrations of carbon dioxide, methane and nitrous oxide that are* *unprecedented in at least the last 800,000 years. Their effects, together with those of other anthropogenic drivers, have been detected throughout the climate system and are* ***extremely likely*** *to have been the dominant cause of the observed warming since the mid-20th century*." (ibid., SPM 1.2, Emphasis in original).

However whilst the primary cause of current and projected climate change is not disputed here, open questions about *how* climate changes remain.

This thesis lays a much greater emphasis than most works on aspects of the statistical methods and reasoning involved. This comes about because the issue of abrupt changes in surface temperature seems difficult for some to accept, despite being at the heart of the identification of significant decadal variability modes.

Abrupt decadal climate regime changes are then examined in more detail.

I include a very brief summary of the foundations of the science greenhouse warming which indicates the early physical sciences roots of interest in climate, culminating in the formation of the Intergovernmental Panel on Climate Change (IPCC). See also Houghton (2009, p.17). This delineates the area of generally accepted science from the more controversial details of how climate changes.

Milestones on the development of the science.

The science of the greenhouse effect is built on foundations first laid in 1824 by Jean Baptiste Joseph Fourier who discussed the role of the atmosphere as a moderating influence on the temperature of the Earth (Fleming, 1999). Tyndall (1861) suggested a mechanism for this moderation; water vapour and CO_2. Arrhenius (1896) first calculated the sensitivity of the global air temperature to a doubling of CO_2 with water vapour feedback, obtaining a value of 3-4°C. Building on the further work of Callendar (1938), Plass's estimate of 3.6 °C obtained 60 years later remained similar (Plass, 1956). Observing the relationship between CO_2 levels and ices ages, he concluded there was "*no possible stable state for the climate.*"

Following publication of results of climate models run by Goddard Institute for Space Studies (GISS) (Hansen et al., 1988), and at the urging of the World Meteorological Organisation (WMO), the IPCC was founded in 1988 (see https://www.ipcc.ch/about/history/) and has since released five reports summarising the state of the knowledge of climate science as it was in 1990 (Houghton, 1990), 1995 (Houghton, 1996), 2001 (Griggs and Noguer, 2002), 2007 (Solomon, 2007), and currently 2013/14 (Stocker et al., 2013c). The physical science basis of global climate change is well documented in these.

Initially the science incorporated heat-balance or radiative-convective conceptual models where temperatures were measured above the ground (e.g. Arrhenius), and this convention applied as models evolved from simple, through highly parameterised multi-layer models (Manabe and Strickler, 1964) to today's highly complex systems which are "are almost as complex as nature itself, making it hard to picture the connections between the most essential processes" (Benestad, 2016). The radiative-convective model remains core theory. In JR2017 we say "*Radiative transfer theory constitutes core greenhouse theory. However, the subsequent process of heat diffusion through the climate system is less well understood, although the understanding that if greenhouse gases are increased, the atmosphere will warm until the radiative balance at the top of the atmosphere is achieved also constitutes core theory. Our conclusion that the atmosphere does not warm in situ will challenge many who consider that to be a basic part of the greenhouse effect.*"

Following JR2017, this thesis agrees with the conventional model of most heat being trapped near the surface by greenhouse gases and reflected downwards, but departs from the conventional explanation of gradual warming in how that heat is dissipated. It also agrees with the conventional model with respect to long-term global mean warming being linearly proportional with the amount of forcing as a first approximation. The contentious issues within the science are therefore largely confined to how observations and model output should be interpreted, the detection and attribution of change on decadal timescales and the mechanisms underlying such changes.

Whilst the conclusion that human activity is responsible for ongoing climate change, and that this is due to greenhouse gas emissions is accepted, the degree of risk this poses is socially and politically contentious.

Abrupt decadal scale climate variability and change

This thesis considers climate change at decadal time scale, as it is expressed at global to regional scales.

Plass's conclusion of no stable state for climate implies little about *how* the climate changes, and probably applies better to the current interglacial period. Mills et al. (2017) suggest that CO_2 outgassing had a role in stabilising the early Earth climate, and recent work shows that life itself can stabilise climate (Dyke and Weaver, 2013). Short term stability is a theme in this thesis.

Defining abrupt decadal change

Decadal scale variability generally refers to climate variability of the order of several decades, but often informal definitions are used. The IPCC refers to "Decadal prediction" as "a new endeavour in climate science" (Kirtman et al., 2013) with a context that suggests that decadal averages are the unit of measure, and implying that decadal variability is variability over small numbers of decades. Drijfhout (2018) discusses the relationship between GMST, and ocean heat uptake and related measures, considering the cause of apparent hiatuses (persistent reductions of trend in GMST) in global climate models. The concept of a well delineated and persistent climate state that switches abruptly, and potentially reversibly, is central.

"Abrupt climate change" is defined by the IPCC WG1 as "A large-scale change in the climate system that takes place over a few decades or less, persists (or is anticipated to persist) for at least a few decades and causes substantial disruptions in human and natural systems." (Stocker et al., 2013b, Glossary). Alley et al. (2003) seem to use the term in reference to various threshold crossing and tipping points, but also uses the term in reference to the quasi-oscillatory ocean systems and ecological changes. Repeatability distinguishes abrupt regime change of the sort discussed in this thesis from singular tipping point behaviour involving threshold crossing.

Contrasting Regime shifts and Tipping Points

Lenton et al. (2008) defined "tipping elements" in the climate as, "large-scale components of the Earth system that may pass a tipping point". The tipping point (see their Table 1) is a threshold and the discussion pertains to elements with at least sub-continental scale that can switch rapidly given minor perturbations (implicitly, across the tipping point). They then discuss policy relevance.

Drijfhout et al. (2015) extend this work with an exhaustive catalogue of tipping elements giving four categories of abrupt changes (see their Table 1). Their categories are (i) unforced switching (e.g. Southern Ocean sea-ice bimodality), (ii) climate change related forced transition to switch, (iii) rapid change to new state (e.g. winter sea-ice collapse, abrupt decrease or increase in sea-ice, local convective collapse), (iv) gradual change to new state, (e.g. boreal forest expansion or forest diebacks). Their first category is not really commensurate with climate variability, it refers quite specifically to sea-ice regimes of long residence times and feedbacks on water column stability. The second refers to similar phenomena which appear only after the climate has passed certain thresholds. JR2017 contrasts decadal abrupt changes with tipping elements as follows, "Note that these step changes are quite different to those catalogued by Drijfhout et al. (2015), who used a different method to screen the CMIP5 model ensemble for abrupt shifts that could be considered as singularities, locating 37 ocean, sea ice, snow cover, permafrost and terrestrial biosphere changes".

Sgubin et al. (2017) consider one contribution to the AMOC collapse tipping point, the North Atlantic Subpolar Gyre. During their analysis of modelled SST trends they showed that the warming signal is spatially non-uniform.

Regime shifts

By contrast, in this thesis, climate regime shifts do not involve essentially irreversible changes or extensive reconfigurations.

The term "regime" itself is not consistently defined in the literature. The first usage of the word regime in a climate context is suggested by Overland et al. (2008) quoting Isaacs (1976) to be ''the assumption is that there are some normal statistics to all kinds of (ocean) conditions. Rather, there are probably a great number of possible regimes and abrupt discontinuities connecting them."

Minobe (1997) defined a "climatic regime shift" as "a transition from one climatic state to another within a period substantially shorter than the lengths of the individual epochs of each climate states (sic)". This use is more consistent with analyses of Hare and Mantua (2000), Vives and Jones (2005), Hope et al. (2006), Swanson and Tsonis (2009) and Jones (2012), where changes have been identified as taking place over one or two years. Several papers have proposed methods for measuring the start to finish time of regime transitions, yielding estimates of one to two years with (Yan et al., 2015, Yan et al., 2016).

Overland et al. (2008) summarise various usages of regimes as: (1) displacement or shifts in time-series, (2) underlying mechanisms, and (3) the distinction between external forcing and internal reorganization of ecosystems. By adhering to their first definition they find regimes changes in the Pacific Decadal Oscillation index (PDO) in 1976, 1989 and 1998. But they cannot eliminate red-noise as being the source of regimes, suggesting that current (at 2008) understanding is unable to attribute a deterministic origin or a single definition to a climate regime. Before this paper Rodionov (2004) had concluded that the PDO is not simply red-noise, given the amount of autocorrelation required to produce the appearance of the 1976 change. Mantua et al. (1997) implicitly uses the third definition of a regime, reorganised ecosystems, in reporting regime changes the PDO correlating with changes in fish-stocks in 1925, 1947 and 1977, remarking that Minobe (1997) found regime changes in a variety of measures in 1925, 1945 and 1977.

The attribution in the literature of regimes to red-noise is a motivation for Chapter 4 of this work.

The initial approach to regimes in this thesis is in line with the displacement definition. I assess shifts in temperature at the surface and in the oceans, in order to inform an improved understanding of underlying mechanisms.

Being able to distinguish between external forcing and internal reorganization of climate are secondary and tertiary goals.

The results can also potentially contribute to an improved definition of climate regimes.

For the duration of this thesis, a regime, unless otherwise indicated, is a period of near-stationary behaviour in a variable (or set of variables) delineated by a brief period (within a year or two) of non-stationarity which are detected as shifts. An abrupt decadal climate change is a regime change of climate variables, where the near-stationary period endures for a small number of decades.

The detection method used in this thesis, the MYBT, treats shifts as being step-like within a single sample interval (for annual data, a single year). The following sections describe (1) the current understanding of climate regimes including their detection and attribution and (2) statistical methods for detection that focus on the bivariate test, while taking account of both false positives and negatives.

Temperature changes and recovery associated with volcanic aerosols

There is a potential for the well observed temperature reductions and subsequent recoveries following major volcanic events to be identified as regime shifts. The four events of interest are Krakatoa 1883, Mt Agung 1963, El Chichón 1982, and Pinatubo 1991. Soden et al. (2002) discuss water vapour feedback post Pinatubo in producing the observed cooling. Schmidt et al. (2014) cautions that overestimated optical depth of aerosols from Pinatubo, ENSOs out of phase with observations, and incorrectly estimated solar forcing, all combine to bias model estimates of the effect of Pinatubo by about one third.

JR2017 considers volcanoes as part of attribution studies to equilibrium climate sensitivity (ECS) based on a 107 member ensemble of CMIP5 model runs. The Krakatoa eruption is associated with a downward step of global temperatures in the 1867-85 decade in a significant number of runs, and the Agung eruption with a downward step in 1956-65. El Chichón is not mentioned and Mt Pinatubo is considered as a potential cause of low ECS correlation due to rebound and as having contributed to the timing of the subsequent shift. El Chichón is considered by Reid et al. (2015), with rebound hypothesised to have combined with rapid warming so as to contribute to a regime shift in the mid-1980s. They estimate a cooling of 0.2-0.3°C which recovered by the mid to late 1980s and thus the net warming for the period of recovery was proposed as a trigger for a regime shift. They then consider why the larger and later Pinatubo eruption did not trigger a regime shift.

Timmreck (2011) uses GCMs to analyse the effects of super volcanoes (>1000 times the emissions of Mt Pinatubo with a frequency averaging 1.4 events per million years), finding that peak deficits in temperature of up to 6°C in the global record can occur one year later (-12°C in the NH), but are compensated by warm air advection. In an earlier study Jones et al. (2005), by scaling the Pinatubo aerosols to super-volcanic scale showed significant but short lived climatic impacts including reductions of GMST for up to a decade and doubling of the AMOC flow. In these and other cases summer cooling and winter warming are observed.

The imprints of Mt Agung, El Chichon and Pinatubo are visible in selected regional records but where present have the typical morphology of a transient response with exponential recovery (Tsay, 1988, see the discussion of transitent change (TC) outliers) reverting to the previous climate within five years at most. The MYBT is not expected to detect these transient events as it is intended to detect abrupt shifts and maintained changes, however the Agung transient may have been detected in zonal analysis (Chapter 5).

Understanding decadal climate variability

JR2017 identifies the interaction of forced warming and decadal climate variability as a key discriminant between H1 and H2.

Abrupt regime shifts may be part of "natural variability" as is implied by their detection in paleo-records and as part of long term climate processes (Minobe, 1997, Kirby et al., 2010, Yasunaka and Hanawa, 2002), or they may indicate moderation of natural variability by exogenous processes (Trenberth, 1990, Vincent, 1998, McFarlane et al., 2000). The latter would imply that analysis of regime shifts may be used to assess sensitivity of parts of the Earth system to climate change. For a summary of these and related issues see Stott et al. (2010), as well as the reports of the IPCC Working Group II (IPCC, 2014a, IPCC, 2014b).

It is also the case that data, previously analysed as smoothly varying, may reveal itself to have abrupt shifts once they are looked for. For instance north Pacific sea surface temperatures (SSTs), assumed to have followed a smooth cooling trajectory from the 1940s to the mid-

1970s, were shown to include abrupt downward shifts round 1970 when analysed for these (Thompson et al., 2010), although in JR2017 and in this thesis the dates are circa 1976.

A number of global climate components have been identified which exist at regional scale (continental or ocean basin) but affect the global climate. Many of these are quasi-oscillatory and occur with varying periods from sub-decadal, through decadal and centennial scales, to multi-millennial. Hurrell et al. (2010) list a number of influential systems.

- The El Niño-Southern Oscillation (ENSO) is an important player operating irregularly over two to seven years (Henley, 2017).
- The PDO, which is defined in the North Pacific interacts with ENSO, although the nature of the relationship is still an open question. A similar phenomenon occurs across the entire Pacific Basin, known as the Inter-decadal Pacific Oscillation (IPO) and may constitute a single complex system (ibid.). Shen et al. (2006) describe the system as having irregular cycles of 75-115 years prior to 1850 and 50-70 after.
- The Atlantic Multi-decadal Oscillation (AMO) may be related to a global heat carrying system, the Atlantic Meridional Overturning Circulation (AMOC) although climate model evidence is not strong (Trenary and DelSole, 2016, Moore et al., 2017). It is described a variation in North Atlantic sea surface temperatures, having 60-90 year variability (Knudsen et al., 2011).
- A roughly decadal quasi-periodic redistribution of air mass between the Arctic and the North Atlantic is known as the North Atlantic Oscillation (NAO). It was described by C.T. Walter and G.W.Bliss in 1932 (Hurrell, 1995) although mechanisms are not well understood see also Hurrell and Deser (2010). Deser and Blackmon (1993), describe the dipole nature of the NAO, and report fluctuations of approximately 9 years prior to 1945, and 12 years after.
- Another system known as the Arctic Oscillation is highly correlated with NAO, and is also primarily a phenomenon of surface winds (Ambaum et al., 2001).

Of these the PDO/IPO and AMO are most prominently aligned with shifts in this thesis. They also figure prominently in the literature on both climate regimes and related eco-system regimes,

Detection of decadal climate variability, approaches and methods

The presence of climate variability at decadal time scales is not itself controversial. Multiple approaches and methods have been used in the research. Research often approaches the identification by combinations of detection of shifts in time-series, and associated eco-system changes in line with Overland's first and third regime types, but also makes use of spatial patterns of variation to focus on regions of interest.

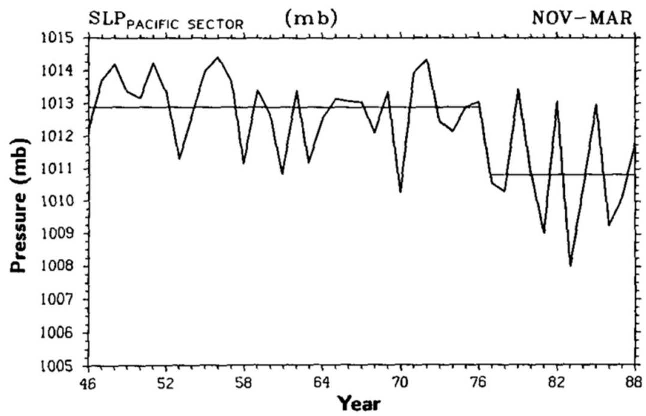


Figure Ch2.1: Time-series of mean North Pacific sea-level pressures averaged over 27.5°N to 75°N , 147.5°E to 122.5°W for the months November-March. Means for 1946-76 and 1977-87 are indicated. Source: Trenberth (1990).

The identification of the PDO relied in part on identification of correlated and quite rapid changes of climatic variables and biological data (e.g. sardine catch rates, as well as salmon and hake especially in 1976/7 (McFarlane et al., 2000)). Trenberth (1990) shows an abrupt downward shift in North Pacific sea-level pressures after 1976, without specifying the break-point identification method used (see Figure Ch2.1). He draws attention to teleconnections between local sea-level changes and global phenomena. Variation in the linkage of these decadal shifts to variation in intensity and frequency of ENSO was hypothesised by Trenberth and Hurrell (1994). Mantua et al. (1997) regressed the PDO index against spatial records of sea-level pressure (SLP) and sea-surface temperatures (SST) to obtain diagnostic patterns, while Minobe (1997) applied empirical orthogonal function (EOF) analysis to the dendrochronology of North America and spectral analysis of sea-surface temperatures to infer the association of shifts dated around 1890 and in the 1920s with "the 50-70 year variability" over those areas, and previously known shifts in the 1940s and 1970s (Figure Ch2.2). As stated above, both derived almost identical change-point dates. Strongly negative PDO values were noted in 1989-1991 (Hare and Mantua, 2000, Mantua, 2004). Changes associated with 1998 were postulated as early as 2000 from principal component analysis (PCA) of the Aleutian Low, and changes in fish-stocks (McFarlane et al., 2000).

Hare and Mantua (2000) also found that the 1989 shift was not as pervasive as 1976, nor a return to 1976 conditions, and suggested that the North Pacific and Bering Sea ecosystems may allow for earlier detection of climate regime changes than climate data alone. In studies of the North Atlantic bio-physical systems, (Reid et al., 2015) detected changes in climate and eco-systems around 1986, using a combination of principal components analysis, a tool called Change-point Analyzer (Taylor, 1997) which appears to perform CUSUM analysis, and a development STARS for multiple steps.

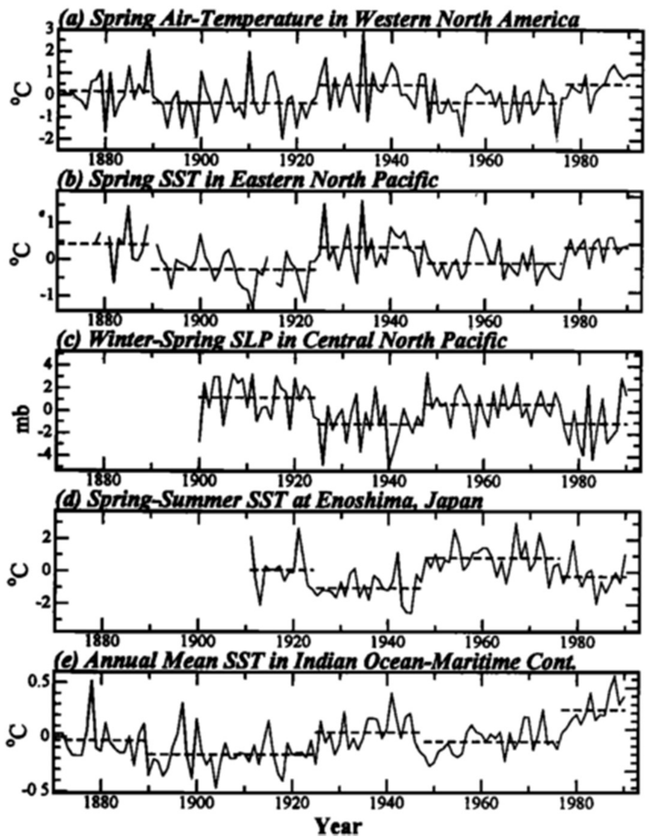


Figure Ch2.2: Coherent step changes in sea surface temperatures and sea level pressure. Source: Minobe (1997).

The AMO was described by (Folland et al., 1984) who used a detrended maximum power spectral analysis to find a dominant peak period of 83 years in GMST and night near surface marine temperatures. Schlesinger and Ramankutty (1994), also discovered an oscillation in GMST records and corresponding detrended North Atlantic ocean temperature records using singular spectrum analysis (SSA). An 8,000 year proxy study relates much of the AMO variation in that time to insolation (Knudsen et al., 2011). Moore et al. (2017) based on proxies (e.g. Labrador Sea algal data), suggests that the onset of industrial warming in 1850 amplified the AMO. (Mann et al., 2014) reviews the methods and finds that, possibly due to data length limitations, the detrended residuals method tend to produce biased estimates of phase, finding that the AMO "*appears likely to have been in a cooling phase in recent decades, offsetting some of the anthropogenic warming*".

Regional Changes

Nicholls et al. (2004) report links between rapid temperature changes and rainfall changes in New South Wales. They separated temperature changes into rainfall associated, and non-associated changes, and concluded that changes in the IPO might be responsible for changes in the relationships. They suggested that forced climate change may play a part.

Jones has shown localised regime shifts over Australia in a number of studies. Jones first selected the MYBT in his PhD thesis for examination of hydrological data in studies of S.E. Australian closed lakes. Jones et al. (1999) warns that non-linear rapid climate change in the area of Jabiluka is possible. Jones et al. (2001) reports on evidence from three lakes of an apparent abrupt regime shift round 1840 and before 1863, which preceded large scale changes in land use. Vives and Jones (2005) (VJ2005) documents abrupt changes in Australian rainfall between 1890 and 1989, along with the detection methods used – LePage test (Yonetani (1993),and MYBT – and shows Australia wide patterns of rainfall shifts corresponding to roughly 1895, (possibly 1922/3), 1945, and 1967-73. Jones (2010) shows an abrupt shift in the relationship between maximum temperatures and rainfall in Victoria post 1995 using the MYBT. This paper must rank amongst the earliest to quantify the contribution of sudden shifts in apparent warming rates (for example, concerning the Melbourne region pp 151, "*The rate of change in this analysis is closer to 1.2°C per century but the bivariate analysis suggests that about two-thirds of that trend can be attributed to the rapid shift in T max of 0.9°C* …").

Jones (2012) published a comprehensive analysis of regime shifts over SE Australia, again using the MYBT and the STARS test. The narrative had become more general, and the extension from regional regimes to global phenomena was further expanded by (Jones et al., 2013).

Jones carefully selected the MYBT for use in the detection of hydrological regime changes, homogenisation of data sets and after VJ2005, detection of temperature regime shifts. There are a number of alternative approaches.

Step detection

Observed regional regimes, approaches and detection methods

Regional shifts in climate and ecological regimes have been associated with the NAO post 1987, in the North, and Behring Seas (Beaugrand, 2004, Alheit et al., 2005). Beaugrand (2004) reports a complex web of changes involving changes in the North Sea, relating to a phase shift of the NAO post 1987, and a subsequent warming. Deser and Blackmon (1993), also used EOF to identify the dominant modes of variability in the North Atlantic SST and air temperatures and explore the dipole nature of the NAO. No comment was made as to whether the change from 9 year variability to 12 years after 1945 might relate to the contemporaneous PDO shift. Hurrell and Van Loon (1997) also show by a frequency analysis changes in the power spectrum and reddening of the NAO signal in recent decades and an association between the North Pacific index (NP) and the Southern Oscillation (SO).

Franks (1999) used a Mann Whitney U test (Mann and Whitney, 1947) to assess the assumptions applied to empirical flood frequency analysis using 40 stream gauges in NSW, concluding that there is evidence of an abrupt change in flood frequency distributions post 1945.

The AO, has been found as a second mode of variation in Northern Hemisphere Sea Surface Temperature (SST) data. For instance Yasunaka and Hanawa (2002), used EOF analysis of SST and state that "*six regime shifts are detected in the study period from the 1910s to the 1990s: 1925/26, 1945/46, 1957/58, 1970/71, 1976/77 and 1988/89*".

Hope et al. (2006) trialled self-organising maps (SOM), empirical orthogonal functions (EOF), and cluster analysis (CA), to investigate winter rainfall in South-West Western Australia, (SWWA) and settled on SOM. A step change in rainfall in 1975/6 was shown. This has been attributed to induced modification of storm tracks, as has a step change between 1995 and 1997.

Jones (2012) has used a bivariate test, the MYBT (in parallel with STARS), and in developing MYBT also trialled the Lepage test (Vives and Jones, 2005, Lepage, 1971). Others (Cai and Jones, 2005, Frederiksen and Frederiksen, 2007, Hope et al., 2006, Schneider, 2004) have used methods for detecting abrupt shifts from a number of disciplines such as homogeneity testing, signal processing, or ecosystem regime shifts (Andersen et al., 2009, Ducré-Robitaille et al., 2003, Harper, 2014, Tsay, 1988), and shown that abrupt regional climate shifts seem to occur.

Methods and approaches used for detection of changes attributed to variability modes involve many methods. Their application to detection of step-like changes is considered further down.

There is a substantial body of literature which has been published, tested and accepted and which includes methods for many statistical models, for single changes and multiple changes. Most but not all multi-change methods use information criteria for model selection, MSBV does not.

Two things are important in respect to the use of a statistical step-detection test. The test must reliably detect steps, and it must be possible to test whether or not they are deterministic.

Detection methods have been variously validated using simulated data, observations plus meta-data, observations and post-hoc examination, case studies and homogenised data with artificial perturbations (Peterson and Easterling, 1994, Easterling and Peterson, 1995).

Almost all of the detection tests used to date assume that the trends and shifts in the data are deterministic, and the non-deterministic components of the data are adequately modelled as normally distributed random noise with or without autocorrelation. However a number of papers attribute regime shifts to the effects of non-determinism in the data, or cannot reject it.

After this, the literature that addresses red-noise in climate is briefly reviewed.

Deterministic and non-deterministic steps in climate data

Whilst smoothly varying processes in climate have mostly been treated in the climate literature as deterministic, the nature of step-like changes is not as widely accepted. Non-determinism in the form of red-noise has been suggested as an alternative to shifts. Red-noise, or integrated white noise, treats a signal as a sequence of random changes added to the previous values and thus the signal is non-deterministic – all variation is random.

Red noise may (a) deceive a step-change method into identifying a shift in a pure red-noise sequence, or (b) inflating the statistical significance of detected shifts.

When regime shifts, signalled as abrupt changes in level, are specifically considered, Rudnick and Davis (2003) show that the composite method (ASD) used by Hare and Mantua (2000) may yield shifts in pure red-noise. Therefore detection alone is not sufficient evidence of a regime change. This is a theme of this thesis.

Percival et al. (2001) applies fractionally differenced (FD) and lag 1 auto-regression (AR(1)) models to the NP index, concluding that regime like progressions cannot be eliminated whilst leaving open the possibility of long memory processes with multiple lags. Furthermore a similar issue exists with area averaging of data, as is also shown in Chapter 4 of this thesis.

The problem of red-noise and regime shifts was also considered by Rodionov who proposed prewhitening,(estimation of and adjustment for autocorrelation) but also found that this decreased the sensitivity of his STARS method (Rodionov, 2006b, Rodionov, 2006a, Rodionov, 2005, Rodionov, 2004).

GMST curves do not merely follow red-noise/random-walks. General red-noise progression was used as a null hypothesis by Trenberth and Hurrell (1994) in determining central estimates of the frequency of decadal oscillations. Such a test neither proves constant frequency, nor addresses step-like changes, but it is evidence against purely red-noise/random-walk signals. See also Boiseau et al. (1999) for a similar analysis of coral reef proxies, and Cai and Whetton (2001) for a similar treatment of Nino 3.4 SST anomalies.

Pierce (2001) examines the North Pacific over the top 50-100m. He is sometimes cited as attributing regimes to red noise, but whilst he states that much PDO variability can be explained as noise, there are some deterministic features present explained by atmosphere-ocean coupling.

Newman et al. (2016) suggest that the PDO, rather than being red-noise like, follows a long memory model. They also suggest the PDO itself is a composite process and hence that different regime changes may result from different processes. Beaulieu and Killick (2018) however find the PDO to follow a short memory (AR1) model in contrast to forced mean shifts. They start by attempting to address the so-called hiatus as a *significant* slowdown in warming [emphasis mine]. They find the so-called hiatus to be unlikely and the PDO to be best represented as an AR(1) process without shifts.

One should note that short and long memory processes imply very different underlying physical models.

Kaufmann et al. (2006) addressed the detection in red-noise data with the Augmented Dickey Fuller (ADF) test (Elliott et al., 1992, Dickey and Fuller, 1981), concluding that structural-breaks (change-points) exist in time-series of GMST, CO_2 and CH_4. Finally this thesis proposes the use of the Zivot-Andrews test (ZA) (Zivot and Andrews, 2002) to address the specific possibility of a false determination of a change-point in pure red noise data (Chapter 4).

Reviews of step-change detection methods

The method of choice in this thesis is the MYBT as a single step detection test and the MSBV developed from it. The test itself has been used over some decades. In the meantime other methods have been proposed. Here I survey reviews of methods reported in various climate studies.

Tayanç et al. (1998), as part of a study of homogeneity in Turkish temperature observations, compared a number of homogeneity tests for temperature data, separating those which use observing station meta-data from those which do not. Whilst they preferred the Kruskal–Wallis homogeneity tests or the Wald–Wolfowitz runs test, supported by Monte Carlo testing, they concluded that meta-data was required to assess inhomogeneities. They made little distinction between non-meta-data based tests including the MYBT test, so the relevance of their work to this thesis is limited to their use of meta-data in validation of their choice of method.

Ducré-Robitaille et al. (2003) compared techniques for detection of discontinuities in temperature series with homogeneity methods using synthetic data. Of eight methods they assessed, three performed reliably, the standard normal homogeneity test (SNHT) of (Alexandersson, 1986), multiple linear regression (MLR) (Vincent, 1998), and Bayesian with reference (Perreault et al., 2000). They also noted that SNHT and Bayesian with reference performed better at detecting smaller discontinuities and that methods developed to (also) detect trends are disadvantaged when a small discontinuity exists towards the middle of the series.

Wijngaard et al. (2003) reviewed four methods, the SNHT, the Buishand Range test (Buishand, 1982), and the Pettit test (Pettitt, 1979) all of which are tests for a single change in level, and the Von Neumann ratio test for homogeneity of variance (Von Neumann, 1941). They referred to the first three as *absolute tests*. Sahin and Cigizoglu (2010) contrasts MYBT with the three *absolute tests*, since the bivariate reference series may add power. However in introducing the SNHT for rainfall data Alexandersson (1986), makes it clear the test uses references from nearby stations as does Alexandersson and Moberg (1997).

Mantua (2004), briefly reviews five methods from the point of view of detection of regime shifts in fisheries. These methods are PCA, average standard deviates (ASD), Intervention analysis (IA) and autoregressive moving average models (ARMA), vector autoregressive models (VAR(1)), and Fisher Information (FI). Most of these methods had a short history at the time of writing. This review is based on re-examination of sub-sets of data previously considered (Hare and Mantua, 2000), and concentrates on the issue of autocorrelation, which had been raised as a confounding issue in the use of ASD, IA and VAR(1) methods "… *provide[s] objective means for identifying and assessing statistical significance of regime shift years* …".

Rodionov (2005), lists a large number of methods for detection of (a) a level change in data with brief summaries of each. This work also tabulates method for detection of, (b) shifts in variance, (c) shifts in frequency structure, and (d) shifts in the system. Note that four methods reviewed by other reviewers as change-point methods, are classified in the last category. These are PCA, ASD, Fisher Information, and VAR(1). Again the MYBT test is not mentioned by name or as the Potter test.

Reeves et al. (2007) performed a review of methods concentrating on change-points in time-series of climate data, including the SNHT and a non-parametric version, the two-phase regression (TPR) (Solow, 1987), without a common trend, and with (XLW), before introducing their own new generalised algorithm (GNL). They tested against five error models with combinations of trend and step – from no tend and no step, to change of trend and a step – and concluded that no single method was preferred if the underlying error model was unknown. They did not review the MYBT method, although they do consider the role of a reference series in change-point detection. In reference to multiple change-points, they say "Many practitioners merely search for the most obvious change-point, correct for that (if one is found), and then reapply the method to the corrected series. This can lead to erroneous adjustments, because the effects of the first change-point are heavily biased when other unaccounted change-points are present… all possible change-point times should be identified jointly before their mean shift magnitudes are estimated". This issue is also addressed in Chapter 3 where the joint identification is given considerable attention.

Beaulieu et al. (2008), reviewed and tested a number of homogenisation methods which allowed neighbour comparisons or reference series to be included. Their selection included SNHT, MLR, TPR, MYBT, Sequential Student Test (STUS) (Gullett et al., 1990), modified in line with recommendations in Ducré-Robitaille et al. (2003), Jaruskova's method (JARU) (Jarušková, 1997, Jarušková, 1996), and a Bayesian method (BAYE). They found high levels of false detection for BAYE, lower rate of 5% for STUS, with MLR and JARU in the range of 1-5%. Detection problems were near the extremes, an issue addressed by Jarušková (1997)[1]. When a single shift was present most methods except STUS and TPR were deemed satisfactory, when two shifts were present the MYBT performed best and when three were present BAYE performed better with MYBT second. They do not specify a segmentation method, and given the findings of Jarušková (1997) this finding may not mean a great deal.

Andersen et al. (2009) reviewed methods and approaches for detecting threshold and regime shifts in ecological data. They include both single dimension and spatial methods. They make

[1] And addressed in this thesis. I was unable to find much work documenting this issue, but Chapter 3 includes sensitivity testing showing that MYBT is vulnerable to a central bias towards the data extrema.

the point that "… although most ecological regime shifts are inferred from changes over time, time itself is never the actual driver … Identification of a change-point in time is therefore the natural first step towards identifying a potential driver …". By "approach" they refer to either exploratory or inferential statistical approaches, and by "method" they refer to algorithmic bases. Included in this review are three methods which can be used for homogeneity analysis, STARS (Rodionov, 2004), Structural Change (Zeileis et al., 2001) and a commercial CUSUM/bootstrap product, Change-point analyser (Taylor, 1997). The MYBT was not included. Whilst no direct recommendation was made, they surveyed uni-variate and multi-variate relations and distinguished between inferential and exploratory methods.

Rao et al. (2009), in work drawn from the image processing community, analyse methods for anomaly detection in dynamical systems, comparing them with symbolic dynamic filtering (SDF). They include a number of neural network approaches, although SOM is not amongst them. They include, as feature extraction techniques principle component analysis (PCA), of which EOF is a variant used in climatology, and kernel regression analysis (KRA). This review is mainly of interest for the reason that it is framed in terms of pattern analysis and feature detection.

Beaulieu et al. (2008) compared eight homogeneity tests, concentrating on precipitation data in southern and central Quebec. Their data had potentially multiple change-points. Whilst they state that they dealt with this segmenting the data, this is further explained in a second paper (Beaulieu et al., 2009). Other authors (e.g. Vives and Jones (2005)) have also used researcher choice in segmenting the data. This issue is central to chapter 3.

Beaulieu et al. (2012), introducing their informational approach, based on use of the Schwarz information criterion (SIC) and the use of two decision rules, review a number of other classes of methods. The refer to change-point analysis for detection of changes in variance (Killick et al., 2010), various linear regression techniques, multiple shifts (Rodionov, 2004, Gérard-Marchant et al., 2008, Seidou and Ouarda, 2007) although they include a piece-wise method (Tomé and Miranda, 2004) which is method for fitting without discontinuities, as is the MCMC based approach of Cahill et al. (2015). They also reviewed non-normality and autocorrelation effects, including Seidel and Lanzante (2004).

Domonkos et al. (2012) provide a short history of statistical homogenisation and homogeneity testing tracing back to the identification of a need identified for temperatures of Milan (Italy) between 1763 and 1834 (referenced as (Kreil, 1848)). They credit Conrad (1925) as providing the first documented method for detecting an inhomogeneity, based on splitting annual precipitation data into two segments and examining the most significant difference. They briefly mention the Craddock test (referenced as (Craddock, 1979)), and many of the tests previously mentioned. Two points of interest are; (a) combination tests to detect shift-like and trend-like inhomogeneities (MLR or newer versions of SNHT) were consistently shown to be less efficient for homogenisation than the best of other methods, (b) likelihood ratio tests (including MYBT) are more sensitive than other approaches. However they seem to include Solow (1987), who restrains a TPR to perform as a broken-stick regression, as well as the MYBT test which is quite different. (See below)

Yozgatligil and Yazici (2015) also used simulated data to assess eight homogeneity tests, considering but not testing the MYBT test, detection tests but elected to consider the SNHT, TPR and the Buishand Range test. They also include KPSS (Kwiatkowski et al., 1992) and ADF tests (see further on these tests below) as they are sensitive to inhomogeneities and the Structural Change (SC) methods (Zeileis, 2005, Zeileis et al., 2001).

Points made by several reviewers are that change detection is intrinsically more reliable for changes located at the centres of data, and that information from reference series increases sensitivity. Also, that correct location of combined step and trend-change is an open problem. Few authors considered methods for segmentation of the data.

The literature surrounding detection of changes in time-series summary statistics, i.e. change-point analysis, is thus quite extensive.

The MSBV is a multiple change-point extension of the MYBT method utilising a reference series for power.

Utilising a reference series

As above, reference series have been found to increase the power of testing. In general reference series are intended to represent the statistical properties of the data in an unperturbed state. In homogeneity tests of rainfall for example, the average of surrounding rain gauges may be used.

Young (1993) used the MYBT for the adjustment of multiple discontinuities in 160 mean sea-level pressure (MSLP) times series focussing especially on Darwin, basing his choice on the review of Easterling and Peterson (1992). He used a variety of interpolation methods to generate reference series and found that the skill of the interpolation was key in the sensitivity of the MYBT, also finding the sensitivity of the test depended on how close the discontinuity was to the end of the data and the length of the series (a general domain problem since commented on by a number of people in reference to many methods).

Easterling and Peterson (1995) is one of the earlier review of homogeneity methods distinguished by concentration on methods for generation of suitable synthetic time-series. They review the MYBT test as published by (Potter, 1981), and the Alexandersson method (SNHT) (Alexandersson, 1986), along with CUSUM (Page, 1954, Brown et al., 1975) and double-mass methods (Chang and Lee, 1974), (referenced by Domonkos et al. (2012) to Köhler (1949)) and conclude that their method is more robust than others tested. However later reviewers, for example Andersen et al. (2009), seem to differ in this assessment.

As seen in the review by Easterling and Peterson (1995), number of change-point detection methods use a reference series, and this adds sensitivity to the testing (Jarušková, 1997). The MYBT, and the Alexandersson test a.k.a. SNHT, both conceived for precipitation time-series do. CUSUM and double mass (Köhler, 1949) generally construct reference series from neighbouring data, two phase regression ((TPR) Liu et al., 1997) and other regression techniques can also incorporate reference series. Note: MYBT marginally outperforms SNHT in this study. Peterson and Easterling (1994) concentrate on the construction of homogenised reference series, with the expectation that the reference series estimates the “true” signal, as is appropriate for homogeneity testing. The problem of testing for homogeneity is being increasingly addressed with sometimes mixed results (Domonkos et al., 2012), a later version of SNHT which allow for trend changes for example is reported as less sensitive.

Andrews and Fair (1988), and Andrews (1993) review structural change in non-linear models. They cover estimators including maximum likelihood and *M*-estimators and the Wald (W) test, Lagrange multiplier-like (LM) tests, and a likelihood ratio-like test. For classic regression models, the Chow test is commonly used (Chow, 1960). The Chow test is the basis of popular structural-change (SC) methods. SC methods test whether a (statistical) model parameter is stable over time. Thus the null hypothesis is that a given parameter is stable over time, tested against the alternative that it is not, which means that in the general case selection of an optimal statistic is hard (Zeileis et al., 2010). In this latter work, model selection (essentially the optimum number of change-points) is based on a modified Schwarz information criterion (LWZ) (Liu et al., 1997).

Domonkos et al. (2012) finds that for homogeneity testing maximum likelihood (including MYBT) tests are in general the best performing approach.

Use of Maronna-Yohai test

Following the publication of the MYBT, the test was illustrated by Potter (1981) using precipitation data. Potter extended the published table of critical values, recommending it as a very valuable tool, and in fact the test is often referred to as the Potter test. Bücher and Dessens (1991) used a reformulated version of the test for detection of homogeneity breaks in Pyrenees temperature data (the derivation was not published but I have derived it, see Appendix 3.1). Jones (1995) selected it as a tool for homogenisation of temperature and hydrological data, noting that Bücher and Dessens (1991) accepted that provided the data and reference time-series were relatively homogenous (i.e. their regression relationship was maintained except for a single change) the criteria of the test were fulfilled.

Gan (1998) used the test to examine the relationship between maximum temperature and precipitation in the Canadian Prairies with temperature as the reference variable. Similarly Jones (2012), used the MYBT to assess changes in relationship between temperature and rainfall variables, applying relationships published by Nicholls et al. (2004).

Vives and Jones (2005) also published the MYBT with repeat testing with random data as the reference series, this converts the test from a relative one to an absolute one in the terms defined by (Wijngaard et al., 2003). At the same time it introduces elements of stochastic resonance where injection of noise assists in extraction of climate signals (McDonnell and Abbott, 2009, Moss and Wiesenfeld, 1995, Benzi et al., 1982).

Beaulieu et al. (2008) found that the MYBT, SNHT, JARU all performed similarly whether data had single or multiple shifts although two Bayesian methods were later found to give greater precision (Beaulieu et al., 2009).

A paper by Boucharel et al. (2011) is of special interest because they used the MYBT to delineate regime changes in ENSO Studies, finding evidence of state dependent ENSO statistics and a see-saw like variability in the tropical Pacific. A later paper, combining sea surface heights and temperatures in the East and West tropical Pacific also shows a surface see-saw signal associated with El Niño (Peyser et al., 2016). This may be an independent finding since the prior paper is not cited in the later one. In very recent work Jones and Ricketts (2019) use the MSBV and a tracking model to show evidence for a state dependent heat pump operating across the tropical Pacific.

The MYBT has been used since its publication, found to be fit for purpose and amongst the better methods when judged by criteria such as sensitivity, precision, and resistance to use outside its ruling assumptions. Issues such as loss of sensitivity towards end points are domain limits.

Many recent papers use MYBT as one of a suite of detection methods for detection of inhomogeneities (Mahmood and Jia, 2017, Hoy et al., 2018). Delvaux et al. (2019) however, cite the non-provability of the reliability of the reference series, and the single change-point detection of such tests as mitigating against their use, preferring their own software called HOMER.

Extending change-point detection to multiple change-points

Methods which specifically deal with multiple change points come from both the econometric and the climate literature.

A number of authors have used univariate and multivariate SC methods based on a package coded in R (Zeileis et al., 2001), for analysis of paleo-climate (Martínez et al., 2015), fishery stocks and cascading regime shifts in the Irish Sea during the 1980s (Lynam et al., 2011), grapevine phenology, showing shifts in 1990-91 in the Veneto region of Italy (Tomasi et al., 2011) and attribution to climate change (Di Lena et al., 2010), Baltic Sea temperature trends, and trend change with step in Arctic sea-ice extent (Stips and Lilover, 2012). The method was used for the detection of climate regime change in temperature and rainfall studies of Italy by Giavante et al. (2009) who proposed a flat-steps model of climate change in that area. This was the method of choice in a paper on arXiv which proposed that evidence for step changes was evidence against anthropogenic climate change (Stockwell and Cox, 2009).

There are two rather different approaches with the same name, “changepoint” or “change-point”. Cahill et al. (2015) and (Foster and Abraham, 2015) adapted a method due to Carlin et al. (1992) who produced a method for the detection of multiple trend changes and named it “CP-regression”. The later work uses the names “change-point” and “changepoint” for their approach to statistical model of non-discontinuous segmentation, a “broken-stick”, or “segmented” regression. However there is a growing body of literature that considers all aspects of generalised multiple change-point methods (Fearnhead, 2006, Killick et al., 2010, Eckley et al., 2011, Killick and Eckley, 2011, Killick, 2012, Killick et al., 2012, Killick and Eckley, 2014), and a web-site (http://www.changepoint.info/). The package, coded in R (Killick and Eckley, 2014) provides methods for detection in level, trend and variance allowing for multiple search methods. They do not cover changes in regression (combined intercept and trend) as such, nor change in autocorrelation for which they recommend AutoPARM (Davis et al., 2006).

The method was used as part of a comprehensive examination of monthly precipitation in south-eastern United States (Wang et al., 2014).

The next sections consider how various detection methods were validated, and what further validation would serve this thesis.

Validation of change-point methods for climate

When developing a new test or adapting one to a new use, there is a need to ensure that the method performs acceptably, given the characteristics of the data and expectations of the use. None of the methods papers I encountered or the reviews of methods ignore this issue, but it would seem that each researcher developed data and methods independently. All used some synthetic data, but not all used real data.

The use of meta-data was proposed at least once. In general, where validation was reported, synthetic data for testing detection of shifts consisted of random data, sometimes with a degree of autocorrelation, and defined shifts and defined trends and trend changes. However climate data is more structured than this. Occasionally authors reported using homogenised climate data to which known perturbations were added, assuming or testing that the homogenised data did not then contain occult confounding change-points. This would have the advantage of testing a method against more structured data than usual synthetic data.

Validation when multiple change points are possible requires not just an assessment of the sensitivity of the detection method to noise but to end effects. Validation of detection of change-points in multi-variate relationships is still more complex; as seen in methods developed for streamflow work (Seidou and Ouarda, 2007).

Table Ch2.1: Validation methods used for establishment of various change-point methods. This summarises the means by which selected methods were validated to ensure fitness for purpose. Does the paper indicate the method was tested against synthetic data? Was it tested against data containing only mean changes? Was it tested in the presence of trends? If real data was tested was any meta-data used to inform interpretation of the test? Was it tested against real data with previous or known results? If real data are homogenised and then shifts are added are they detected? Were case studies published?

Paper	Method (possibly with comparisons) or Review only	Simulated or synthetic data	Means tested	Trends tested	Multiple Change-points	Use of meta-data	Performance metric	Real data assessed	Homogenised real data plus synthetic	Case studies
(Peterson and Easterling, 1994, Easterling and Peterson, 1995)	Method	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
(Jarušková, 1996)	Method	Yes*	Yes	Yes	No	No	No	Yes	No	Yes
Seidel and Lanzante (2004)	Theory/ method	No	No	No	Yes	Yes	Yes	Yes	No	Yes
(Rodionov, 2004)	Method	No	Yes	Yes	No	No	No	Yes	No	Yes
(Fearnhead, 2006)	Method	No	No	No	Yes	No	No	No	No	Yes
(Seidou and Ouarda, 2007)	Method	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
(Beaulieu et al., 2008)	Method	Yes	Yes	Yes	Yes	No	Yes	No	No	No
(Gérard-Marchant et al., 2008)	Review	Yes	Yes	Yes	No	No	No	Yes	No	Yes

*For derivation of critical values.

In Table Ch2.1 above, the validation methods as gleaned from a range of papers is tabulated. Case studies are not considered as validation. Papers which are silent on validation are not listed. Econometric papers are not considered here since the thesis domain is climate.

Columns indicate the following

- Does the paper indicate the method was tested against synthetic data? Although presumably developers of method do so during development, an indication of what data was used is helpful.
- Was it tested against data containing only mean changes?
- Was it tested in the presence of trends? Trends are an important nuisance parameter in testing changes of mean.
- If real data was tested was any meta-data used to inform interpretation of the test? This applies especially to homogeneity tests and validation of homogeneity tests as there are multiple sources of inhomogeneity.
- Was it tested against real data with previous or known results? Does it behave as expected, improve or shed light on a previous method?
- If real data are homogenised and then shifts are added are they detected? This is a method for testing performance in data as close as possible to the target domain but the homogenisation should be conducted with care.
- Were case studies published?

From this table the works of Peterson and Easterling (1994) and Easterling and Peterson (1995) would seem to have covered the validation of methods in more detail than most.

One author in particular (Jarušková, 1996) highlights the technical difficulties associated with testing for multiple trend changes, and with detecting changes at the data extrema.

Criticism of an earlier paper submission not only suggested that further sensitivity testing of the MYBT was warranted (now addressed in Chapter 3), but raised the entire issue of statistical induction. This is addressed in JR2017. In that paper severe testing (Mayo and Spanos, 2006) is used to unite the hypothetical part of the work and the findings, and a philosophical proposal, the Theoretical-Mechanistic/Statistical-Inductive (TMSI) is introduced.

This thesis also borrows inspiration from an approach to misspecification testing (Mayo and Spanos, 2004), and as a result Chapter 4 is dedicated to post detection analysis of the data in which change-points are detected.

Post detection tests

A ruling issue in this thesis is that detection of abrupt shifts by any method is based on assumptions about how physical systems ought to operate, how that relates to data collected about them, and what alternative hypotheses should be considered. Similarly use of the tests documented here does not ignore the assumptions about the data that frame the tests, rather, these ruling assumptions are incorporated into the interpretations of the tests.

Post detection testing serves two purposes. The first and most immediate is the detection of possible false positives. The second is to do with identifying non-determinism.

The need for the detection of false positives in this thesis is due to the potential for misspecification between climate data and the ruling assumptions of the step-detection method, and so a less sensitive, presumably better specified test is performed post detection. The appearance of non-determinism can come about either because false negatives or sub-detectable events exist, or because the underlying process is in fact dominated by red-noise. The impact of non-determinism is a difficult issue because there are trade-offs between the ruling assumptions of the tests and their power. As seen in Chapter 4, the thesis follows an error-statistical, misspecification approach in discriminating the statistical results (Mayo and Spanos, 2004).

Once a multiple change-point method has selected change-points, it may be that portions of the data are affected by data quality issues that make individual change-points suspect.

The MYBT assumes statistical stationarity (with no more than one exception at one point in time). Other detection methods are similar in assuming limited stationarity. Hence the detection and classification of non-stationarities in the data is important.

There is a small but important body of literature which focusses on the validity of the statistical methods. There has been something of a tension in the literature about multiple testing. For instance recent papers that specifically claim change-point methods are subject to the fallacy of multiple testing (Rahmstorf et al., 2017, Cahill et al., 2015), but this was anticipated and previously rebutted (Mayo and Spanos, 2004). There is also a growing body of literature which attempts to refine or ameliorate the uncontrolled use of null hypothesis testing. For instance Haig (2016) argues that NHST should not be used in science, preferring what he calls neo-Fisherian (the role of experiment is to solely attempt to disprove the null hypothesis on the basis of the data, without selection of an alternate, and without the concept of type 1 error) or error-statistical approaches, owing to Mayo (1996) (of which the severe test (below) is a component). It is often assumed that autocorrelation in climate data should be compensated for on the basis that it can inflate regression significances , but this has been disputed (Mizon, 1995), and if anything continuing autocorrelation should narrow the confidence bounds on the timing of an abrupt change.

Whilst a number of papers have used regression methods to detect and remove natural variation in order to characterise trends (Rahmstorf et al., 2012, Foster and Rahmstorf, 2011), up until now few use econometric unit root analysis to investigate climate data. Socio-economic modelling of climate risks has used these methods (Liddle and Messinis, 2015).

Probative Tests

Once a sequence of change-points are found, and the chosen statistical model signal is applied (segmented, dis-joint, or curvilinear), the residual series should be featureless if all assumptions hold. Systematic variation, may be detectable by White's test (White, 1980), as used in JR2017, or equivalently the studentized Breusch-Pagan test (Breusch and Pagan, 1979) used in this thesis. This might be taken as a signal of an unaccounted for variable or parameter, or a model misspecification.

In this thesis, the detection method assumes change-points marked by step-like shifts, but the analysis allows for trend like changes. Error probabilities from the detection are inappropriate. Equally, error probabilities based on a misspecification are invalid.

Analysis of variance (ANOVA) can be used to test the specific hypothesis that either a change of offset, or a change of trend exists as was the case in Rahmstorf et al. (2017), but interaction between the two parameters must be taken into account (Walpole et al., 1993). ANCOVA (Tabachnick and Fidell, 2007) can be used to assess the impact of the assignment of specific change-point relative to no change, essentially replicating the Chow test (Chow, 1960) within a more general framework.

Unit root tests

Unit root testing was stimulated in the econometric literature as part of the consideration of an economic assumption, "The Unit-Root Hypothesis", that macroeconomic series follow a unit root (e.g. a random walk) process with random economic shocks being persistent (Nelson and Plosser, 1982). Investigation of this claim, led to the development of a number of statistical tests, each with its own assumptions and uses.

Unit root behaviour applies to data which has autocorrelation plus random noise, and which converges at least transiently on a fully integrated, maximally auto-correlated, white noise process (Stock, 1994, Glynn et al., 2007, Kaufmann et al., 2006). Perhaps more simply (Stern and Kaufmann, 1997) put it thus, "Time-series can be characterized in many ways. … There are two types of trends - deterministic trends and stochastic trends. A stochastic trend is a random walk process that may or may not contain deterministic or stochastic drift. *A time-series that contains a random walk process is termed a unit root process*" [emphasis added]. This can lead to the identification of a chance period of rapid unidirectional random change as a step change in the mean, and a drift like behaviour as a change in trend. It is also possible that the temperature series may be dominated either transiently or in total by unit root behaviour. Stock (1994) identifies two categories of unit roots, moving average (MA) and autoregressive (AR) and identifies them as integrated lag zero or I(0), and integrated lag one or I(1) respectively. Data can only be taken to be stationary if the roots of their characteristic equation (which includes autoregressive components) are all less than one. Mathematically, a unit root process is one where at least one root of the characteristic equation is unity.

Transient unit root behaviour could indicate some sort of regime change as part of the system behaved as if it was decoupled from normal forcings, I(0) behaviour; or as if it were forced by lagged change as I(1) behaviour, or more coupled to other sub-systems (Tsonis et al., 2007). Importantly in this work, it would indicate a potential deception if a time-series contained both a unit root (an endogenous feature) and an exogenous change.

There is now a growing literature (Estrada and Perron, 2014, Estrada et al., 2013) concerned with applying econometric methods to climate data. Unit root tests have been used to analyse HadCRUT3 global and hemispheric data for structural change (Coggin, 2012). They have been used to analyse for structural breaks in the greenhouse gas and GMT time-series, concluding that climate change has affected the mean of the series but not the variability, but differentiating between trend changes in the Northern and Southern hemispheres (Stern and Kaufmann, 1997).

Here I introduce three methods sourced from the econometric literature to assess the evidence for various manifestations of unit root behaviour. The outcomes of the tests are reinterpreted in the light of the ruling assumptions about each test, as evidence for non-stationarity.

The ADF tests a null hypothesis of unit root against an alternative of stationarity after compensation for auto-correlation (Elliott et al., 1992, Dickey and Fuller, 1981). It has been used to assess stationarity in climate series, in remote sensing of vegetation indices (Goetz et al., 2005), and a study of Granger causality and comparative hemispheric response (Kaufmann and Stern, 1997).

The KPSS test is a test for either trend stationarity or level stationarity against an alternative of unit root. The sense of the test is inverted compared to the ADF test since the null hypothesis of each is the contrast hypothesis of the other.

The Zivot-Andrews test (ZA) tests for the presence of unit root against an alternative of exogenous change. Gay-Garcia et al. (2009) included it in a survey of the statistical properties of climate models. They showed it to be susceptible to the rejection of the null of a unit root in the presence of a structural break, an expected result, and they preferred to use a Perron-Yabu test (Perron and Yabu, 2009). However this work uses complementary tests, testing both step and trends, whilst the independent estimate of time provided via the ZA is of interest. Perron (1989) argued against the "The Unit-Root Hypothesis" (above) and proposed a test for the presence of structural breaks (i.e. exogenous changes) at a prescribed time in the presence of possible unit root, concluding in part that macroeconomic time-series identified earlier as following a stochastic path with persistent shocks were more deterministic with persistent structural breaks and transient stochastic ones. Zivot and Andrews (1992) revisited the test, internally computing a break-time. Thus it is a transformation of Perron's UR test which is conditional on structural change at a known point, into one which first must select a candidate change-point.

Step-like changes in a time-series are by definition non-stationary, hence would not be expected to pass tests for level stationarity. The contrast hypothesis of KPSS tests is formally unit root behaviour, but the ruling assumption ignores exogenous changes such as steps and trend changes; similarly the contrast hypothesis of the ADF. Thus the unit root result needs to be interpreted, and also requires further confirmation. The ZA provides further discrimination in this case, but it too requires careful interpretation.

Interpreting unit-root tests given presumptive steps.

These tests are all framed as null hypothesis statistical tests (NHST) with unit-root as one alternative. I reinterpret them as evidence of stationarity or non-stationarity and secondarily attempt to classify non-stationarity as more likely due to deterministic (e.g. undetected or ignored changes), or non-deterministic (e.g. random walk).

1. Such tests have the following components

a. Ruling assumptions – the specific statistical behaviour that is shared between null and alternate hypotheses, the nature of error terms etc. For example in this work, significance testing of a trend will include the assumption “The data contains a signal and noise, the noise is i.i.d., this signal can be represented as y=a+bx.” The ruling assumptions include the presence of a limited number of alternates (usually one).

b. A statistical function and associated probability, the value of which is compared to a “significance” and used to decide between alternative hypotheses. The probability is mapped to the statistical function, and it is this mapping which depends on ruling assumptions.

c. The null hypothesis and a single alternate, a threshold probability or significance often treated as a behavioural rule (the researcher will accept or reject H0 in favour of H1 based on the probability value).

2. Most tests in use attempt to accumulate evidence against the null, treating it as evidence for the alternate, and the probability represents the likelihood that the null case holds given the evidence against it. Short data sets tend to reduce the opportunity to accumulate evidence.

3. Climate signals can almost never be completely described as a single signal and noise.

4. If the ruling assumptions are violated, as in the real world they mostly are, the effect on the decision process depends on the nature of the violation, the data etc. The statistical function changes its relationship with the probabilities, and this should be accounted for.

For instance the KPSS test for trend stationarity, a unit root test, has

I. Ruling assumptions that include the absence of deterministic steps, which is important in this work, but also absence of other features. It also assumes that the signal can be decomposed into a linear combination of deterministic trend and a random walk (the sign of a unit root) and i.i.d. error.

II. A null hypothesis of stationarity, and an alternative of unit root.

If a shift is present then the method will effectively treat this as non-trend stationarity, hence as a unit root. I reinterpret the test as (H0: Data are trend stationary, H1: data are not trend stationary). Given the issue of small data sets, H0 has a rider, “or insufficient evidence exists to reject non-stationarity”.

The ADF test assumes unit root, and rejects it in favour of trend stationarity plus an autocorrelation structure and i.i.d. noise.

I. The ruling assumption still includes absence of steps or trend changes and other features.

II. The null is unit root, after accounting for autocorrelation, against the alternative of trend stationarity plus autocorrelation.

Also note

a) The values given to trend and the autocorrelation lags are all affected if a shift is present hence it may be less likely to class a shift and non-stationarity.

b) Computing the autocorrelation reduces the sensitivity of the test and it is more applicable to longer sequences.

The test is reinterpreted as H0: Data are not trend stationary (and this is not due to autocorrelation). H1: data are trend stationary (with or without autocorrelation). H0's rider is "or insufficient data exists to account for autocorrelation and trends"

The Zivot-Andrews is actually the test that is most applicable since a provisional change-point is present. However the ruling assumptions must be noted.

I. The ruling assumption is that there is at most one deterministic change-point. If there is then the underlying process must be stationary. **Note:** the test first removes its own estimated change-point and then operates much as the ADF would on the residual series. Also note its internal change-point may also be a trend-change or a strong transient.

II. H0 is unit root with drift (interpreted as endogenous change). H1 is a single deterministic change point with stationarity (interpreted as an exogenous change)

   Violations of the ruling assumptions may give false findings either way.

III. The presence of more than one deterministic change-point in the absence of unit root can register as a unit-root.

IV. The presence of a deterministic change in the presence of unit root can register as stationarity.

V. In the absence of any obvious shift, the test operates as a simple unit root test similar to the ADF.

VI. Because the test is being applied where there is a presumptive step already identified, I reinterpret the test.

    a. If it returns stationarity it may still be non-stationary data with a deterministic change. If it returns non-stationarity, then either the step-change is a misidentification (in which case the residuals after the step is removed will still be non-stationary), or at least one more change-point exists (in which case there are three possibilities for the residuals (see 8))

    b. If there are possibly two or more change-points then the residuals will still contain at least one change-point, and the underlying residuals are either stationary or non-stationary.

Application of the tests to residuals

In Chapter 4, I also apply the same tests to the residual after the internal trend and shifts are accounted for. The same interpretations apply as above For example, if the ZA test initially was interpreted as evidence of non-stationarity but the residuals are stationary the most consistent explanation is a single undetected change-point was present in the data in addition to the one found by MSBV.

The Theoretical-Mechanistic/Statistical-Inductive approach

This thesis utilises a framework outlined in JR2017, the theoretical-mechanistic/statistical-inductive approach, which requires a carefully reasoned matching between scientific hypotheses about the physical world with statistical hypotheses. There is good reason to do this. As Haig (2016) notes regarding psychology, “*tests of statistical significance (ToSS) … have been widely popular in psychology for more than 50 years and in statistics for more than 80 years*”, before going on to detail practitioner discomfort with ToSS and in particular with a specific framing of ToSS, null hypothesis significance testing (NHST). He makes ten recommendations the first of which is that NHST should not be used in research, the second favouring “defensible forms of ToSS” including *error-statistical* approaches (Mayo, 1996, Mayo, 2004).

Somewhat earlier than Haig, an often unrecognised limitation of reasoning was addressed by Mayo and Cox (2006) who say, ‘*The defining feature of an inductive inference is that the premises (evidence statements) can be true while the conclusion inferred may be false without a logical contradiction: the conclusion is "evidence transcending."’*, going on to delineate two *traditions of using probability citing (Pearson, 1955): " For one school, the* ***degree of***

confidence in a proposition*, a quantity varying with the nature and extent of the evidence, provides the basic notion to which the numerical scale should be adjusted." The other school … suggests that "it is through its link with* ***relative frequency*** *that* ***probability*** *has the most direct meaning for the human mind*" (ibid)' [emphasis added].

Under the degree of confirmation approach, probability is used to provide a *post-data assignment of degree of probability, confirmation, or belief in a hypothesis*, while in the second, probability is used to assess the *reliability of a test procedure to assess and control the frequency of errors in some (actual or hypothetical) series of applications (error probabilities or error-statistics).* The first is also called *degree of confirmation* (DC), the second, *error-statistical* (ES) (Mayo and Spanos, 2006).

The TMSI approach builds a hierarchy of models between theory and data following Haig (2016), and employs the concept of severe testing (Mayo and Spanos, 2006). It is explained in detail in Section 2 of JR2017.

Severe testing

Mayo and Spanos (2006) propose a severity criterion which supplies a meta-statistical principle for evaluating statistical inferences, where the *severity* of testing is not assigned to hypothesis *H,* but to the testing procedure. Severe testing is based on the intuition that "*Data x_0 in test T provide good evidence for inferring H (just) to the extent that H passes severely with x_0, i.e., to the extent that H would (very probably) not have survived the test so well were H false.*" (Mayo and Spanos, 2006).

As stated in the Introduction, JR2017 proposes two physical hypotheses, H1 and H2, describing the interaction of externally driven warming interacts with internally driven natural decadal variability with H1 holding the two are independent and H2, they interact. JR2017 proposes six tests in order to differentiate the two by severe testing. These are quoted in the Introduction under the heading "A probative framework".

Severe testing is beginning to be picked up, for example Katzav (2011) suggests assessment of climate model projections have not been severely tested and could be, for example to address issues of model tuning in such projections (Katzav et al., 2012). It was applied to an analysis of optimal fingerprint methods in climatology (Katzav, 2013). Severe testing forms a core component of JR2017 and a conference paper (Ricketts and Jones, 2017 henceforth RJ2017).

JR2017 lays out six tests that together formed a regime of severe testing for distinguishing H1 and H2. (In JR2017 "steps" are the values returned by the MYBT, approximating the difference in the means before and after a change; “trends” are computed by OLS for the data before and after the change and “shifts” are the dislocation at the time of change.)

Test 1 Patterns of step changes in observations and alignment with known events: Major regime changes in observed surface temperature data were reported in JR2017, globally in data from five different sources and in two differing zonally averaged data sets. The dates are commensurate with other findings (Bartsev et al., 2017, Yan et al., 2016, Yan et al., 2015, Reid et al., 2015), and these are expanded upon below. These likely originate in decadal and regional processes (Rodionov and Overland, 2005, Overland et al., 2008, Reid et al., 2015, McCarthy et al., 2015, Alheit et al., 2005, Freitas et al., 2015, Trenberth, 2015, Trenberth and Fasullo, 2013). When the effect of shifts was separated from trends, shifts predominated, more so in mid-latitudes and SST. Satellite temperature records showed changes that aligned with surface temperatures.

Test 2 Reproduction of observed patterns of step changes in GMST by models forced by historical emissions: A high level of correlation in dates was found between observations and an ensemble of Global Climate Model (GCM) analyses. 58 of 107 GCMs forced by an intermediate warming scenario, RCP4.5, showed shifts within one year of the observed change after 1996.

Test 3 What is the relationship between different components of change and equilibrium climate sensitivity (ECS) in GCMs: For the period 1861 – 2005, correlations (r^2) between warming and steps, shifts and trends was 0.87, 0.43, and 0.13 respectively, but with there was no correlation with ECS. For the period 2004-2095, r^2 values were 0.96, 0.54 and 0.49 with ECS values being 0.65, 0.52 and 0.18 respectively, so that trend was dominated by steps and shifts

Test 4 Can step-like change be identified using attribution methods: Regional attributions showed that warming commenced abruptly rather than smoothly in SE Australia, the UK and US and at times aligned to changes in the appropriate hemisphere.

Test 5 Do other climate variables also undergo step changes: Similar timing has been shown in tide gauges records, rainfall, ocean heat content, forest fire danger indices (Hennessy et al., 2005) and other climate variables and impact variables (Jones et al., 2013)

Test 6 Are temperature time-series more step-like or trend-like: By multiple metrics temperature time-series are more step-like than trend-like (see JR2017, Table 6 for a comprehensive set of metrics).

JR2017 satisfies this programme of testing, and this thesis address Tests 1, 2, 5 and 6. It is also important in the TMSI to ensure that counter-theories are addressed and that counter proposals can be discriminated.

In considering the role of error statistics (Mayo and Spanos, 2011) highlight the behavioural aspects of conducting ToSS and NHST testing. A decision must be made, based on some sampling, for example a batch of bolts produced in a factory (to use their example), is to be to accepted or rejected based on a sample from the batch. This is a *behavioural* decision and involves (or should involve) consideration of the consequences of the decision, and the properties of the test. When NHST is used in climate, the framing is generally “There exists a null hypothesis (usually that some dataset is randomly constructed), and an alternative that some specific feature exists. A statistical test is selected that provides a probability of the alternative as opposed to the null. (I will) reject the alternative in favour of the null case unless the probability falls below some threshold”. The acceptance or not is a researcher choice – not forced one by mathematics.

Applying Theoretical-Mechanistic/Statistical-Inductive approach to decadal climate

Climate time-series data are inevitably highly complex, and contains imprints of many features. H1 and H2 are often either silent, or in agreement, on many of these. For example both agree that anthropogenic warming entails inbound solar radiation (predominantly short-wave), to which the atmosphere is transparent, being absorbed and re-radiated back from the surface at infrared wave-lengths to which the atmosphere is not-transparent. A back of envelope calculation suggests one inbound higher energy photon has the same order of energy as 20 outbound ones. Both agree that absorption of the outbound radiation occurs in the atmosphere resulting in warming but differ sharply on what happens next. The difference in inbound and outbound radiation at the top of atmosphere has been measured at between 0.5 and 1.0 W/m^2 (Abraham et al., 2013). Both are consistent with estimated absorption into the oceans of about 70% of this radiative deficit. Both would stipulate that this involves transition through the atmosphere/ocean interface, i.e. the sea surface. Where they differ is that under H2 such absorbed radiation (at least a portion of it) interacts with systems responsible for what we measure as climate variability *and* such systems respond to this extra heat. Under H1, either the systems that control variability modes do not interact at all with the surface heat flux, or they equilibrate to it without it affecting their variability, and/or they dispose of it rapidly in some occult fashion.

Statistical model fitting

Statistical model fitting is used widely in climatology. Seidel and Lanzante (2004) analysed three alternatives to simple trends – segmented trends (which they called "piece-wise linear", a.k.a "broken-stick"), steps ("flat steps"), and step and trend models ("sloped-steps", which I tend to refer to as "disjoint segmented"), finding that no single model stood out on the basis of p-values. They concluded that detection and attribution studies should consider abrupt changes. If alternative statistical models cannot distinguish between H1 and H2 on the basis of *p*-values, then many of the standard tests are not severe enough. JR2017 lays out six tests which together form the basis of a severe test of H2 against H1. This thesis shows that the testing is adequate given the data, and that statistical and spatial features uncovered are consistent with prior observations, and with plausible mechanisms.

Detection and attribution

Following Jones, and as expanded in JR2017, the signal within the trajectory of change is apportioned statistically to a combination of trend-like changes and shift-like changes. These are taken to be reflective of two different hypothetical types of physical processes, one of which may give rise to abrupt shifts and the other not. Some terms are ambiguously used. "Jump", "step", and "shift" are all used to refer to the y-axis displacement between two trend-lines in a disjoint time-series. For the duration of this thesis, consistently with JR2017, a "step" is the difference in the mean between two zero-trend segments, and is the value estimated by the MYBT. In some literature this is also called a "level-shift". An "internal shift" or simply "shift" is the difference between two regression lines at the time of dislocation. The word "jump" will be used with respect to climate state as the system jumps from one state to another. "Internal trends" are the individual trends of the disjoint regression.

Detection and attribution are defined by the IPCC in the IPCC guidance paper, as follows; *"Detection of change is defined as 'the process of demonstrating that climate or a system affected by climate has changed in some defined statistical sense without providing a reason for that change. An identified change is detected in observations if its likelihood of occurrence by chance due to internal variability alone is determined to be small'…"* and *"Attribution is defined as 'the process of evaluating the relative contributions of multiple causal factors to a change or event with an assignment of statistical confidence'* (Hegerl et al., 2010)". The IPCC Fifth Assessment report goes on to explain, "*In general, a component of an observed change is attributed to a specific causal factor if the observations can be shown to be consistent with results from a process-based model that includes the causal factor in question, and inconsistent with an alternate, otherwise identical, model that excludes this factor*" (Bindoff et al., 2013).

Under H1, attribution of temperature increases to anthropogenic causes involves apportionment of response between anthropogenic trends and “natural” trends with other components, including shifts, being treated as noise or as nuisance parameters (Kaufmann and Stern, 1997, Stern and Kaufmann, 1997, Hegerl et al., 2010). Under H2, shifts and their timings become primary components of a complex signal, and attribution between natural anthropogenic causes involves shifts and their timing, and residual trends. Jones (2012) employs an inverse linear method involving maximum and minimum temperatures and rainfall to detect, and then attribute changes in linear relationships amongst these variables to climate regimes. JR2017 also separates model temperature time-series into shifts and trends, correlating these with ECS. After 2006, shifts explain 2.9 times more of the variation in ECS. JR2019 also a similar attribution scheme to show a differential contribution between the Eastern and Western Tropical Pacific to ECS in climate models. In the same paper, observed ECS is estimated as 3.2±0.6°C, in line with but with narrower uncertainty limits than given in an analysis of models by Andrews et al. (2015).

Model misspecification

This thesis’s use of misspecification testing (M-S) concentrates on determining that the model is adequate for purpose. Within the M-S approach, hypotheses of interest are “H_0: the assumption(s) of statistical model ***M*** hold for data ***z …*** against not- H_0.where not- H_0. Would consist of all the ways ***M***‘s assumptions could fail” (Mayo and Spanos, 2004). In practice one must consider specific departures from H_0.

The MSBV operates by using the MYBT as a detection mechanism, and the MSBV selects specific solutions from the family of solutions using a variation on what Mayo and Spanos call a “causal structure search”. M-S tests that the assumptions of the statistical model hold for the data. This is different from the often used approach of estimation, and removal of extraneous signals, using regression methods, often against derived indices such as the IPO index. Doing the latter makes an implicit assumption that the confounding signals are in constant fixed phase with their indices – essentially an assumption of H1. Autocorrelation is either dealt with similarly or by binning, and in general changes of autocorrelation are not considered as signals. In this thesis the approach is to bin to annual averages as a means of minimising autocorrelation, and to examine a segment of data within which a change point is found, and determine if the signal itself, and the residual after removal of internal shifts and trends, show signs of deceptive non-causal drifting behaviour. Detection is based on level changes, post detection assessment is based on a less constrained model, so that the issue of cannibalisation is addressed.

Given the dependence on the pairing of physical and statistical models, misspecification testing (M-S) was proposed as an approach to determining whether the assumptions needed to reliably model statistical variables are met (Mayo and Spanos, 2004), for example whether an assumption of independent trials holds. The authors differentiate between model specification and model selection. *They say, "Far from increasing error rates, multiple tests, if appropriate, may serve to cross-validate and fo*rtify other tests, so that the model inferred as statistically adequate has passed a reliable test." Despite this, a recent paper suggests quite the opposite, that multiple tests of multiple hypotheses are dangerous, and they include serial model selection in this, suggesting Bonferroni adjustments (Rahmstorf et al., 2017). However Bonferroni adjustments are useful only in restricted circumstance, specifically when multiple independent causes are postulated (Perneger, 1998).

Serial feature detection in any time-series is a form of model selection from a family of related models, reliant on model-specification. The model selected can be coded as a set of time-points, and the model selection utilizes some metric of the time-series deducible from the specification. The coded model varies in complexity as features are added or deleted during model composition. Here, selection of individual features requires a DF approach whereas their admission into a statistical model is generally an ES problem. Emergence of model selection algorithms piles up significance test results (Mayo and Spanos, 2004). Practitioners have not resolved roles of error probabilities and therefore model selection methods not aligned to ES (ibid). This can lead to pre-data (ES) probabilities being applied to post-data degree of confirmation, ES design used for a DF problem, which require priors. For instance, use of the Akaike Information Criterion (AIC) and similar for model selection by ranking has been criticized as giving rise to unreliable inferences (Spanos 2010). However such information criteria control model selection in several popular systems for multiple change-point detection (Killick and Eckley, 2011, Zeileis et al., 2001).

Regime changes discussed in other recent publications

This section briefly surveys recent publications which demonstrate the uptake of the concept of regime shifts in climate, some definitional issues encountered as the concept is taken up across different domains.

deYoung et al. (2004) develop what they call a pragmatic definition of an ecosystem regime shift as an abrupt shift from a quantifiable ecosystem state". They state "While conceptually straight forward, understanding and identifying a regime shift concept has proven difficult. If we can define an ecosystem state, then it seems clear that a change in state, in the dynamical sense, should be measurable and that some such changes in state can be substantial, abrupt and persistent …”, also noting that “the concept is more easily defined and accepted in the physical modelling community”. Amongst their findings are regional scale climatically induced shifts in the NE Atlantic and North Sea due to upper ocean warming and advection, and very large scale physical changes related to top to bottom food web changes. When they looked at sub-millennial time scales they were unable to distinguish climatic regime changes from red-noise (Rudnick and Davis, 2003). Chapter 4 of this thesis is thus of special importance, as it feeds into later work which suggests that decadal variability is very likely to be deterministic.

Alheit et al. (2005) present evidence of an ecological regime shift when levels of copepod and other species changed abruptly in the North Sea and Central Baltic, coincident with NAO changes from negative to positive (1987-1989). The effects on copepods were attributed to the effects of temperature changes on water stratification, and severely affected other fish stocks.

Reid et al. (2015), in a comprehensive paper, find wide-ranging changes in and around 1987 involving most ocean basins and continents, multiple fish-stocks, and flowering dates. They also comment on an Eastward movement of the time of onset from 1985 in the Pacific to 1988 in Asia. They define regime shifts as “Regime shifts are abrupt, substantial and persistent changes in the state of natural systems.”

Belolipetsky et al. (2015) finds a staircase like signal in post 1950 HADCRUT4 global climate data after removing the estimated influence of ENSO. The adjustment for ENSO was conducted at the grid level which has the intended effect of removing regional biases. They find changes after 1987, 1988 and 1998. They did not find a change corresponding to 1976/77, although it is widely documented, this is discussed. I suggest that the adjustment method has differentially removed a signal present in the Pacific cold tongue (see their Fig. 3 (c)). Bartsev et al. (2017) follow similar methodology and speculate on the mechanisms of the three big shifts (excluding 1976), firstly considering four cases. These are (a) an unknown multi-stable parameter of the climate system, (b) multiple bi-stable systems, (c) transitions between attractors in a chaotic system, (d) interconnected oscillators forming a super-system and shifts are the transitions between states.

Varotsos et al. (2014) found evidence of two steplike changes in SST in the zone 30N-60N 1925/1926 and 1987/1988, using STARS. Varotsos et al. (2019) find abrupt shifts in the lower troposphere 1986, 1994 and 2014 by the same method. These results are commensurate with results published in JR2017 except that the data used did not allow 2014 to be tested.

Zonal analyses

This thesis extends the analysis used in JR2017 and RJ2017 to the remaining zonal records of observed climate.

It also extends JR2017 to zonal analyses of a large subset of the temperature records from the atmosphere ocean global models (AOGMs) submitted for the IPCC fifth climate assessment report (AR5) (IPCC, 2013).

Major regime changes in observed surface temperature data were reported in JR2017, globally in data from five different sources and in two differing zonally averaged data sets. The dates are commensurate with other findings (Bartsev et al., 2017, Yan et al., 2016, Yan et al., 2015, Reid et al., 2015). These likely originate in decadal and regional processes (Rodionov and Overland, 2005, Overland et al., 2008, Reid et al., 2015, McCarthy et al., 2015, Alheit et al., 2005, Freitas et al., 2015, Trenberth, 2015, Trenberth and Fasullo, 2013).

Zonal data were analysed in JR2017 as a means of assessing spatial coherence and prevalence, and shedding some light on specific claims. As stated above, Maher et al. (2014) associate Pacific tropical cooling with a negative phase of IPO in climate models, and Thompson et al. (2010) find a widespread, cooling over the Northern Hemisphere and North Pacific at the start of the 1970s. However in JR2017 and this thesis a regime shift comes after the mid-1970s in the Northern and Southern Tropics, and is preceded by a shift just before 1970 in the Southern mid-latitudes.

Zonal analyses are less frequently reported in the literature, and when they are it is often as part of a study of meridional influences of atmospheric warming over time (e.g. Lu et al., 2008, Gu et al., 2016). And, as in these papers, at much finer zonal scale than the datasets reported in JR2017. The difference between equatorial and polar warming rates was analysed using GISS zonally averaged temperatures to make the point that surface warming was not uniform (Belcu et al., 2015). As noted above the analysis of Varotsos et al. (2014) was conducted zonally and although a different detection method was used, their results are similar to JR2017. The zonal analysis reported in Chapter 5 and following from JR2017 serves very much as a tractable intermediate, but in fact indicate that climate regimes are not purely global phenomena. This was explored using global climate models.

Spatial analyses

Spatial methods such as EOF are often used as part of a regimes analysis but of themselves do not detect abrupt changes.

Two recent papers simply assume the presence of regime shifts and analyse gridded spatial data for the duration of local change processes (Yan et al., 2016, Yan et al., 2015). Following He et al. (2012), they assume that state changes take a measurable time and exploit a four parameter logistic curve for detection of magnitude and duration of a change with the production of gridded maps showing coherent structures. However apart from these it is difficult to find work which attempts to locate loci of changes spatially.

Using altimetric data of the South Pacific, and combining it with ocean salinity and temperature data a derived from ARGO floats, a spin-up of the Southern sub-tropical gyre between 1993 and 2004 is shown (Roemmich et al., 2007). This gives a 12cm increase in sea-surface height, and changes extending down past 1800m at the gyre centre. The paper suggests that this is linked to annular wind circulation and consequently all gyres and ocean circulation systems would show similar effects.

Chapter 3: The Multistep Bivariate Test

Introduction

The principal change-point detection method used in the rest of this thesis is based on the method first introduced as an homogeneity test by Maronna and Yohai (1978) (the paper henceforth will be referred to as MY78). The Maronna-Yohai bivariate test (henceforth MYBV) is a likelihood ratio test that uses a reference variable, and a single test variable. An adaptation of the test was published by Vives and Jones (2005) (henceforth VJ2005), and was designed for detection of step-like shifts in climate records. I have extended it to deal with multiple step-like changes, as suggested by Jones (2012). It will be referred to as the Multi-step Bivariate test (MSBV).

A preliminary version of the MSBV was published under the name "Probabilistic Bivariate" (PBV) (Ricketts, 2015a) , and formed the basis of analysis reported by Jones and Ricketts (2017b). Since the publications of the PBV and JR2017, the algorithm has been slightly simplified, an optional alternative rapid termination procedure suited to analysis of large data sets has been included, and all three originally described variants of the MYBV have been implemented in the MSBV.

The rest of this chapter is structured as follows.

The MSBV test is introduced, commencing with the relationship between the variates in the MYBV test and the three variants published in MY78.

The main features of the test are summarised, followed by a reiteration of the central equations of the MYBV test (see Box Ch3.1). The extension to multiple steps is then described, as are the algorithm and the decision rules incorporated. Various empirical tests of the method are described. This is followed by a brief comparison to two other multiple change-point methods which have been used in climate studies, although neither is designed specifically for multiple steps, but instead for combined steps and trend-changes.

Three case studies, taken from later in the thesis, and which demonstrate the use of different aspects of the MSBV. These are: (1) a comparison step-like shifts in the global mean annual temperature and annual mean of the northern mid-latitudes, to demonstrate the general method for analysis of single time series. (2) The method is demonstrated at finer scale with spatial analysis and illustrated with some additional detail for one year of step-like change. This also shows that the phenomena have inherent spatial coherence. (3) A variant method suggested in MY78, when two correlated variables each of which may have a change-point are to be analysed, is demonstrated using global mean ocean temperature data at two different depths. Discussion, and consideration of future directions follows.

The bivariate test

MY78 assumes both the test and reference variable to be independent and both to be independent and identically distributed (i.i.d.) series, save for no more than one step-like change in the test variable. The role of the reference variable in the MYBV test is to adequately represent the characteristics of the test variable under a null hypothesis of no discontinuity between the reference and test variables. The exact null is a domain specific issue. In applications of the test, the null hypothesis, and thus an appropriate reference series is a domain specific consideration. For most applications of the MSBV the null hypothesis is strict stationarity and the reference series is a flat random time-series.

MY78 identifies two related statistical models. When the reference series is correlated with the test series, this is known in MY78 as Model I, and is the principal focus of that paper. This is the most analysed case, e.g. Potter (1981). When it is uncorrelated, it is known as Model II, and has different critical values. MY78 also deals with the case of two correlated reference series, either of which may contain a single shift independent of the other series. MY78 stipulates that the change-point for the pair be the most likely change-point when either is used as a reference for the other. Thus three uses of reference series can be identified.

a. The *one way homogeneity* test based on MY78 Model I. The reference series represents the expected trajectory of a time series without a step change. For example a comparison of weather recorded at an observing station compared to averages of surrounding observations. Potter (1981), for example, derives reference series for rainfall collecting station data from composites of neighbouring collecting stations. In effect the composites are assumed to represent an estimate of an independent variable, and the individual station data are a dependent variable. The MYBV is one of a number of tests available (Buishand, 1982, Alexandersson and Moberg, 1997, Moberg and Alexandersson, 1997, Yozgatligil and Yazici, 2015). A correlated series from a separate independent variable may be used as the reference series. For example, as part of a regional attribution analysis of continental mid-latitude areas it had previously been established that "… annual average minimum temperature (Tmin) is correlated with maximum temperature (Tmax) … and Tmax is correlated with total annual rainfall". In these cases, Tmin is treated as the dependent variable and Tmax as the independent reference variable, whilst separately, Tmax was treated as dependent with total rainfall as an independent reference (Jones and Ricketts, 2017b Section 3.3).

b. The *two way homogeneity* test based on Model I with two correlated variables either of which may contain a change-point. This would apply when there is a potential co-dependency between two variables or if dependency is undetermined. This is also useful when the two series are responsive to the same driver(s).

c. The *step-change* test is the case on which this work has largely to date concentrated, Model II with a flat random reference where this is an adequate representation of the statistical null hypothesis of no change. The statistical null hypothesis, in turn represents the expected behaviour of a steady state regime. Development followed the reasoning of VJ2005 with the test of a time-series being reiterated multiple times against a resampled flat random reference series.

Main features of the Multi-step Bivariate test

The system extends the MYBV test for a single step change to multiple step changes.

The system is written in the Python language (v 2.7.1), with the core bivariate test module nested in a rules base framework.

The system consists of two components, (a) a rules based algorithmic framework which takes test and reference time series, and given an evaluation routine such as the MYBV test, produces a sequence of change-points; plus (b) a core of bivariate evaluation routines based on the MYBV test with options that support all three modes of operation.

For a given time series, and a threshold probability, the MSBV returns a stochastic-statistical model in the form of a list of break points, each of which meets a nominated threshold *p*-value for a step-change, each element comprising the time of change, the step size, the test statistic, and the *p*-value of the level change.

By contrast with other methods for multiple change-point detection the MSBV does not use an information metric for statistical model selection. Rather, provisional change-points are added between existing change-points on the basis of their *p*-value and all subsequent change-points are reassessed; and the entire analysis is reiterated until a stable consensus is reached. Also, when running with random reference series it uses multiple evaluations throughout, resampling the reference variable and selecting the most common (the modal) time of change.

Although work using this framework published to date has used flat random reference series and multiple iterations to detect steps as per VJ2005, provision has been made for the *two way* detection of shifts in paired time-series where neither is necessarily an independent reference, or *one way* detection of inhomogeneities given a presumptive errorless reference. The previously published versions of the test (the PBV) ran as a step detection test with flat random reference series only.

When searching a single time series with a random reference, multiple trials are performed to locate change-points. This means that we do not use a continuous probability distribution to compute a central tendency, but return the most common change-year found, the mode. This is an important feature. If the segment under examination has two frequently detected points, one will represent the most likely and the second the next most-likely but their average would be meaningless. The same strategy was adopted by Beaulieu et al. (2008) in inter-comparisons of homogenisation techniques for rainfall data.

Methods, governing equations and algorithm.

This section briefly describes the algorithmic framework, its genesis and current form. The equations initially presented in MY78, were published with minor typographic differences by Potter (1981). JR2017 and this work use the form of the bivariate test published in Bücher and Dessens (1991) and reiterated in later work (Jones, 2012, Ricketts, 2015a). For convenience the equations are repeated here (Box Ch3.1,below), and a derivation that links the form in Potter (1981) to that in Bücher and Dessens (1991) is provided in Appendix 3.1 (equations A8-A15). Flowcharts are provided for the main parts of the MSBV.

Extending to multiple steps.

The MSBV can be run in the three modes of use defined above, and extends the MYBV test in two main ways. It (a) extends the test to detect multiple shift points via an algorithmic framework, and (b) utilises a resampling strategy to explore alternatives when a random reference series is used. It will be shown that this greatly increases the precision of the test.

Extending any test for a single change in a time series into a test for multiple changes involves multiple iterations, each of which can either establish or modify the time of change-points, the computation of a revised model and the evaluation of a halting criterion to terminate the process. The MSBV test like others, for example structural change methods, examines change-points in the times bounded by the immediate prior and posterior changes. But rather than an exhaustive search, it revises sets in a causal order by traversing them from earlier to later times. Limiting each change-point evaluation to data bounded by provisional, earlier

Assuming that (x_i, y_i) are i.i.d. random vectors of length *n*, let x_i be a stationary reference time series and y_i be a test time-series which is assumed to correlate to x_i except for a single shift at some time i_0.

Step 1. Standardize series.

$$\overline{X'} = \frac{\sum_{j=1}^{n} x'_j}{n}, \overline{Y'} = \frac{\sum_{j=1}^{n} y'_j}{n}, S'_x = \left(\frac{\sum_{j=1}^{n}\left(x'_j - \overline{X'}\right)^2}{n}\right)^{1/2}, S'_y = \left(\frac{\sum_{j=1}^{n}\left(y'_j - \overline{Y'}\right)^2}{n}\right)^{1/2}, \tag{1}$$

$$x_j = \frac{\left(x'_j - \overline{X'}\right)}{S'_x}, y_j = \frac{\left(y'_j - \overline{Y'}\right)}{S'_y} \text{ for all } j \leq n. \tag{2}$$

Step 2. Compute test statistics.

$$S_{xy} = \sum_{j=1}^{n} x_j y_j \tag{3}$$

$$X_i = \frac{\sum_{j=1}^{i} x_j}{i}, \; Y_i = \frac{\sum_{j=1}^{i} y_j}{i} \text{ for all } i < n \tag{4}$$

$$F_i = n - \frac{{X_i}^2 ni}{(n-i)} \text{ for all } i < n \tag{5}$$

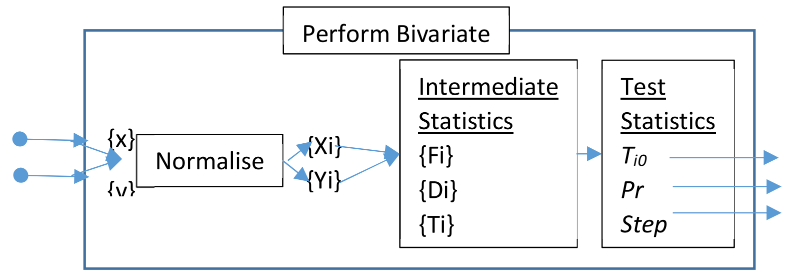
$$D_i = \frac{\left(S_{xy} X_i - nY_i\right)n}{(n-i)F_i} \text{ for all } i < n \tag{6}$$

$$T_i = \frac{\left[i(n-i)D_i^2 F_i\right]}{\left(n^2 - S_{xy}^2\right)}, \text{for all } i < n \text{ and } T_{i0} = \max(T_i) \tag{7}$$

Let i_0^* be the value of i for which $T_i = T_{i0}$, the time after which a change occurred. Its successor is the first time of the new regime. D_i^* is defined as the maximum likelihood estimator of a shift at i_0^*. T_{i0}^* is the test statistic that tested against some constant, discriminates with a specified probability, a null hypothesis H_0 of no shift against H_1 that a shift exists (Maronna and Yohai, 1978). A mean shift can be computed as $\Delta\bar{y} = D_{i0}^*.\acute{S}_y$. For the null trend case, analyzed in Maronna and Yohai (1978), critical values of T_i are given for probabilities of (0.25, 0.1, 0.05, and 0.01) for the null hypothesis of no change, given time series lengths *n* of 10, 15, 20, 30, 40 and 70. Potter (1981) provides these for *n*=100.

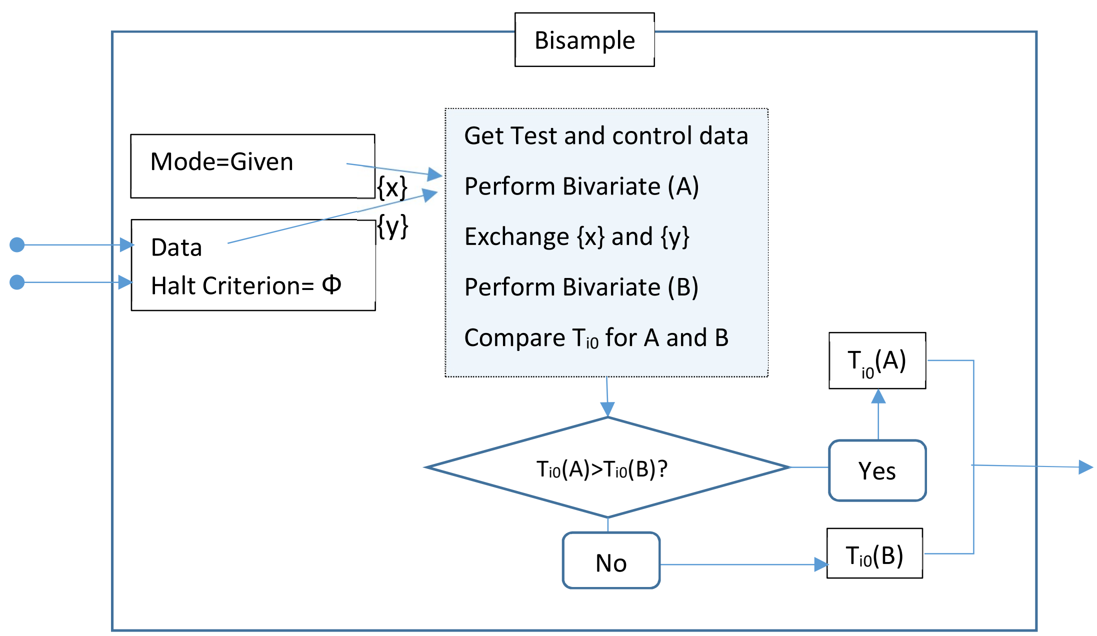
Box Ch3.1 : The Maronna-Yohai test equations, from (Ricketts, 2015a) and following MY78. See Appendix 3.1 for an extended derivation from the forms published in MY78 which did not include a standardisation step.

established change-points, imposes censorship on the decision process since a change-point determination does not simultaneously consider the statistics of other, presumably related, segments.



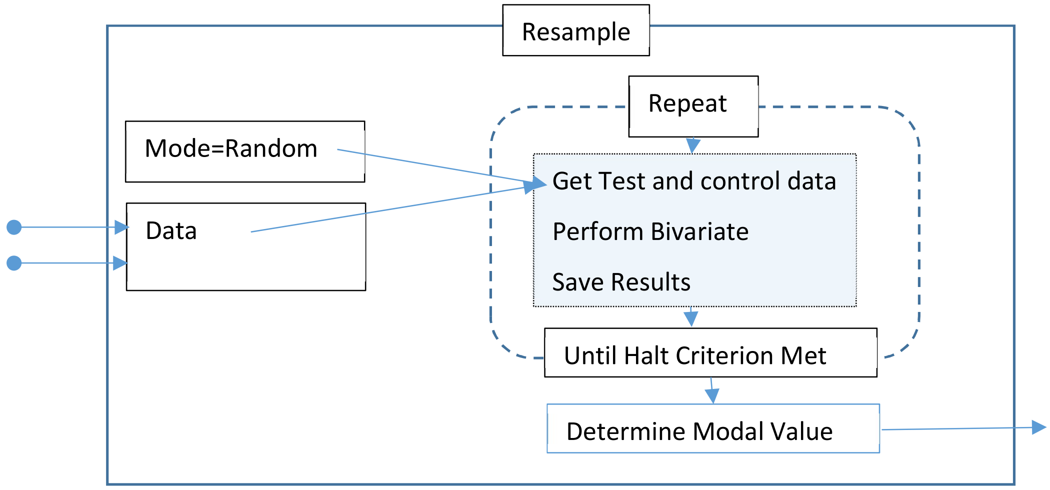
Flowchart Ch3.1: Bivariate test. A step of the size returned is accepted in a segment of length N, if Prob T_{i0} < t(crit, N).

When running with a deterministic reference as an homogeneity test, the bivariate test is run once for each determination of a potential change-point. When operating as a two way homogeneity test the bivariate test (Flowchart 3.1) is run twice on each segment, interchanging the test and control variable, with the returned time of change being the overall most likely time of change (I call this "bisampling"), (see Flowchart 3.2).



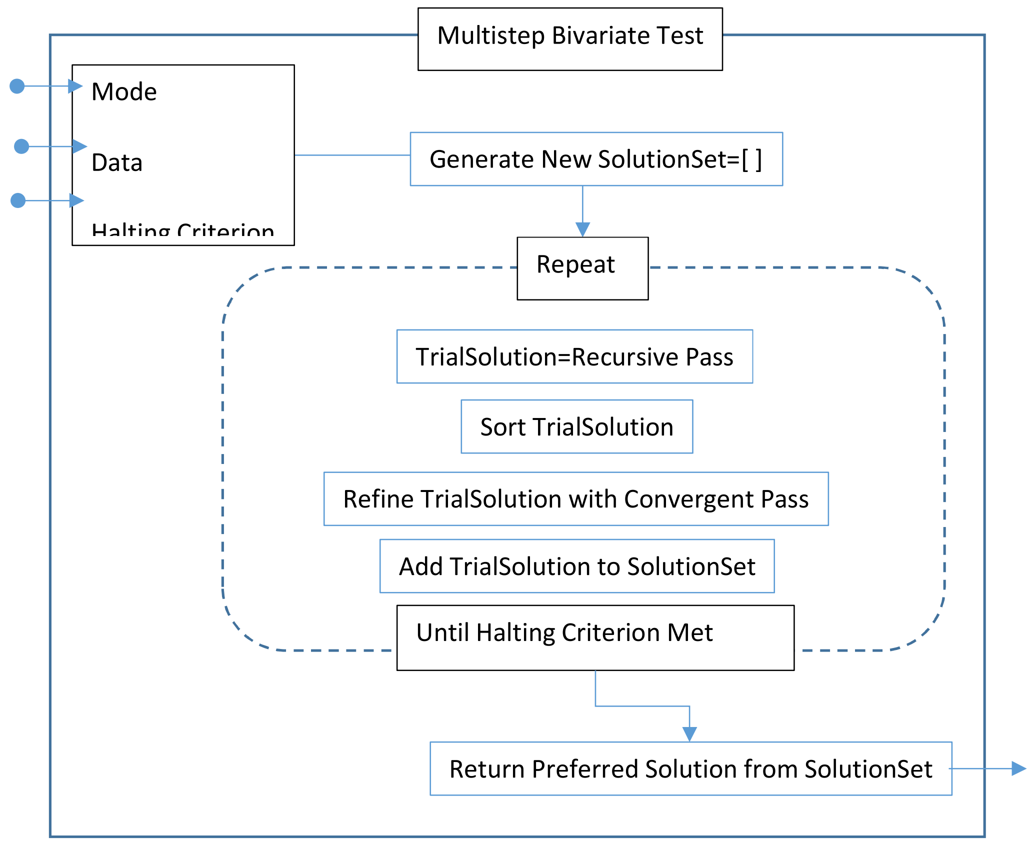
Flowchart Ch3.2: Bisample. Does not return modal values, it simply performs the bivariate test twice, interchanging variables, and returns the time of the greatest departure.

When operating as a step-change test as described in VJ2005, the system is less deterministic and takes on some aspects stochastic resonance (Benzi et al., 1982) where injected noise improves accuracy of feature detection (see Flowchart 3.3). In this test, noise is injected into the reference series, not the test data.

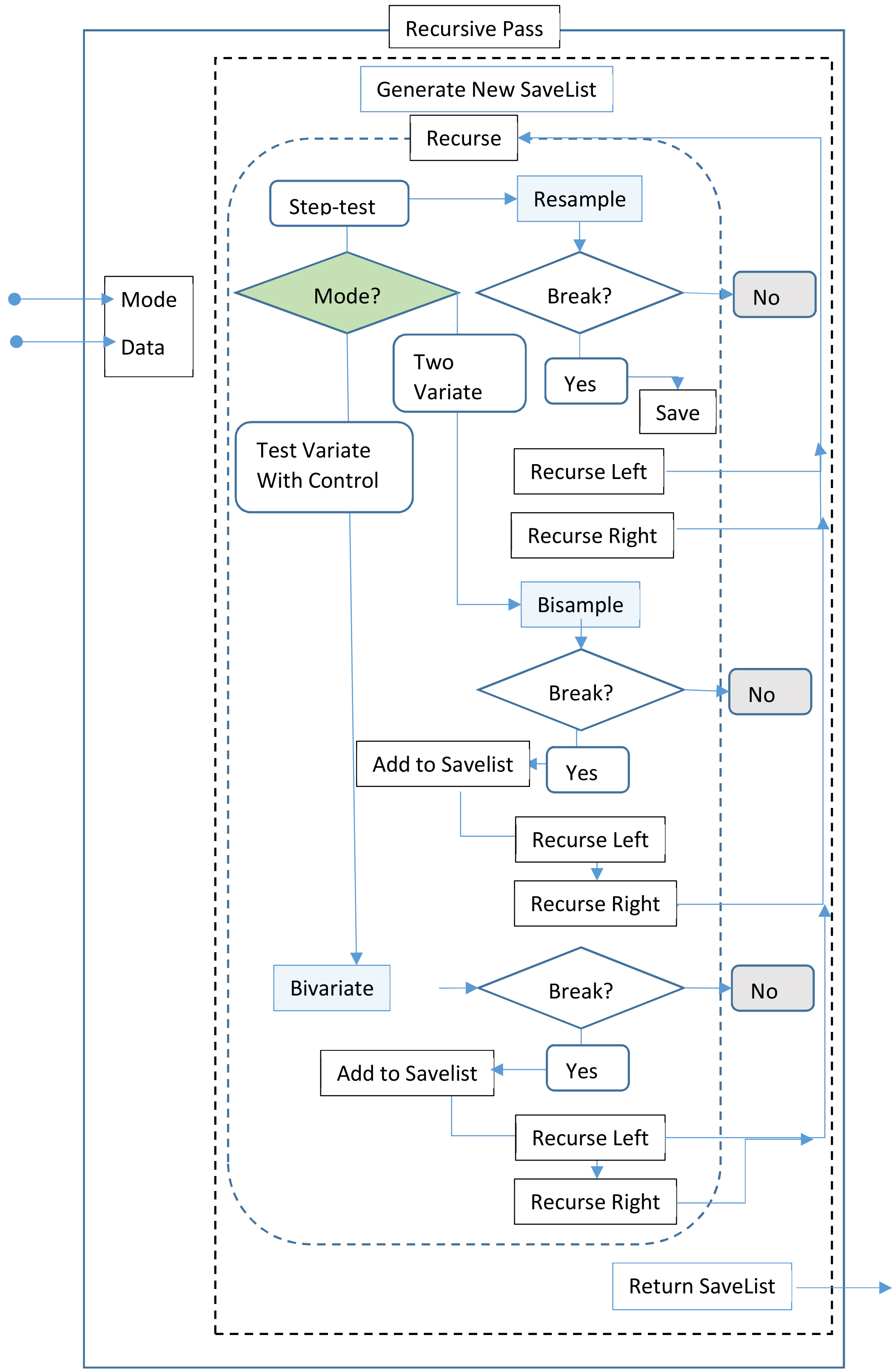


Flowchart Ch3.3: Resample procedure. The control data is randomised on every iteration. This is the default behaviour described in VJ2005 and J2012.

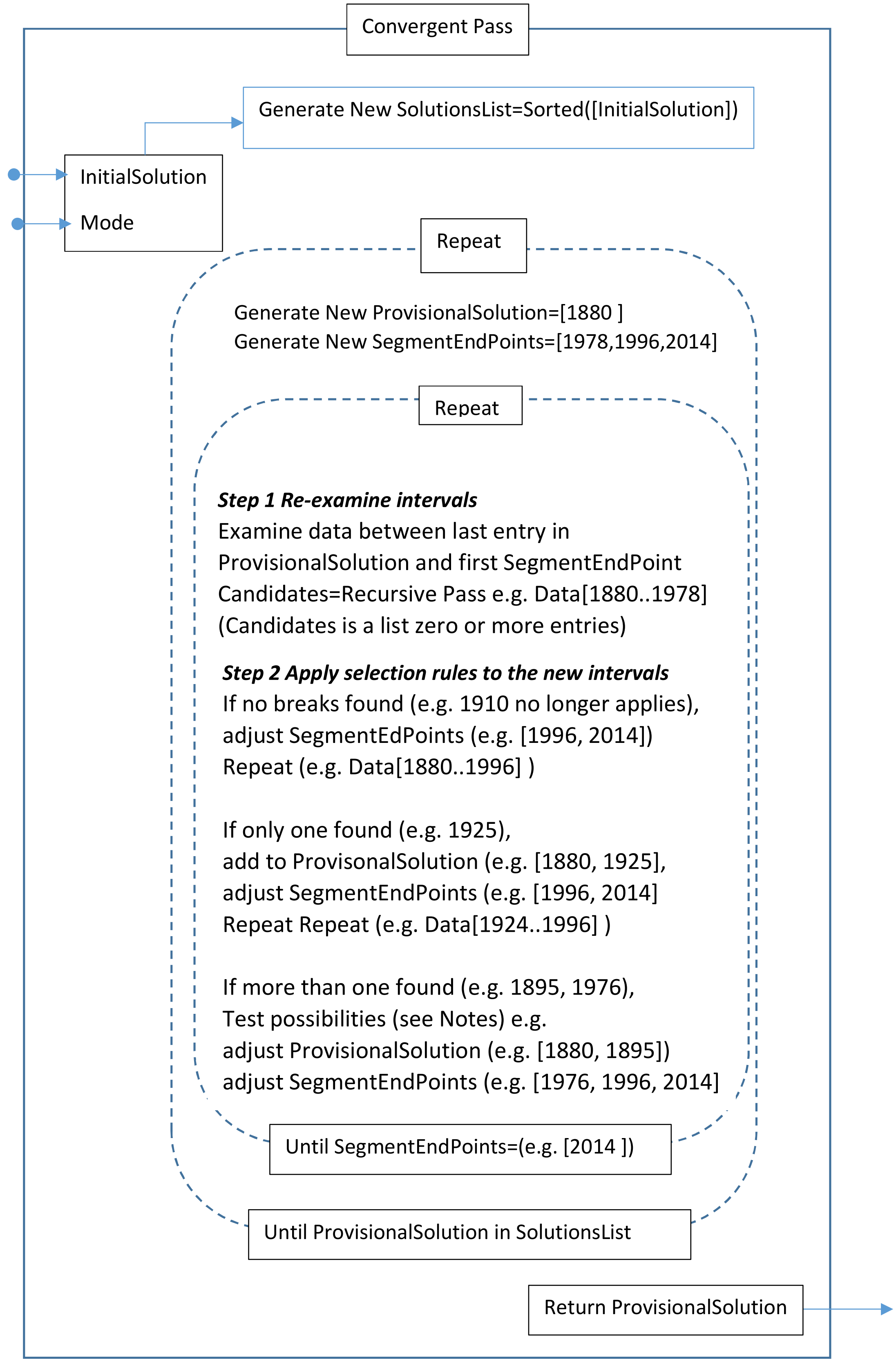
A single time series analysis by MSBV consists of two passes; a *screening pass* which produces an approximate list based on a recursive analysis, and a *convergent pass* (Flowchart 3.6) which refines the list by adjusting breaks in time order until a stable solution is attained. Together these form a single iteration (see Flowchart 3.4). The recursive procedure (Flowchart 3.5) is used in both passes, and is similar to *binary-segmentation* (Killick, 2012).



Flowchart Ch3.4: A single iteration of the Multistep Bivariate Test



Flowchart Ch3.5: Recursive procedure returns an unsorted list of possible breakpoints. The green diamond labelled mode selects three alternative modes of use hence three variations on the bivariate test. For step-detection resample mode is used (Flowchart 3.3), for two variates, bisampling (Flowchart 3.2), homogeneity testing – a single bivariate test.



Flowchart Ch3.6: Convergent pass, annotated with example dates.

Algorithms

Outer loop. A single iteration of the test consists of a screening pass then convergent pass (see below). It returns a list of break-years that segment a time series such that each shift is statistically likely at a selectable level (default is p=0.01) based on $T_i^{critical}$. Whilst most often, all iterations return identical lists, analysis of some time series will several variations – mostly as individual shift dates move by a year or so. Within this loop, all breaks are determined by applying a *resampling test* (below).

Halting criterion. Normally the outer loop is iterated 100 times, and from this a small set of step-series is returned (most commonly just one), and the modal set is retained.

Optionally, if speed is required, I attempt to choose

(a) If the first seven iterations yield identical sets then that is returned since this has a *p*-value < 0.01. If not, then the test continues until

(b) More than 9 iterations and either (i) exactly two distinct sets have been found and the *p*-value returned by a Yates Chi-squared test <0.05, or (ii) the likelihood ratio of the modal set to the next most frequent set >20:1.

(c) More than 50 iterations have occurred and a distinct modal value is apparent.

This will be referred to as the Rapid Assessment Stopping criterion, and the MSBV run using this may be called the Rapid MSVB (RMSBV).

Resampling test. This operates on provisional data segments. For the PBV, the bivariate test is repeated 100 times using different random sequences and the i_0 values and means of the associated T_{i0}, and shifts are collated. For the RMSBV, when speed is needed, the same metric is used as for the halting criteria. On the *screening pass* only modal values are examined. On the *convergent pass* the modal and the second modal values (if present) are returned. The first modal value (i.e. most frequent) is returned as i_0 for those runs, the second is logged. The mean T_i of modal i_0 values, denoted $\bar{\bar{T}}$ and the mean shift for those values associated with i_0 is also returned. A segment contains a breakpoint in position i_0 if $\bar{\bar{T}} \geq T_i^{critical}$.

Binary segmentation. This is a recursive segmentation technique, (Killick, 2012). The entire time series is analysed for a single break-point using the resampling test. If $\bar{\bar{T}} \geq T_i^{critical}$, then the segment up to and including i_0 is analysed for an earlier break, and the segment after i_0 is analysed for a later break. This process is repeated for the sub-segments so formed until no breaks are found with p-values below the selected threshold. The result is a series of break-points which are then refined on the convergent pass. It should be noted that as break-points found on this pass are returned on the basis of a recursive process, end point effects may perturb the results, including probabilities.

Screening pass. This pass produces a provisional list of break-points using binary segmentation, which serves as a starting point for the convergent pass (below).

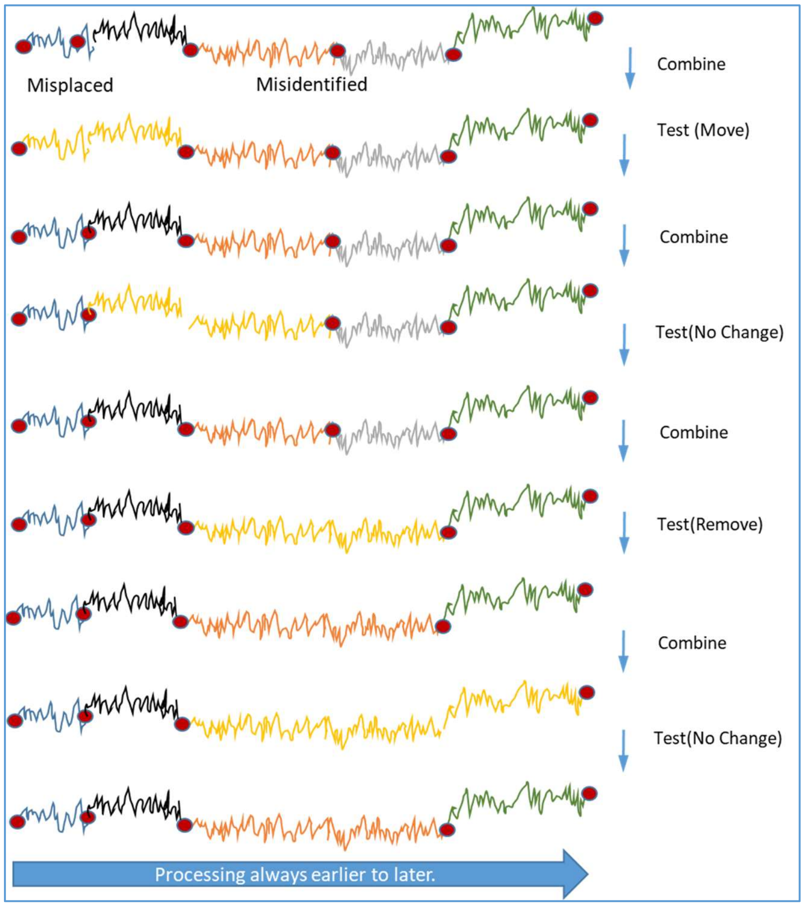


Figure Ch3.3: Schematic of one iteration of the convergent pass. Each shift-point (red circle) is revised by combining the sub-segments each side and testing the result to locate all shift-points within. For the rest of this example assume that no more than one are found. The earliest such point becomes the revised location within the combined sub-segments. The newly revised point becomes the left hand bound for the next iteration. In this case an initially misplaced point is modified and a false positive is eliminated. The convergent pass continues until a solution is discovered for the second time, and that is returned as the result.

Convergent pass. The list of *n* break-points breaks the original time series into $s=n+1$ segments. The algorithm then works its way from earliest to latest segments combining consecutive segments into one, and then searching within that segment using binary segmentation to produce new estimated break-points (usually but not necessarily no more than one) (see Figure Ch3.3). These estimates are processed by the decision rules below. There are two special cases, segments 1 and *s* which are analysed individually at either end of this process to cover the impact of end point adjustments. This pass iterates until a break-point list is found for the second time, and this is returned as the result.

A small set of decision rules are used during processing (see Box Ch3.2). These depend in part on the use of the recursive routine during the convergent pass. The rule set has been slightly refined since the PBV was published. The application of the rules is illustrated with a worked example (Box Ch3.3).

Decision rules.

Decision Rules

1. **Prohibition:** A tunable prohibition period defaulting to seven years is applied after a break-point before another point will be accepted, to minimize false positives associated with sub-decadal variation or transience. Break-points are disallowed within the same prohibition period at the end of the data.
2. **Mode:** If resampling of random reference variables is used, the change-point time (T_{i0}) returned is the most common – the modal value – of those detected over all iterations, subject to the next rule.
3. **Well defined mode:** If resampling is used and the modal value returned by resampling is >= 90% of those detected or the modal value is >50% and the second modal value > 20% then the modal year is accepted, else it is dropped on the basis that it is possibly artefact.
4. **Prefer later findings:** If a segment contains a single break-point that break-point replaces any previous one.
5. **Mergers:** If a segment no longer contains a break-point then the segment and the next are merged and treated as a single segment on the next iteration.
6. **Multiple possible change-points:** Within the convergent pass when a segment if a segment contains more than one point the earliest two are retained and the rest discarded. The two points are then trialed using a resampling test to determine if the interval up to the later of the two still contains the earlier break, and if this is still present, it is retained, otherwise the other is retained.

Box Ch3.2: Summary of rules applied during the convergent pass processing of the MSBV.

Tunable prohibition. There are three principal reasons for imposing a prohibition period. The first is that the dominant sub-decadal variability mode is due to ENSO, generally assigned a mean periodicity of 3 to 4 years, but varying up to 7 years. The second is that the principal sub functions of the test involve summing of either exclusively random numbers (X_i) or partially random numbers (Y_i), and any resulting random steps will be unlikely to persist after seven samples. The intent of the MSBV is to detect persistent step changes – but to ignore transients lasting for a small number of years. Consequences of this rule are seen towards the later end of a data set, and this may be confusing when earlier and later collections are compared. For example an earlier collection of zonal ocean temperatures 60S-30S with complete data only up to 2014, had a change-point at 1996. The same collection with complete data to 2016, drops the 1996 date and shows a shift in 2008. This is a consequence of the seven year prohibition with 2008 becoming a permissible change point so that the 1996 change-point, previously considered probable between 1979 and 2014, is not as probable within the interval 1979 to 2008. This could be ameliorated by reducing the prohibition for the last segment. However this was not performed in this thesis.

1. **Use of mode rather than mean.** This is a critical point. The data may contain two separated years which are very nearly equally likely. During resampling as per VJ2005, both of these may occur, but the mean of the two would then represent a year which never occurs, and in fact may maximally unlikely. If the modal year is chosen subsequent processing of the resulting sub-segment may uncover the other year, or it may prove insignificant.
2. **Well defined mode**. This eliminates stochastic drift. If the modes are neighbours then they represent a single shift, if they are well separated then the less strong one may be examined on the next iteration (see previous rule).

The following rules deal with the consequences of the convergent pass where on examination of a data segment using the recursive search, zero, one or more potential breaks may be found.

3. **Retention of a later determined single break.** This simply a consequence of the segment bounds being potentially revised.
4. **Merge of an empty segment with the next.** This will occur primarily due to a change early bound. The newly merged segment is tested again immediately. This may itself return the change-year that had just been dropped, in which case the next sub-segment to be examined will have that as the lower bound and processing continues normally.
5. **Multiple possible change-points.** In practice this occurs within long data sets responding to strong forcing, when the screening pass may miss a number of possible years. It is a consequence of the recursion finding break-years in order of T_i value. The rationale for this rule is to trial both of these possible shifts as possible segment end points, since they will be treated that way on a subsequent pass.

1. In this example, the *screening* pass has produced 3 points in the range from 1850 to 2014. This gives a list of years and the order in which they were found as follows, [1850,1900(2),1970(3),1987(1),2014]. This came about because, first 1987 was detected between 1850 and 2014; then 1900 was discovered between 1850 and 1987; then 1970 was discovered between 1900 and 1987. However until the interval from 1850 to 1970 is re-examined, 1900 may not be the year that best fits. Hence the list is revised on the convergent pass.
2. The *convergent* commences with the provisional list found above. It first takes the interval 1850 to 1970 and attempts to find *all* possible shift points within. Let us suppose it adjusts 1900 to 1902, and finds nothing else.
3. The provisional list is then [1850, 1902].
4. The next iteration tests for *all possible years* in the interval 1902 to 1987, another interval which has not yet been tested, but within which 1970 is assumed. Let us suppose that this time the test returns *two* years, [1919 and 1970]. As it happens 1919 was first found within the interval 1902-1987, and then 1970 was found within 1919-1987. Rule 6 is applied to test that if 1970 is real, 1919 is still found in the interval 1902 to 1970. Let us suppose it is not. The provisional list is then [1850, 1902, 1970].
5. The last iteration on this pass tests the interval 1970 to 2014, again an interval not yet tested. Let us suppose that 1987 is all that is found. At this stage the break list is [1850, 1902, 1970, 1987, 2014].
6. Another convergent pass is required because a year has been amended.

Box Ch3.3: Worked example in which rule 6 is used. See also Figure Ch3.3.

Assessments of the MYBV

The NCDC zonal data version v3.5.4.201504, see [APPENDIX-DATA].

Although the MYBV test has been amply reviewed and assessed it was considered useful to survey its behaviour when shifts are small and/or when the underlying physical model and its representative statistical model does not completely conform to the assumptions of the detection test.

Appendix 3.1, equations A16 to A18 define the limit behaviour of the MYBV test with multiply resampled random reference variables. The reason for the precision of the MYBV test can be understood as the product of two elementary functions, the limit of a cumulating function of the residuals, in the terminology of MY78, Y_i^2, for all $i < n$, and the limit of a hyperbolic function that weights for distance, $\frac{i(n-1)}{(n-i)}$, for all $i < n$. The product is a spire shaped function with a discontinuity in the derivative at the break-point. Hence the test statistic is maximal both in its value and in its derivatives at the point of change, and from this follows a very high degree of precision in the presence of noise.

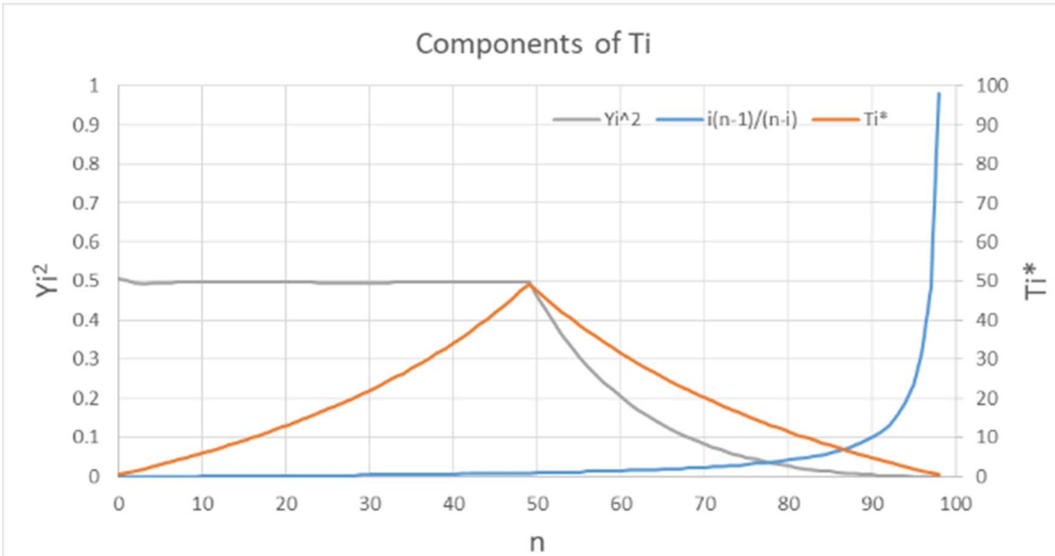


Figure Ch3.4: Illustrative relation between running mean normalised observations (Y_i^2), the distance weighting (blue), and their product, the T_i* function.

Figure Ch3.4 illustrates this with the example of a sample of n=100 points, a standard deviation of σ, and a shift of 2σ at the mid-point. Figure Ch3.5, bottom pane illustrates that even when the step change is a sixth of the total change the location of the step-change is still precisely measured, however by contrast in a more extreme case (step of 1/16th total change) deceives the test.

Further consideration of equation A18 shows that if the data contains only a uniform trend then the T_i function will be an inverse parabola and data with a trend change or combined step and trend will show some combination of hyperbolic and parabolic curves. Data without a step change but with a small trend change is potentially deceptive. See the top pane of Figure Ch3.5, where the T_i function is plotted for four data sets which have the same total change end to end, but follow four different trajectories including the above cases. All the T_{i0} values would be considered significant. However empirical testing shows that the timing (T_{i0} values) returned by the MYBV is resistant to moderate trends, and J2012 and RJ2017 give two methods that were applied to climate data to establish that the data was not deceptive for the MYBV – the shuffle test, and the window test.

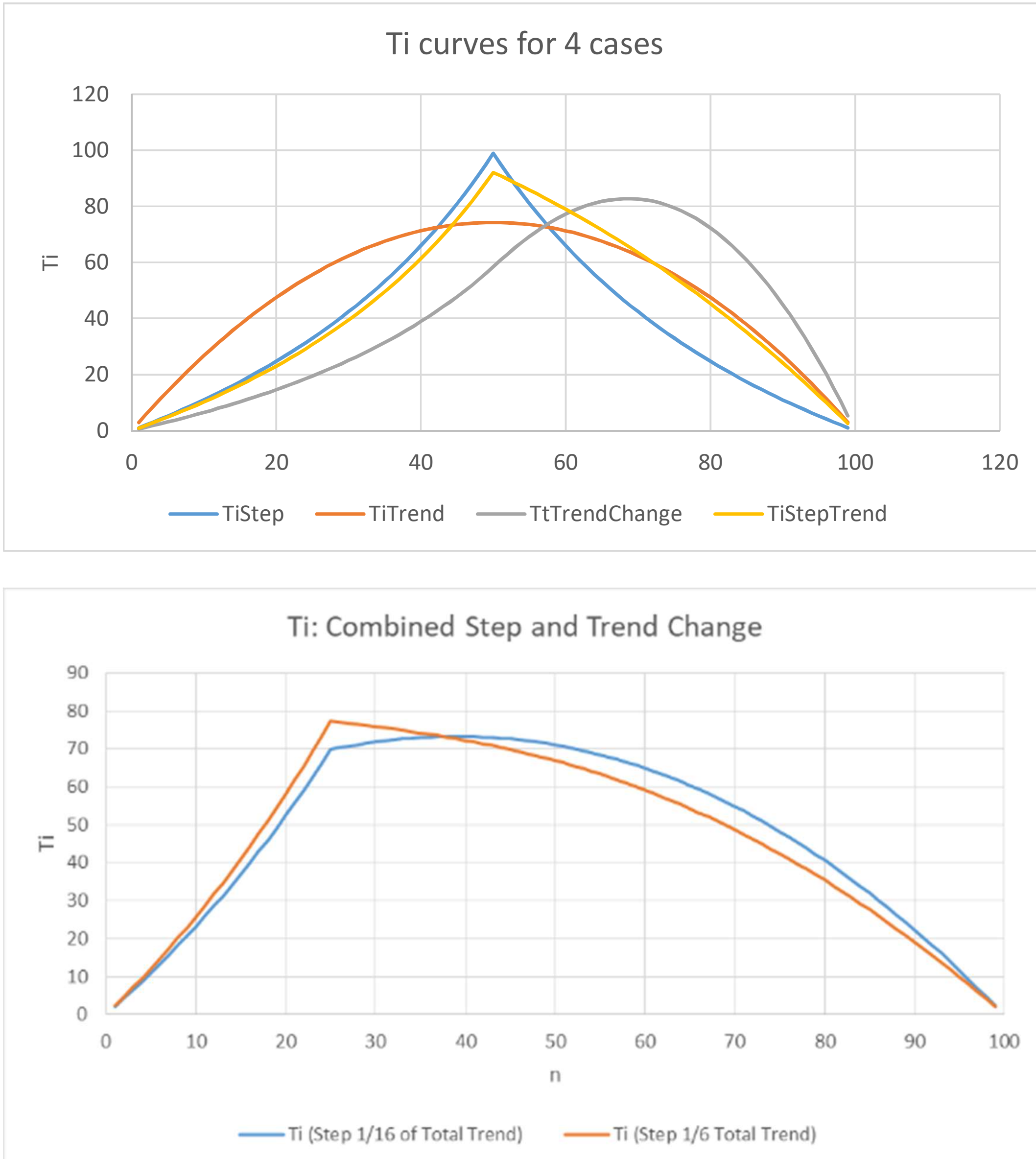


Figure Ch3.5, Top pane: Ti functions computed for illustrative cases. Blue: A step change in an otherwise flat time-series. Orange: a uniform trend. Grey: a trend change without a step. Yellow: a step and a trend change together. Bottom pane: Illustrative Ti functions given combine step and trend change. Time change as defined as time of maximum corresponds time of inflection until the variation due to trend strongly dominates. The step change in both cases is at index 25.

This last possibility does not seem affect current observations since the trends and the trend changes have not been severe to date. As an illustration, a diagnostic plot from testing of the MSBV is shown (Figure Ch3.6). Data is the GISTEMP3 Northern mid-latitude mean zonal land and ocean temperature data supplied by NCDC (30-60°N), taken from analysis published by Jones and Ricketts (2017b). It compares the supplied data to a hypothetical step and trend model with the same variance, shift and trends as deduced from the data. Variability of the T_i function is illustrated by the blue and red blurs. Each point in the blurs is the T_i value at that year for one of 1000 iterations. The range of variability of the T_i function of the supplied data is shown in blue, simulation of randomised data with the same trends either side of the same internal shift in red. Solid lines indicate the mean of the blurs, blue for the supplied data, red for the hypothetical case and also in green the mean response for the pure linear trend computed for all data. This diagnostic matches the post 1996 land and ocean shift shown in the bottom panel of Figure Ch3.11 further down. Of particular note, if there was no change-point in the data the T_i function would follow the green curve and maximise circa 2003. Clearly the selection of max(T_i) is not perturbed by trend and 1996 is a close estimate of the time of change, although the year after may have been a candidate.

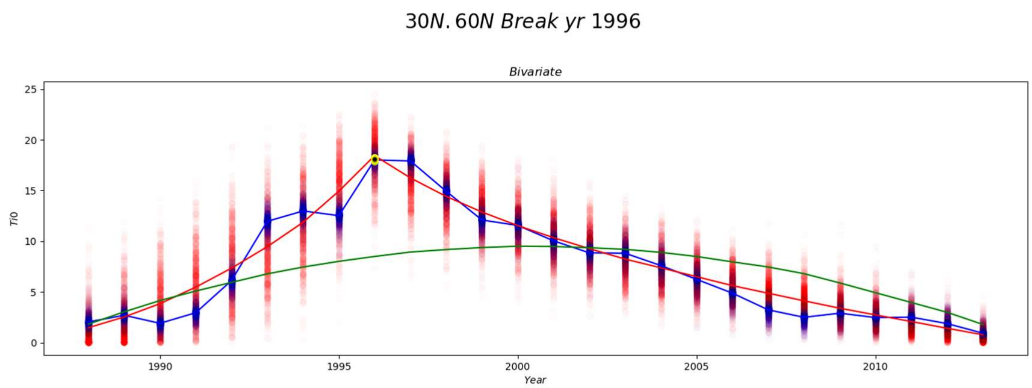


Figure Ch3.6: The T_i function for sample data, showing both the T_i function computed for sample data and the effect of the a random control variate. The test was repeated three different ways. The blue dots represent the variation of the T_i function for 100 iterations, the blue line is the mean of them. Black dots with yellow circles are the T_{i0} function selected by the MSBV for the first 100 iterations. In this case, the same change-time throughout. The red circles the T_i function 1000 Monte-Carlo iteration, and the red line is the mean. For this internal trends and shifts either side of the change-point were subtracted from the initial data to yield a residual, normally distributed noise with the same variance as the residual was added back to the same internal trend and shifts, and the T_i function was plotted. Lastly a single trend was computed from the whole of the initial data, the same normally distributed noise was added in and another T_i function computed. This is shown in green.

Testing of the method

This section proceeds by first testing the sensitivity of MYBV test itself to the location and size of a shift, and size of the data segment. It then tests the performance of the MSBV against synthetic climate data.

Sensitivity testing uncertainties of timing.

Confidence limits on timing: In the literature on homogeneity testing for which the MYBV test was proposed, the emphasis is mostly on the detection of an inhomogeneity in a data segment, with associated *p*-values determined by the size of the inhomogeneity. The best estimate of the time of change is also required, but the theoretical uncertainty bounds on the timing are generally not discussed. Rather, for example, bounds of ±2 years are stipulated as “well-located” (Beaulieu et al., 2009). The MYBV test has been analysed several times either singly or in reviews of similar methods (Beaulieu et al., 2009, Beaulieu et al., 2008, Ducré-Robitaille et al., 2003, Tayanç et al., 1998, Alexandersson and Moberg, 1997, Easterling and Peterson, 1995, Bücher and Dessens, 1991, Buishand, 1984, Potter, 1981, Killick et al., 2010). Computation of confidence intervals on change-points is complex and Hušková and Kirch (2008) shows that bootstrapped confidence intervals fall within the limits of asymptotic intervals as given by for example Bai (1997 see equation 17) in discussions of general change-point models.

Monte-Carlo simulation was used to survey the impact of segment length and step size on T_{i0} (timing) confidence intervals. Normally distributed data with standard deviation of σ was used to generate segments of length 20 to 100 in increments of 10. Step-changes of 0.5σ to 3.0σ in increments of 0.5σ were added at the mid-point.

To assess only the precision of the estimate of location of the change-point, the MYBV was run 10,000 times with new simulated data each time, for all defined segment lengths and step sizes. The construction of this test allows a sample standard deviation on the timing to be extracted.

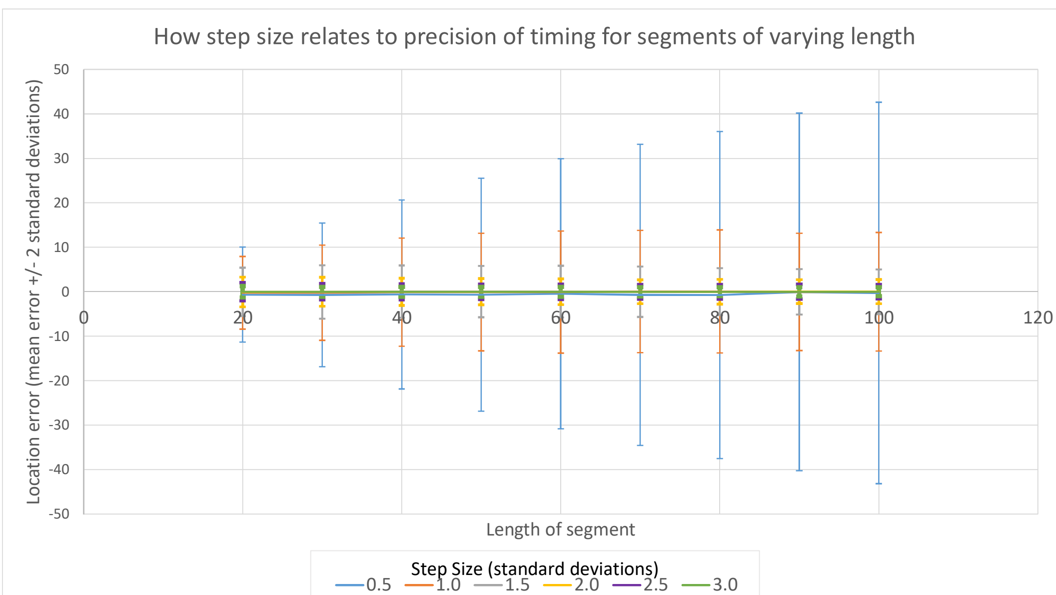


Figure Ch3.7: Confidence intervals for the detection of a step-change at the mid-point of data of various lengths. Once the step-size exceeds 2σ the 95% bounds shown as bars are less than two years.

As is illustrated in Figure Ch3.7, confidence bounds are inversely related to step size, and at the design threshold of the MSBV (reliable detection of 2σ steps), the 5% confidence limits are close to ±2 years. This does not include the effect of position or trend on the precision, and if there is a central bias, the error bounds may be optimistic.

Therefore, separately, I assessed the effect on precision of change-point location by Monte Carlo testing of 10,000 runs with shifts at 20% and 80% of the length. I then assessed the smallest step-size that needs to be present for the precision to be ±2 years whilst independent of its position in the data. There is still a dependency on segment length and for most segment lengths the necessary shift size is ≥2.5σ. However if MYBV is repeated as it would be when resampling in the MSBV the step-size required to give 100% accuracy is much smaller, ≥1.5σ for segments of length 30 or less, and 1σ for longer ones (see Table Ch3.2).

Table Ch3.2: Non-central shift points. Top row, the size of a shift that can be detected by a single run of the MYBV with 95th percentile spanning ±2 years, independently of the time of shift. Bottom row, the corresponding size of shift if the median value of a series of runs is used.

N	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
Shift size in σ that yields precision of ±2 years at least 95% of the time	3	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Shift size in σ that yields the correct value as a median of multiple repetitions	1.5	1.5	1	1	1	1	1	1	1

Testing for central bias in the absence of trend: The T_i function, by its construction is non-linear, greater in the middle of the data (see Figure Ch3.5). The published critical values are independent of T_{i0} and therefore there is potential for a bias to the middle to occur. This would express itself as either a tendency for T_{i0} values to bias to the middle. This effect has already been seen in the prior section. Hence it is important to know the sensitivity to shift-size of any such bias in time of change. I tested the MYBV by Monte Carlo, using random data (standard deviation denoted σ) of lengths 10 to 90 at intervals of 10. Each data segment had a single step of size σ times a random value between 0 and 5, at a location randomly selected between 0.2 and 0.8 of the length. I recorded the medians of deviation from expected location, the median T_{i0} and *p*-values. Shifts with *p*-values less than 0.01 centred on their expected location and with satisfactory accuracy. The method shows a bias to the mid-point when shifts have p-values > 0.05. Steps of 2σ or greater were accurately located with *p*-values <= 0.01, independent of segment length or location of shift (see Figure Ch3.8 for an example of a change-point at the centre in a time-series of length 100). In general steps must be 2σ or greater in order to be accurately located by the MYBV in data that conforms to the assumptions of the test (i.i.d., no trend).

The effect of trend and the effect of the resampling strategy as used the MSBV were tested.

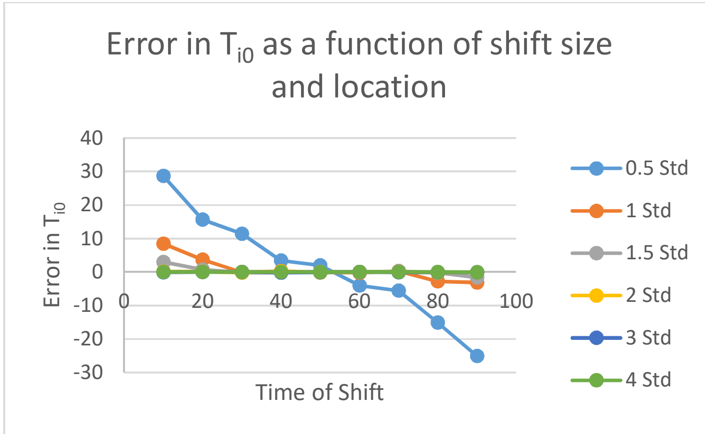


Figure Ch3.8: Illustrative examples of how the bivariate test behaves with shifts of different sizes. This plots the defined time of change on the X-axis against the mean error of 10000 determinations on the Y-axis. The 0.5 standard deviation shifts did not reach statistical significance at P=0.01 and the 1.0 standard deviation shifts at index 10 and 90 were significant at P=0.05 but not at P=0.01. Shifts of 2 standard deviation are detected without a displacement error.

Testing for the effect of trend on bias and precision: The MYBV test assumes zero trend, as indeed do other tests (Rodionov, 2005, DeGaetano, 2006), and the form of the T_i function is sensitive to trend (Figure Ch3.5). Therefore it is important to assess the sensitivity to trend of the MYBV and the MSBV.

Using segments of length 100, I tested the T_i and T_{i0} statistics returned by MYBV with uniformly trended data of that rose 4σ over the segment, to which had been systematically added a step of between 0.5σ and 4σ in 0.5 σ increments at the mid-point. I also tested data with the same steps and initial trends of zero, but where the trend changed at the same time as the shift. Imposition of uniform trend strongly modifies the T_i values leading to under-estimation of *p*-values and inflated significance.

When the step is small there is a bias towards the centre of the data in the estimation of T_{i0}, increasing with trend. As above for small trends and steps greater than 2σ the bias is negligible. However there is a relation between the degree of trend and the smallest shift that can be located without bias. If the increase due to trend over a period is *T,* the size of the shift required for zero bias *T* increases by nearly *T*σ/2. The major effect of a trend change is that the errors are skewed towards the side of the greater trend, whereas for uniform trend they are symmetric and centrally biased (See Figure Ch3.9).

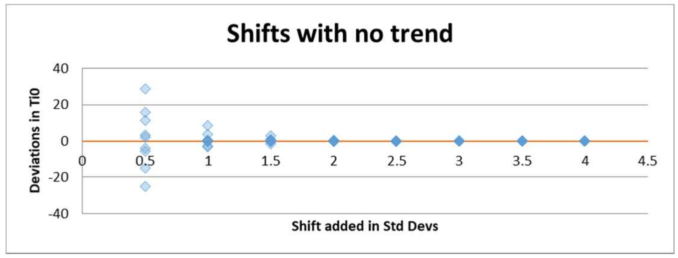
Secondly, to assess the sensitivity of bias and precision of the MYBV impacted by trend, another Monte Carlo test was performed. For each set of parameters 10,000 iterations of all combinations of the parameters were tested by MYBV. Parameters were segment length, location, shift and trend. Data segments of lengths 30, 60 and 90 were generated. Locations of 0.2, 0.5 and 0.8 of the length were selected for a step-shift. Step sizes of 0σ to 5σ in increments of 0.5σ were added. Then uniform trends of 0 to 8σ in increments of one were applied. The error of the T_{i0} values were collated for each combination of parameters and the 2.5th percentile, median and 95th percentile were tabulated. Again the median indicates the performance of the MSBV and the percentiles provide the spread for the MYBV against random references.

Table Ch3.3: Sensitivity of the relationship between imposed trend and minimum shift of a shift than can be accurately located by median of multiple trials e.g. MSBV (first set) or the MYBV run singly.

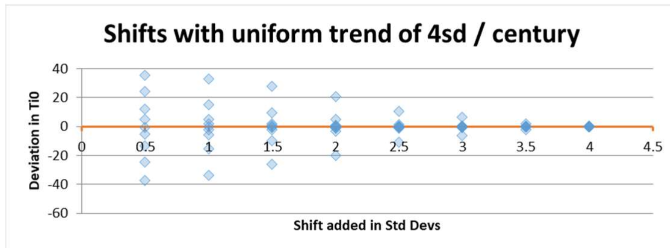
	Size of trend (σ/Century) that can be located within ±2 years by taking a median for increasingly large steps.			**Size of trend (σ/Century) that can be located within a symmetric 95% confidence limit of ±2 years for increasingly large steps.**		
Shift (σ)	30 Years	60 Years	90 Years	30 Years	60 Years	90 Years
0.5						
1	5	2	1			
1.5	8	5	3			
2	≥8	8	5			
2.5	≥8	≥8	7	0	0	1
3	≥8	≥8	8	8	6	4
3.5	≥8	≥8	≥8	≥8	7	5
4	≥8	≥8	≥8	≥8	8	7
4.5	≥8	≥8	≥8	≥8	≥8	8
5	≥8	≥8	≥8	≥8	≥8	≥8

Table Ch3.3 summarises the pertinent points. The main aim is to identify shift sizes which are always accurately detectable regardless of various imposed trends (“resistable trends”). For each combination of length, location, and shift the trend nominated as “resistable” was the greatest one for which the median error was ≤2 years, with all smaller trends also the same (these are the first group in the Table. Similarly the 95% span, shown as the second group.

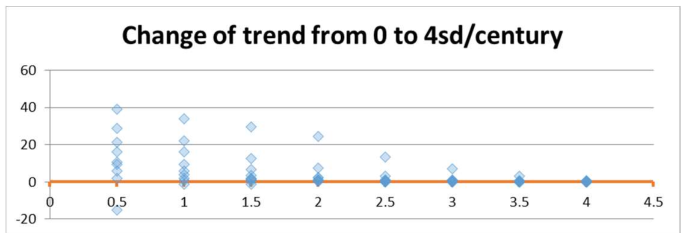
From the above results, the MSBV (with resampling) would be expected to accurately locate shifts of 1σ with trends of up to 5σ/Century in a time series of thirty years, but individual MYBV tests would not be deemed reliable with any trend until the step-size is ≥ 2.5σ. One should note the seemingly counter intuitive result that longer time-series are apparently more sensitive to trend. This is because more trend accumulates in longer series while the change-points are placed proportionately to the length.



A.



B.



C.

Figure Ch3.9: The effect of a uniform trend and of a trend change on the central estimates, and the spread of results, for shifts at random locations is shown here. In (A) shifts of 1.5 standard deviation are acceptably located by the bivariate test. In (B) a uniform trend of 4 standard deviations over the time series means that steps of 3.5 standard deviations are the smallest that can be reliably located, and the spread is symmetric. In (C) the same trend change is introduced with the step change. In this case the spread is asymmetric.

To assist, Figure Ch3.10 below illustrates the type of spread of MYBV results when small and large shifts are affected by moderate and greater trends.

In this illustration the top pane shows the sorts of spread that can occur with a smallish shift of 1.5σ. The red crosses indicate the location of the shift (0.2, 0.5 and 0.8 of the length) in the data, the solid lines a trend of 4σ/Century, the dotted lines are trends of 8σ/Century. The lower pane shows a substantial shift of 3.5σ. In the case of the smaller shift, central bias occurs, and increases with trend, and, as the median is displaced, the MSBV may be affected. In the other case the shift is sufficient that no bias occurs and both the MSBV and the MYBV would locate the change-point acceptably.

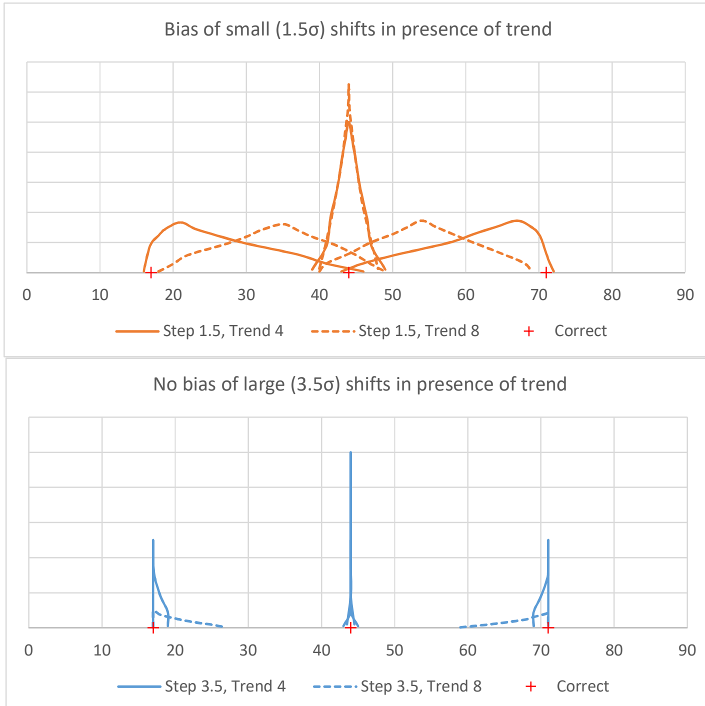


Figure Ch3.10: Illustration of the effect of trend in the MYBV in a sequence of 90 years shown as empirically derived probability distributions. Six cases are illustrated on each pane. Shifts located at 0.2, 0.5 and 0.8 of the length, in the presence of either 4σ (solid lines) or 8σ (dotted lines) trends. Top pane: shifts of 1.5σ, bottom pane: shifts of 3.5σ. Red crosses denote time of shift. Maxima in each curve correspond to medians and if p-values and processing rules permitted would be returned by the MSBV. Each curve represents the 95th percentile span. Central bias is shown in both panes but in the bottom pane the higher step size is sufficient avoid bias in the median.

Sensitivity testing of the MYBV shows that (a) small shifts may be reported closer to the mid-point of the data, but not shifts exceeding the design threshold; further, a shift of 1.5σ in a non-trending dataset should be located with adequate precision. (b) In the presence of uniform trend the shifts must be progressively larger for central bias not to occur. (c) Where trend is involved it is the cumulated trend – the actual rise or fall within the data – as a function of standard deviation that determines the minimum size of shift that can be accurately located.

Testing the MSBV for multiple steps in climate-like data

In order to assess the performance of the MSBV test with climate-like data containing multiple shifts including those below the expected thresholds, five families of artificial data sets ("A" to "E") representing a 200 year random sequence with autocorrelation (lag 1 25%, lag 2 10%)

Validation methods (Chapter 4).

Of the validation methods outlined in the next chapter, two are briefly summarised.

ANCOVA is a method for comparing two regression lines. If the regression lines fitted to data before and after are compared, ANCOVA can assess whether they are likely to be different. If they are not, then the change-point is suspect. However in the absence of trend ANCOVA is less sensitive than the MYBV.
The CP-index categorises shifts on the basis of ANCOVA and the surrounding trend.

Stationarity tests are can be used in combination to classify the data as stationary with a single change-point, or possibly as non-stationary in which case the change-point might be an artefact of red-noise or another one missed.

Box Ch3.4: Validation tests fully documented in the next chapter but mentioned here.

were tested and analysed with the validation methods. Each family consisted of four combinations of data, steps, and quadratic trend with the third series (A3 to E3) having steps and the fourth (A4 to E4) also having quadratic trend. The average size of shift amounts to 1.5 standard deviation, (which is less than the bivariate test would be expected to reliably detect in the presence of trend). This analysis also forms part of the case study reported in Chapter 4, see Table Ch4.8 for specification, Table Ch4.11 for results, Figure Ch4.4 and discussion therein. To summarise, the test detected shifts as expected. Of 40 shifts imposed on a quadratic trend, 11 were of less than 1σ, 21 between 1σ and 2σ, eight were greater than 2σ. A step in set B4 was detected in 1944 prior to the first imposed shift in the series, but is identified by ANCOVA (see Box Ch3.4) as having *p*-values greater than 0.05, and diagnosed by the CP-Index as a continuation of trend. Series B3 and B4 had shifts in successive years (1974, 1975) and this was identified, as expected, as a shift in 1974. The last two shifts in series E3 and E4 at 2071 and 2080 were both small (<1σ) and a shift was identified between them after 2074 with *p*-values < 0.05 by ANCOVA. This would count as a misplaced point. Change-points within seven years of a prior point or the end of data were not detected (as expected). Shifts of between 1σ and 2σ were more often detected than not – again an expected result. All data segments bar one (D4 at 2084) were diagnosed as stationary. So false negatives and misplaced points are presumptively attributable to small shifts combining with trend.

MSBV compared with other methods

For comparative purposes the same data were also tested using the default *breakpoints* function from *strucchange.R* with its default Aikake Information Criterion (AIC) (Ludden et al., 1994, Akaike, 1974) as a penalty (henceforth SC) and the *cpt.meaanvar* from *changepoint.R* with its default of Modified Bayesian Information Criterion (MBIC) (Zhang and Siegmund, 2007) as a penalty function (henceforth CP). SC also reports the 95% confidence ranges for the estimated year of change, (see Appendix 3.2, Table A3.2.26). SC is demonstrably prone to false positives in the absence of an expected change. As expected the MYBV is also prone to false positives in the presence of trend. Yet in the combined presence of trend and shifts neither method would be preferred. CP is the most conservative in the configuration I used and most prone to false negatives. Use of the AIC for model selection has previously been criticised as giving unreliable inferences, ignoring relevant error probabilities, and tending to overfit (Spanos, 2010). The MBIC is less widely used, being initially developed for use in genomic hybridization (Zhang and Siegmund, 2007), and I have not located a critique outside the general criticism of model selection methodology.

Case studies

Publications to date, based on this framework, have been concentrated on detecting regime changes in global and zonal temperature records, and the MSBV has been used with resampling of random controls.

Here I illustrate the MSBV as it used in published material, firstly with random reference variates, illustrated with the global and a zonal temperature record, then secondly, the related spatial versions which uses a truncated stopping criterion. I then illustrate the use of all modes of analysis using global ocean temperatures at 100m and 700m since 1955.

Global and Zonal Surface Temperatures

Here I show a sample analysis of zonal and global records brought forward from Chapter 5 which also incorporates the post-detection validation tests of Chapter 4. This is based on an analysis of NCDC zonal data version v3.5.4.201504 (for brevity NCDC). A full analysis is available in Chapter 5. Earlier results were published in Ricketts (2015a) and Ricketts (2015b), and an extensive analysis in Jones and Ricketts (2017b).

Here are shown in Figure Ch3.11 the graphs derived from the MSBV analysis of global (90S.90N) land, ocean and combined land-ocean data; and one zonal data set, the Northern mid-latitudes, 30 to 60 degrees North (30N.60N).

Points of interest include …

- In general land temperatures rise faster than ocean temperature, and such changes are dominated by upward steps. The overall differential is predicted by global climate models and seen in other analyses of observed data.
- Globally, land temperatures rose by 1.42°C of which 0.78°C is attributed to internal trends mainly after 1978, and 0.63°C occurred in three distinct shifts of 0.18°C, 0.20°C and 0.25°C. Global ocean temperatures rose by 0.53°C comprising 0.43°C in five upward and downward shifts, and 0.11°C due to a mix of cooling and warming trends. This is consistent with a state dependent heat buffering mechanism.
- In the northern mid-latitudes (a.k.a. NML, 30N-60N), land temperatures rose by 1.71°C, of which 0.46°C can be attributed to trends and 1.26°C to four shifts, the last of which in 1996/7 was of 0.54°C. Ocean temperatures changed more often with a both trends and shifts showing a mixture of mix of upward and downward values. Temperatures rose by 0.73°C (trend 0.29°C, shifts 0.44°C), with the last two shifts in 1988/9 and 1997/8 totaling 0.49°C. The combined land-ocean temperatures show only three change-points, with a 1.09°C total change (trend -0.05°C, shifts of 1.14°C).
- The inset in the top pane of Figure Ch3.11 shows the detail of the step change in Ocean temperatures circa 1986/7. When tested by ANOVA, separately the change of trend has a *p*-value > 0.7 but the change of mean (the step) has $p=0.028$. When tested together using ANCOVA to test the disjoint model against a simple trend the presence of a regime change at this time is not strongly supported ($p=0.075$). However there is stronger support for this year in the Northern latitude mid latitude data shown below it.
- The great Pacific reorganization. The result of the global analysis support the notion that a trend change occurred together with a shift in ocean temperatures 1976/7 with consequential shifts over land a little later, but this does not show in the NML.
- The so-called hiatus. Both datasets support substantial shifts in 1996/7 or soon after. The global oceans show a shift in 1986/7 and another 1996/7. The NML ocean shows shifts in 1988/9 and 1997/8. The mid 1980s event has attracted less attention than the other events but has recently been analysed (Reid et al., 2015, Reid, 2016). The 1990s event is somewhat complex and analysed in Chapter 6. A sample year is shown in the next section below.

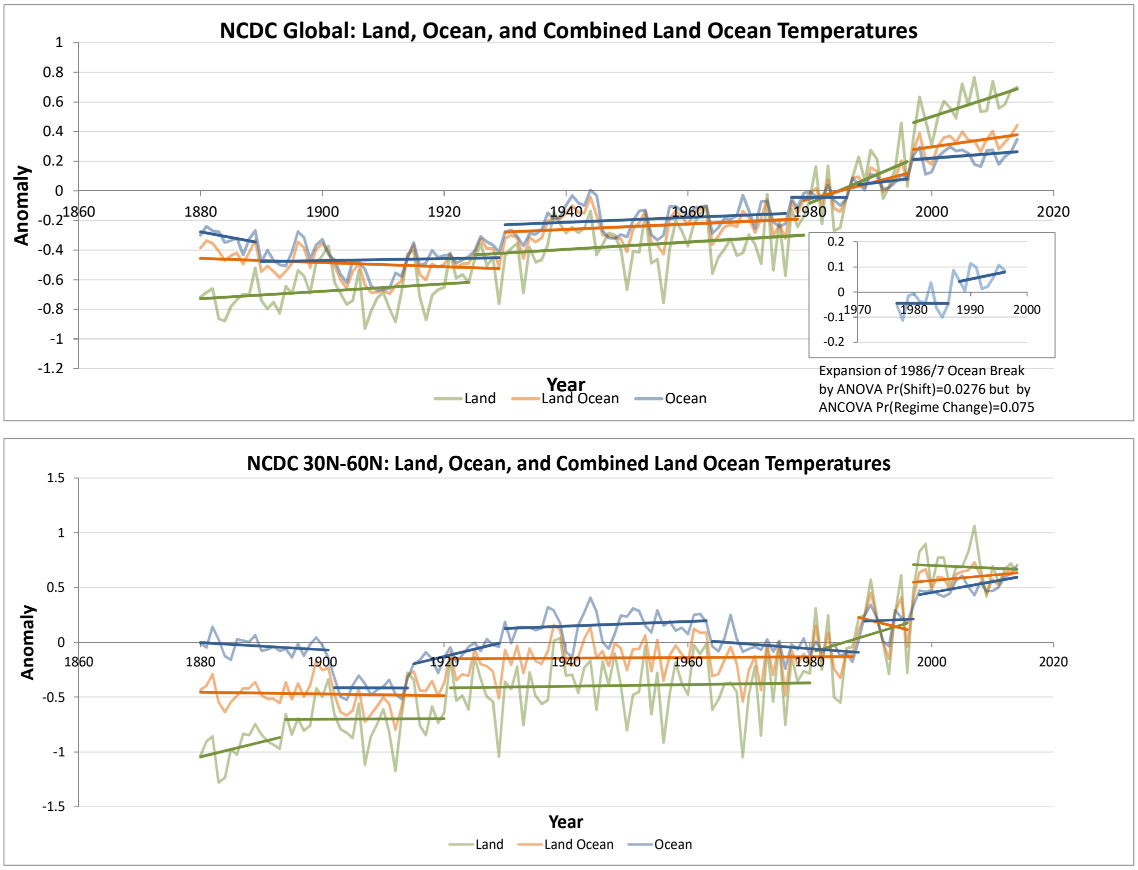


Figure Ch3.11: Top. Globally averaged temperature anomalies together with Land and Ocean splits, analysed by the MSBV. Land shows nearly twice the warming of the oceans. Bottom: The corresponding analysis of the Norther mid-latitudes.

Spatial analysis of observed temperatures.

When analysing gridded data and where it is deemed acceptable to trade precision for time, the MYBV is run with modified rapid halting criteria (RMSBV). This illustration is detail taken from preliminary work more extensively reported in Chapter 6. The application of the MSBV to gridded data enables one to undertake more detailed analysis of events. Figure Ch3.12 illustrates the patterns of change after the year 1997. The zonal evolution of the Northern mid-latitudes after 1996, which appears to be very step like (Figure Ch3.11, bottom pane) can be seen to be more complex. Zonal and spatial views are analysed in the two subsequent chapters.

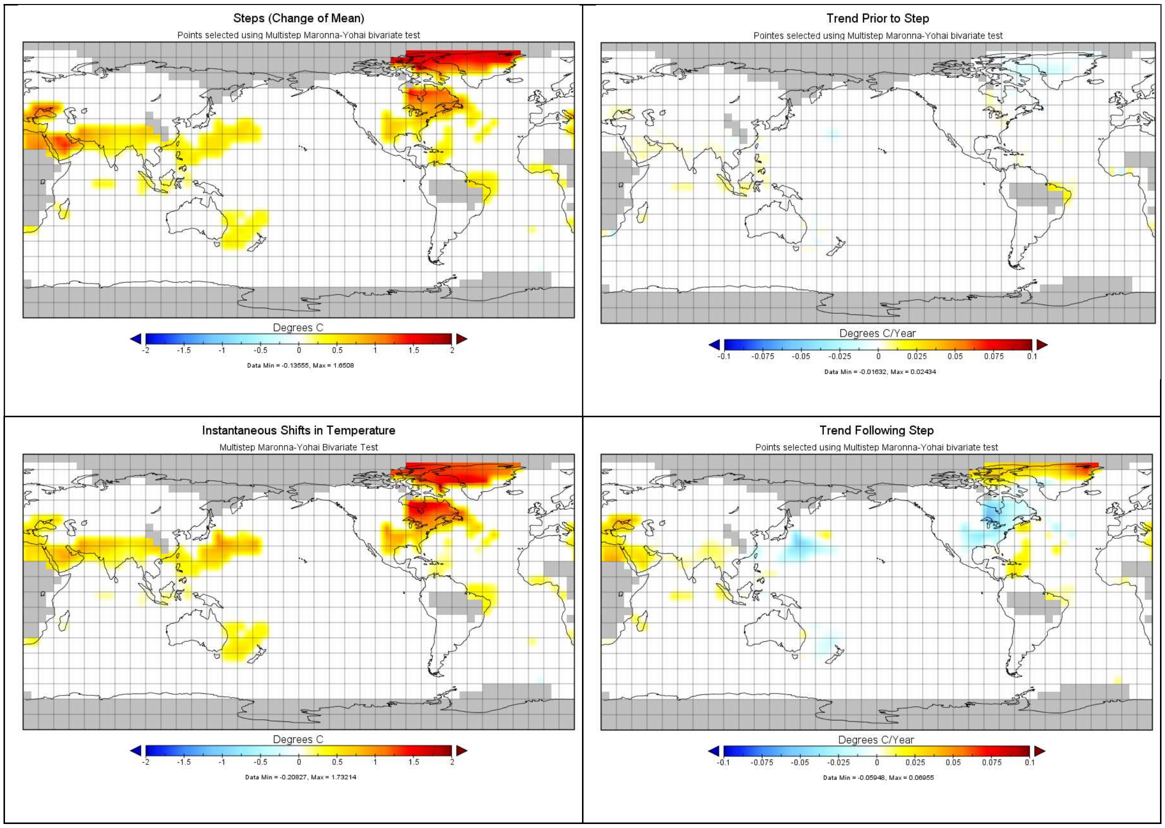


Figure Ch3.12: Here I illustrate the patterns of change associated with 1997/8 (that is, with 1998 as the first changed year). Top left (Steps) plots the steps as computed by the MSBV test – approximately the difference in means between the segment before and after the change. Bottom left (instantaneous Shifts) plots the temperature difference between the prior and the posterior segments for the year of change. Top right shows the trends associated with the locations which show a step change prior to change, and bottom left shows trends after.

Firstly I show the overall pattern of steps returned by the MSBV in the top left pane, and lower left the internal shifts. One should note that only change-points are plotted. Trends are shown on the right, before (top right) and after (lower right). Analysed this way one reason for the controversy surrounding the so-called hiatus become clearer. There was an abrupt rearrangement of the global heat map, following the change of phase of the PDO. This affected mainly areas which had not until then been contributing greatly to the global progression (see top right). Of the areas which showed an upward step, there is a slight preponderance of reduced trend, with some areas showing increased trends. The event proceeded over several years, from circa 1995/6 to 1998/9, and was preceded by an event around 1994 in the Eastern North Atlantic that quite likely corresponds to a phase change of the AMO.

Regimes in Ocean temperature relationships at 100, and 700m depth.

In Chapter6, Section 2: Vertical ocean structure, use is made of the MSBV, running as a *two way* test, see Figure Ch6.47 to Figure Ch6.49 in that section. The work goes on to show that, for example, during the surface shifts associated with the event of circa 1997, corresponding shifts in the 100m and 700m ocean temperatures also occurred, at the very least suggesting that a restructuring of the oceans was also involved.

This illustrates the results of a *two way* bivariate test of globally averaged annual temperatures at 100m and 700m depths (Figure Ch3.13). Data produced by NODC was sourced from KNMI, and consisted of anomalies (1981-2010) of monthly temperatures between 1955 and November 2015. Here annual averages were derived and the MSBV running as a two way test was used to locate change-points.

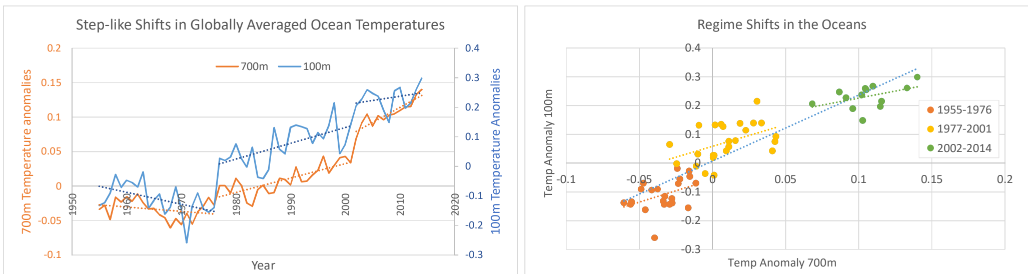


Figure Ch3.13: Change-points in global annual average ocean temperatures for depths of 100m and 700m. Change points are determined by MSBV with each variable acting as a reference for the other. Left pane. The time series and change-points. The changes in1976/7 has p-values <0.01 for both variables but the one in 2001/2 has a p-value < 0.01 only for the 700m temperatures. Right pane. The relationship between the two variables before and after the two changes. The relationship between the two changes after 1976 but not after 2001.

Two change-points are identified, 1976/7 and 2001/2. ANCOVA was then used to test whether the regression relationship between the two variables changes at the same times. By ANCOVA, the first change-point also reflects a change of regression between the two variables ($p<0.01$) but the second does not ($p=0.7$). After 1976 there is a step-like change in both temperature time-series ($p<<0.01$), and after 2001 there is another in the 700m temperatures ($p<<0.01$), but for 100m temperatures ANCOVA yields $p=0.09$. Hence round 1976 the regression relationship between the two variables changes with trend changes in both and a shift in 100m temperatures. Around 2001, 700m temperatures shift up again ($p<<0.01$) but 100m temperatures do not shift sufficiently to register as a step-like change, neither is there evidence that their regression relationship is changed (Figure Ch3.13 right pane). This may simply reflect the variability of the 100m temperatures, in which case further accumulation of data may clarify this.

Since it would be unphysical for the temperature changes at any depth to be due to step-like changes in total ocean heat, these must relate to heat distribution. This is entirely in keeping with the widely reported great Pacific reorganisation of 1976/7 which is observed in marine and terrestrial eco-systems, atmospheric circulation changes, tidal gauges and surface temperature records. The 2001 event is of interest since the 700m temperatures clearly shift, but how and if it relates to surface temperature regimes cannot be determined from this analysis. However Chapter 6 Figure Ch6.49, a spatial version of the same test shows that quite extensive reorganisation in the oceans took place contemporaneously with surface temperatures as seen in Figure Ch3.12 above.

In this case the constancy of the relative warming trend between layers of 1.5°C/°C for 100m/700m temperatures across two regime shifts suggests that the overall relationship between layers is unchanged globally and that the shifts most likely represent changes of vertical ocean configuration, possibly related to surface regimes. When analysed by MSBV in the same way as global temperatures were, as a *step-change* test, a shift after 1986 is detected in the 100m in addition to those after 1976 and 2001. The 700m time series, tested the same way has an additional shift after 1990.

Discussion and conclusion

Advances

This chapter has detailed the extension and automation of the Maronna-Yohai bivariate test for detection of multiple abrupt changes in a time series. Although the main published work utilising this has been with various climatic temperature series, observed and modelled, the framework extends easily to bivariate series and to gridded data. Both the MSBV and the rapid MSBV can be performed using either a random control variate as per MYBV Model II or with a correlated control variate as per MYBV Model I. One can also perform the latter with two perturbed variates.

The automation of the Maronna-Yohai bivariate test has solved a problem of researcher choice/bias and also enabled users to explore the nature of decadal scale progression of Earth's temperature records in some detail. In this and related work we have been able to compare and publish analyses of various estimates of the mean annual global temperature, mean annual zonal temperatures and comparisons of land and ocean splits (JR2017). We have been able to explore and publish the mean annual global temperature records from CMIP5 global temperature records (ibid). Because of the probabilistic nature of the test running with random control data, the framework can be used to discover alternate plausible break-sets.

To the best of my knowledge this is the only step-change detection software that can run over gridded data at informative scales and handle both univariate and bivariate relations. A recent paper has used a method based on a ramp model to analyse abrupt changes in sea surface temperature time-series, reporting on duration of change events (Yan et al., 2016), however the computational costs and limits are not reported.

MY78 analyses two statistical models and also briefly extends the method to two potentially perturbed but presumptively correlated series. This latter case does not appear to have been reported in the literature.

Sensitivity testing has been performed and more detail shed on the mathematical underpinning, especially as it pertains to the use of random reference series. As a result the limits of the method are better understood and previous published work substantiated.

It may seem counter intuitive to use a step-only detection method when analysing for a regime change where a trend change is not impossible. This has been tackled from both theoretical and experimental angles. In this work, the precision of determination of timing of an event is a major concern. The reason that a step method is preferred is shown by examining the error behaviour of the detection of a level change. It can be shown the error bounds on time of change are a function of standard deviation of the time series, whereas determination of trend-changes are a function of the standard deviation of difference series, which converges on double that of the time-series. The papers introducing Jaruskova's method (Jarušková, 1997, Jarušková, 1996), make the related point that if a time series may contain multiple level shifts then each can be identified, but if it has no level shift or if this is ignored and there is more than one trend change, there is no proper way to isolate the time of *any* of the trend changes; detection of both steps and trend changes in a time series is hard (Jones, 2012, McDowall, 1980). Sensitivity testing shows that the multiple resampling strategy adapted from VJ2005 increases the precision of determination T_{i0}, and given data that conforms to the assumptions, error bounds of ±2 years suggested by Beaulieu et al. (2009) are achieved with even quite small shifts (1.5 standard deviations). It also shows that if data contain trends then the step-changes need to be bigger to avoid bias.

The case studies

Both the zonal and the spatial analyses highlight the importance of using modal values and resampling. The Great Pacific Reorganisation, reported as occurring after 1976 most likely occurred over oceans around 1976, but not until after 1978 are persistent step-like changes seen principally over land. Analysis of a composite of land and ocean signals sometimes shows a bimodal response with both years represented.

Very approximately, the data reported previously (JR2017), in which for the most part internal trends are negligible, has a standard deviation of 0.12 to 0.16. So shifts of 0.2°C or greater should be regarded as adequately located.

The example of the abrupt changes in relationship between maximum temperatures and rainfall in RJ2012, demonstrates that to be useful a control variate does not have to be, as in homogeneity testing, the same type of variable. Rather, the analysis takes advantage of a one way dependency, based on prior research.

The example of the 100m and 700m ocean temperatures with two determinate time series both where there may be a two-way dependency, shows that the choice of whether a random variate or a deterministic one is treated as a reference makes a subtle difference due to the relative insensitivity of the test to variation in the reference variable.

The shift-dates found correspond remarkably to those found in surface temperature records. 1976/7 for both depths corresponds to the great Pacific reorganisation 1986/7. The shifts after 1986 for 100m and after 1990 at 700m, not found in the *two-way* test, none the less may relate to the 1986-1988 surface temperature event. The change after 2001 may be at the tail end of more complex events associated with the so-called hiatus.

Conclusions

For present day temperatures and other variables, the MSBV is fit for purpose in that it will locate step-like change-points with satisfactory precision. However it is limited once trend begins to dominate the signal. Thus secondary testing of the change-point and of data is required, and this is addressed in the next chapter.

Chapter 4: Characterising and validating discontinuous change points in climate time series.

Introduction

This thesis utilises a statistical model of disjoint segmented linear regressions separated by step-like change-points. Change-points are detected by the multistep bivariate test (Chapter 3) based on the Maronna-Yohai test (Maronna and Yohai, 1978) (MYBT).

In this chapter I extend the characterisation of individual change-points in climate data, first published by Jones and Ricketts (2017b) (hereafter JR2017), by supplementing the resampling and window tests documented therein, and considering potential deceptions such as extreme trend, and extreme autocorrelation. This also gives a basis for attribution of changes between internal shifts and internal trend changes, and probes the assembled multiple change points for evidence of undiagnosed features. It addresses the gap between the assumptions made for the purposes of change-point detection and subsequent reasoning about those change-points.

Change-points are statistical entities, detected under an error statistical (ES) approach, and as such the statistical model used makes simplifying assumptions about the data. This is a general issue that affects analysis of any climate data, and is true of any detection method applied to them. Reasoning about the physical world also requires a probative, degree of confirmation (DC) approach where multiple lines of evidence may be considered (see the discussion of the Theoretical Mechanistic/Statistical Inductive (TMSI) framework in Chapter 2).

Model specification delineates families of statistical models. For example, in this work the family of segmented linear regression models is used. The choice of specific parameters from within a specified family is termed model selection, and would in this case include the selection of specific change-points. For physical problems, the family would be misspecified if the available parameters do not properly reflect the physical processes.

In their paper on misspecification (M-S) testing, Mayo and Spanos (2004) use an example of a linear regression model to move from an ES approach to a DC one, and address the problem of validation in regression models. Three general forms of M-S are recognised:

1. Functional form misspecification in which a statistical model includes the correct parameters but inside an incorrect function. For example, as x^2 instead of x^3 or $\sin(x)$.
2. Missing parameter misspecification in which a variable is omitted
3. Irrelevant parameter misspecification in which unnecessary parameters are introduced.

The immediate goal in this thesis is to examine the meaningfulness of two proposed functional forms in temperature data, linear trend (identified with H1), and step-like change (identified with H2). Thus the work of Mayo and Spanos (2004) is of interest, and strongly influential, but since their exemplar misspecification is of a different class (an irrelevant variable, whereas we are primarily interested in functional forms and missing variables), it is not used prescriptively.

I propose that, since the full functional form underlying the data is unknown and complex, it is infeasible to exhaustively test differing aspects of the data. Instead I take advantage of much previous work which has bounded the domain, and concentrate on tests which examine possible aspects of the data that may cause step-like change-points to be falsely identified or temporally misplaced. Likewise, the parameters representing step-like shifts and trends should not be biased since H2 (interaction of warming and natural variability) is differentiated from H1 (non-interaction) by their relationship. A suite of tests is proposed to determine whether the assumptions needed to reliably model step-like changes are met.

The suite of chosen test are all framed differently and in combination give a basis for nuanced reasoning about the meaning of the test results. So one does not simply reject a change-point on the basis of a single ancillary test, but rather searches for reasons as to why a change-point detected by a test known to be reliable in its domain, is discounted by another test based on different assumptions.

The tests can be automated, and good progress has been make towards automating the reasoning further. This involves consideration of the ruling assumptions of each test and their behaviour when presented with a variety of features outside those assumptions. I have found no guide as to how to best do this, but each of the tests was individually trialled by interested users after it became available. Their reviews and observations were incorporated and cross checked.

With a view to automated reasoning two indices were produced. One addresses the issue of the second last paragraph. The MSBV is sensitive to strong trend, and furthermore trends can change, and so an index composed from three tests is used to provide a coarse, graded assessment of aspects of trend. This then suggests further lines of examination. A second index has proved very useful. Data segments can be classified as to whether there is residual non-stationarity when the change-point is accounted for, and whether there is evidence other deterministic changes. A key finding from later in the thesis is that at progressively smaller scale data, data segments in which steps have been identified, become increasingly likely to be classed as having stationary residuals without other changes. This in turn bears directly on the scientific question posed at the start, as this is consistent with H2, and not consistent with H1

Confounding deterministic factors

In the literature a number of confounding climatic processes have been identified. By this I mean that any and all of the processes may be present, and must be accounted for. These include quasi-periodic sub-decadal processes such as El Nino/Southern Oscillation (ENSO), and decadal scale processes including Pacific Decadal Oscillation (PDO) (Trenberth, 1990, Minobe, 1997, Bjerknes, 1969, Trenberth and Hurrell, 1994), Atlantic Multi-decadal Oscillation (AMO) (McCarthy et al., 2015), North Atlantic Oscillation (NAO) (Schlesinger and Ramankutty, 1994), Arctic Oscillation (AO) (Thompson and Wallace, 1998). Additionally there may be influence due to volcanic and anthropogenic aerosols. Other influences, more often described by their statistical impact than their physical origin, include particularly auto-correlation and moving average behaviours, which can potentially vary in turn leading to transient episodes of red-noise like behaviour. Lastly, changes of seasonality can reduce the effectiveness of simple seasonal anomaly removal. The most common strategy for dealing with the last is to use annual averages, and that is what is done here.

Initial assumptions used in data preparation can influence the outcome of tests. One method for isolating the trend component of data is to identify likely factors, model them as independent additive influences, and remove their influence using multiple regression, and often using an index as a proxy for a phenomenon, e.g. Foster and Rahmstorf (2011). This however cannot be guaranteed to not interfere with a step analysis, especially if the steps are aligned with one or more of the quasi-periodic elements removed. This would happen if, for example, the ENSO influence was removed by a regression method against an ENSO index, and the PDO's temperature influence (for example) was expressed as phase-dependent biases in the ENSO index (Verdon and Franks, 2006). Trend and cyclic and quasi-cyclic behaviour is certainly expected in a climate system and likely to have physical explanations. However since the climate system is neither fully deterministic, nor as fully characterised as possible it is likely that an analysis focussed on specific features will treat some deterministic components of the signal as random noise. Periods that look to be integrated noise may still be mainly deterministic when appropriately tested – a missing variable misspecification.

The rest of this chapter proceeds as follows.

(a) Three classes of testing undertaken are briefly described.
(b) A brief glossary to introduce the terminology used in this section.
(c) The suite of methods is introduced. The MSBV used for detection provides a time and probability of change, on the assumption that the effect of trend on the timing of detection of a change-point in the data is negligible. This assumption is assessed by testing that the identified change-point has explanatory power in the presence of trend and trend-change.
(d) A misspecification due to model selection, shown by heteroscedasticity, is tested for. Four tests of the assumptions of stationarity in the data are then documented.
(e) Lastly I present an analysis of the type of simple time-wise averaging that is often used in climatology to produce, for example, mean global temperature records from multiple observed spatially distinct time series. I introduce the term "compositional misspecification" to cover the impact of averaging spatial time series when there are either multiple distinct processes or propagating change present.

Types of tests

In what follows, three classes of testing useful during analysis of step-like change-points have been identified.

Tests of probability of individual change-points.

The detection test used in this work, the MYBT on which the MSBV is built, assumes zero trend, but trend changes may occur and in fact must be properly attributed. Analysis of covariance (ANCOVA) is used post-detection of a change-point by MSBV (a constrained statistical model) to ensure that the presence of the change-point provides explanatory power in an unconstrained disjoint linear statistical model. It does not attempt to locate an alternative change-point under a less constrained model – a completely different manoeuvre. As used this is exactly equivalent to a Chow test.

If the data conforms to the assumptions of the MSBV then MSBV is a more powerful test than ANCOVA. An index (CP-index) that combines the significance of the trends either side of a change-point and ANCOVA was devised to assist with automated reasoning about changes in the presence of trend and changes in trend.

Tests of probabilities associated with sets of detected change-points.

The full set of change-points in an entire sequence is tested here by the studentized Breusch-Pagan test (hereafter SBP test) for homoscedasticity of the residuals of the disjoint multi-segment model (JR2017 utilised the equivalent White's test (White, 1980)). An adequate model explanation of a time series, under the assumption of i.i.d. error, should have a featureless residual. Since we know that the i.i.d. assumption is an approximation at best, this test has a null of homoscedasticity, rejected in favour of heteroscedasticity at low probabilities.

Tests of stationarity

In these tests the segment containing a provisional change-point is tested for features (particularly non-deterministic ones) that may deceive tests for shifts and trends. The MYBT, ANCOVA, and where used, ANOVA tests have ruling assumptions of serial independence. The MSBV, and other multiple break tests assume some form of censorship between provisional data segments (determination of change-points within provisional bounds includes only the data within the bounds); but tests of the overall model assume homogeneity of error, thus of variance (e.g. the Akaike Information Criterion or AIC). The SBP also assumes this. All of these above tests are formalised as null hypothesis statistical tests (NHST) and as such they each are subject to their own ruling assumptions. The ruling assumptions are incorporated in the interpretation of the tests.

Autocorrelation is variously treated; some propose its estimation and removal (Rodionov, 2006b), some warn against this idea (Mizon, 1995). Some treat it as a short term process and a cause of deception in change-point analyses (Beaulieu and Killick, 2018), others have treated it as a persistent signal (Percival et al., 2001). In climate signals, autocorrelation often appears to be time varying. Therefore in this thesis I apply the MSBV without adjustment for autocorrelation and perform post-detection analysis to determine whether the detection test is likely to have been interfered with. Almost all autocorrelation tests, including the unit-root tests mentioned below, assume absence of step-like changes.

In this work, both the raw data, and the residuals after removal of internal steps and trends, are tested. The rationale for testing both derives from the formulation of the tests themselves, since in these tests, the deterministic and non-deterministic components are separately parameterised. Climate data is known to contain complex lag and correlation structures – and detection tests can be sensitive to these due to the governing assumptions of the tests themselves. The set of tests chosen are from the econometric literature, and each is framed as a null hypothesis significance test (NHST) with its own specific assumptions. Each test poses either H_0 or H_1 as presence of an assumed non-deterministic unit root progression (see Chapter 2) in data, and the alternatives are chosen from a small range of deterministic features. Crucially, each must be interpreted in the light of its own ruling assumptions.

The full process applied to a single time-series is then …

(a) the MSBV is applied to delineate provisional change-points. The resulting statistical model would be accepted as the best estimate (i.e. further testing of change-points not warranted) if the time-series of the residuals was *known* to be i.i.d., *and* underlying physical processes were fully deterministic, and fully reflected in the time series. However this should not be simply assumed.

(b) The segment containing each provisional change-point is tested to ensure that to a feasible extent, physically plausible types of deception are not present, and that change-points are deterministic, not stochastic quirks.

(c) The set of detected change-points is treated as a disjoint segmented model and the residuals examined for evidence of a misfit of model to data.

The program of tests thus sharpens the error-statistical reasoning component of the TMSI framework introduced in JR2017 and discussed in Chapter 2, in line with the severe testing requirement.

Terminology

This work borrows from econometric (e.g. Parker, 2018), and signal processing (e.g. Granger and Morris, 1976, Smith, 1997), approaches to time-series analysis. It applies them to climatological signals. The differing approaches are grounded in their own distinct literature, resulting in a variety of similar terms for various phenomena. Papers grounded in economic or physical *systems* also vary in terminology from those grounded in corresponding theoretic/statistical *analyses*. The following are some explanatory notes.

Deterministic and non-deterministic trends in time-series. "Trend", in a general sense is the tendency of the values to cluster about some smoothly differentiable function. In the deterministic case this is usually a function of time or other variables indexed by time. Often a linear trend is referred to. The residual – that not explained by the trend – is often referred to as the *error*. If there is no deterministic trend then any tendency of the signal to vary over time is *stochastic* trend due to the *error* terms. The simplest and most mathematically tractable type of error is white (serially non-correlated and usually Gaussian) noise. A white noise time series is described by two parameters, its mean, and variance, with successive values being independent. A process is also statistically *stationary* if its mean, and the noise parameters are unchanging. If it has a constant trend it is *trend stationary*. If it has zero trend it is *level-stationary.*

Integration order: Complete independence of successive errors leads to the description of "integrated order zero", or I(0) noise. (Also referred to as "independent and identically distributed" or i.i.d.). More generally, the integration order defines the number of times a difference must be applied for a signal to become stationary (which is a technical requirement of ordinary least squares (OLS) regression). If the error accumulates over time it is *red-noise,* or integrated order one, I(1) noise. The literature also refers to *difference-stationarity*, where a signal is not stationary but the difference series is; and *trend-stationarity*, where the residuals about the trend are stationary. The first case would be *non-deterministic*, the second, *deterministic*.

Systems may contain both deterministic and stochastic components. A signal in which there is a deterministic component and a stochastic component describes a situation where the random error series accumulates separately from the deterministic part.

Unit Root: The term itself comes from the signal processing literature, and the representation of the characteristic equation of a process on the complex plane. An equation has a unit root if the random component of a signal from one time is fully carried forward into the next time. See the discussion of on Page 96. Unit roots are characterised statistically as being non-stationary, and in particular having a variance that expands over time. A signal with a components of I(0) noise may be referred to as having an I(0) or moving average (MA) unit root. If the unit root is I(1) this may sometimes be called an auto-regressive (AR) unit root.

Exogenous and endogenous change. These are econometric concepts. In short, an endogenous change is one that follows naturally from the system without external influences and an exogenous change is one that has been imposed externally. If a characteristic equation of process has both a unit root and deterministic trend, a one-off injection of noise (a shock) will cause an apparent persistent change since it is carried forward thenceforth. This (*endogenous change*) may superficially resemble a change of deterministic parameters of the process (*exogenous change)*. Many tests for stationarity assume no exogenous change and will diagnose non-stationarity. The cases are differentiated by attempting to remove the influence of a deterministic change, and so a test that diagnoses a change as exogenous has stationary residuals, and a diagnosis of endogenous change corresponds to non-stationarity.

Complex deterministic signals may be found to be non-stationary in some cases due to violation of the ruling assumptions of the tests used. Their change-points would be exogenous, and segments between change-points would be stationary. I refer to this as deterministic non-stationarity. By contrast stochastic non-stationarity is due to endogenous changes and segments between change-points would remain non-stationary.

Validation suite methods

Statistical significance of change-point

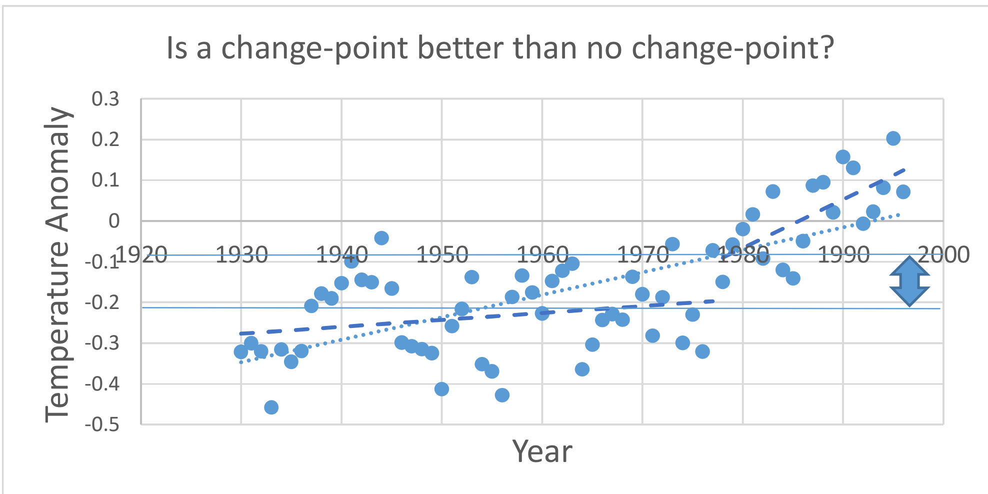


Figure Ch4.14: The ANCOVA test is a post-detection test that probes whether a change-point improves the goodness of fit, is it required at all? The dotted line represents the OLS regression line for the data as a whole, the dashed lines represent the OLS regression lines of data taking into account a single change-point. The double headed arrow is the internal shift. Note that the step computed by MYBT would be approximately the difference in the means of the two segments.

For my purposes, the segment of data which contains a change-point is internally partitioned into two sub-segments – with one internal shift and two internal trends. An internal trend is the ordinary least squares linear trend of a sub-segment, and the internal shift is the dislocation or step between them at the change-point. The MSBV – the detection method – assumes an internal shift only. Its statistical model has three degrees of freedom, an initial mean (or intercept), a time of change, and shift or change of mean. ANCOVA assumes both an internal shift and change of internal trend, and has one more degree of freedom, initial intercept and trend, post-change intercept and trend, but time of change is stipulated by the MSBV.

```
#the following call sets up ANOVA to assign significance to trend
#changes and shifts
#(see http://r-eco-evo.blogspot.com/2011/08/comparing-two-
#regression-slopes-by.html

lm(formula = Anom ~ Class * Year, data = data)

#aov() calls the linear models routine lm()
#compute residuals for a two segment model using Class to split them
mod1<-aov(Anom~Year*Class,data=data)
#and a single segmented model
mod2<-aov(Anom~Year,data=data)
#compare the residuals from both models with ANOVA
anova(mod1,mod2)

Which yields for example

Analysis of Variance Table
Model 1: Anom ~ Year * Class
Model 2: Anom ~ Year
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     62 0.54913
2     64 0.77291 -2  -0.22378 12.633 2.5e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Box Ch4.5: The core R code. Year2 is the Year expressed as an offset from the year to be tested, Class is categorical, 'Pre' or 'Post', Anom is the mean annual temperature expressed as an anomaly as defined by the data providers.

Figure Ch4.14 demonstrates this. ANCOVA is used to test that the two segment linear model, formed about the change-point, and shown as the dashed lines, is preferable to a single linear regression shown as the dotted line. The null hypothesis is that the change-point makes no difference – for our purposes, that the change-point is a possible false positive.

Here, the R statistical language is used. ANCOVA is implemented via a pair of ANOVA tests, one based on a model with data partitioned into pre or post change, the other with no change (see snippet of R code in Box Ch4.5). The R code takes a temperature anomaly ("Anom"), the year relative to the change year ("Year") and a classifier ("Class") comprising ("Pre" and "Post"). Formally, ANCOVA takes the null hypothesis H_0: a single OLS regression explains the variation in the data as well as would two disjoint regressions, one either side of the break. Rejecting H_0 at a given *p*-value is evidence that the change-point partitions the data into differing regressions with a corresponding likelihood. When the change-point is established via the MSBV (*p* <0.01 in this thesis) and *not* supported by ANCOVA, it is evidence that the change-point does not correspond to a combination of shift and trend-change. In the presence of underlying trend, the apparent change of mean found by the detection method might be attributed to ongoing trend. However, note that changes of trend are hard to prove in short data segments, and the ANCOVA itself may reject genuine change-points on this basis. For that reason, a more complex and informative measure (the CP-index, next section) is proposed. ANCOVA and the MSBV, both being based on an assumption of i.i.d. data, means that stochastic trend or drift must also be checked for.

During the analysis separate ANOVA tests are applied as consistency checks, to assign probabilities to shifts, and to trend changes, assuming independence. And these are shown for interest in the various tables in Appendix 5.1.

Reasoning about change-points from different lines of evidence: CP-index

The MYBT, if its assumptions are met, returns lower probabilities (higher significance) than ANCOVA, even though it first selects a change-point. However data may meet the assumptions of ANCOVA but not those of the MSBV. In this thesis a sufficient number of change-points are examined in the analysis of some problems that a compact, automatable summary statistic was useful. I define a CP-index as follows. If the trend prior to the change-point is different from zero with *p*<=0.05, that counts as 1; if the post change trend is different from zero with *p*<=0.05, that counts as 2; if the *p*-value from the ANCOVA <= 0.05 that counts as 4. This gives a range of values from zero to seven. Zero represents an MSBV <= 0.01 and an ANCOVA of >0.05, something that can occur in shorter data sets. Values of one or two indicate that one segment of the two has a trend. I take a value of three (ANCOVA *p*>0.05 but both trends *p*<=0.05) as indicating the trend is such that the MYBT value may be deceptive. Four would indicate that the shift is big enough that it meets the assumptions of the MYBT and ANCOVA. Five, six and seven all indicate that trend is present but the shift can be accepted.

Heteroscedasticity Testing

In JR2017, Whites's test was applied to entire records, and to 40-year windows as part of a statistical inter-model comparison showing that step changes in general have more explanatory power than trend-only models. These tests were conducted separately with extensive researcher effort. Here I use the equivalent studenized Breusch-Pagan test (SBP) from the R ‘lmtest’ package (Hothorn et al., 2015), with no auxiliary variables specified. It is applied, as in JR2017, to the entire record. The test estimates the residuals from a fitted model and has a null hypothesis of homoscedasticity in the residuals with an alternative of heteroscedasticity. Three stochastic models are applied for comparative purposes. A linear model is fitted to determine whether by this test, a linear model is sufficient, where a determination of homoscedasticity would be unexpected if any exogenous change exists in the entire time series. Similarly a quadratic curve is fitted to test whether the data record conforms to an accelerating trend, as this is a condition that can interfere with change-point testing. The derived disjoint regression model deduced from the MSBV is then fitted to determine whether there is evidence for unaccounted processes. These tests are applied to the entire time-series after a change-point model has been obtained. When an underlying trend is smooth, any linear segmented model, disjoint or not, applied to it is a misspecification, with potential to be deceptive. This is because there is no available procedure for identifying a preferred starting point (Jarušková, 1997, Jarušková, 1996). Any piecewise model assumes discrete change-points as is assumed by this suite. This means that similarity across comparisons under differing approaches to change-point detection (e.g. structural change analysis, CP regression) do *not* serve as assurance the change points are valid. For instance structural change tests assigned more change-points when tested on curve only series than did the bivariate test, consistent with (Jarušková, 1996). Formally stated the SBP test has a null hypothesis of homoscedasticity, rejected in favour of an alternative of heteroscedasticity.

The sample code run in Box Ch4.6 shows, in this case, clear indications that the data is consistent with a disjoint step and trend statistical model as illustrated by Figure Ch4.15. The example data is taken from one member (D4) of the artificial dataset DS2 discussed further on. In this case, the residuals of a step and trend model obtained by running the MSBV show no detectable heteroscedasticity whilst the single linear trend most certainly does. Figure Ch4.15 omits the quad-model for brevity. The residuals from the single linear regression show a clear structure, whereas those from the step and trend model do not.

```
#specify three models using lm
breakmodel<-lm(Anom~Year*Class,data=data)
linearmodel<-lm(Anom~Year,data=data)
quadmodel<-lm(Anom~Year+I(Year^2),data=data)
#test the models without auxiliary variables
bptest(breakmodel,data=data)
bptest(linearmodel,data=data)
bptest(quadmodel, data=data)

#which given accelerating trends and nothing else might yield …
studentized Breusch-Pagan test
data:  breakmodel
BP = 10.028, df = 5, p-value = 0.9567
data:  linearmodel
BP = 4.9908, df = 1, p-value = 0.0018
data: quadmodel
BP = 1.351, df = 2, p-value = 0.0021
```

Box Ch4.6: Sample R code and result given an analysis of quadratic trend embedded in noise. The change-point method has returned three change-points. The break model shows homoscedastic residuals, neither linear nor the quad model do so.

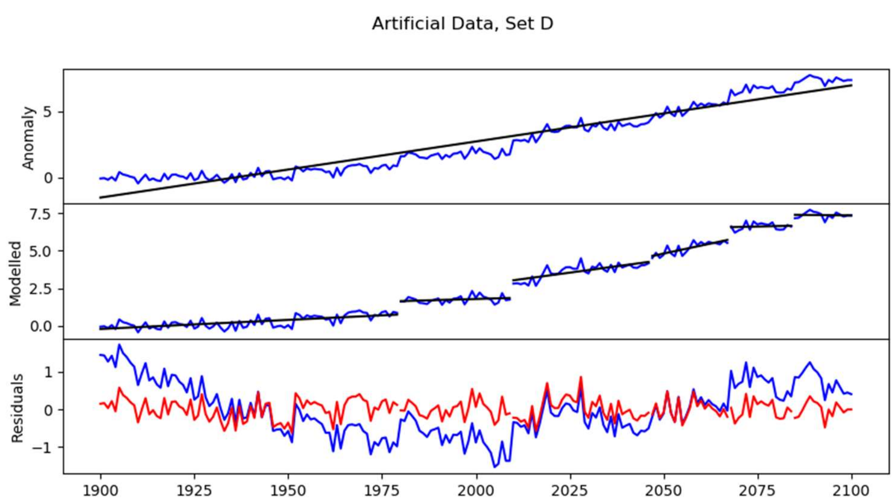


Figure Ch4.15: Heteroscedasticity and homoscedasticity in the residuals of a single segment regression and the step and trend model deduced from the MSBV. Top pane; data and the OLS trend –line. Middle pane; the same data showing the change-points from the MSBV. Bottom pane; in blue the residuals corresponding to the top pane, and in red, those corresponding to the middle pane. The studentized Breusch-Pagan test indicates that the residuals of the linear fit (blue) are not homoscedastic (p=0.0018), but those of the step and trend model most probably are (p=0.96)

Deceptive features detectable with unit root tests

The application of deterministic tests such as OLS to non-deterministic data progression is a misspecification; the results may be deceptive and one should not infer a causal structure. Unit root tests probe the data for features that can imitate deterministic structural changes due to stochastic behaviour of an unrecognised red or near red progression. As will be shown later however, unit-root tests may also be used to detect non-stationary sequences. Because multiple tests are performed, and because they each have differing ruling assumptions, the tests are interpreted in terms of evidence for and against stationarity in the underlying processes.

Unit roots, non-stationarity, and climate

Whilst Foster and Abraham (2015) assert that autocorrelation can be ignored in annual data, it may give inflated trend significance. Transient unit root behaviour, if it occurred, could indicate some sort of regime change, temporarily decoupled from normal forcings. If, in addition, measured noise was not persistent this would show I(0) behaviour; or, if it were fully persistent, as I(1) behaviour. In regional signals in which this occurs, the region may also have become coupled to other sub-systems (Tsonis et al., 2007). This could indicate that the underlying physical model is incomplete and that a missing variable misspecification has resulted. On the other hand, persistent unit root behaviour means that a deterministic change-point analysis is suspect. For example it has been shown by Monte Carlo methods that a test for deterministic trends will find deterministic trends in about 85% of realizations that contain only a stochastic (unit root) trend (Chang et al., 2016).

The Earth system is constrained so that the temperature cannot solely follow a pure random walk – at worst it would follow a Brownian bridge (i.e. sequences where the end-points are meaningful and accepted as deterministic but the path is apparently a random walk, (e.g. Fischer et al., 2013)). However the composition of summary deterministic signals, such as the GMST, involves manipulations that can produce data that existing unit root tests will identify as containing unit roots, and furthermore deceive deterministic tests in much the same way as random walk data. This issue was extensively examined and is addressed later.

Detecting unit root presence

The previous discussion has shown that random walk/unit root progression may be present in climate data because of transient physical conditions, or because the data is unrelated to the physical processes assumed (M-S due to irrelevant variables). Additionally there may be features in the data that do not correspond to any of a shift, a trend change, or unit root behaviour (M-S due to missing variables), and UR tests are potentially sensitive to this. This source of deception must also be dealt with. Other features may be present in the data but not detected. For instance, a step-like shift well above a detectability threshold may be present together with a number of small, deterministic shifts below detectability, and this latter may be taken to be evidence of stochastic drift by a test.

The unit root based tests used here all inherit in one form or another the Dickey Fuller (DF) model (Dickey and Fuller, 1981) which describes a time series with autocorrelation in terms of its previous values as .

$$Y_t = \mu + \beta t + \rho Y_{t-1} + e_t \quad (1)$$

… where ρ represents the portion of the signal (Y_{t-1}) carried forward by autocorrelation, β represents the (deterministic) linear trend, μ represents the intercept, and e_t is the i.i.d. error with zero mean and a constant variance σ^2.

In Equation 1, if $\rho = 0$ then this describes a deterministic trend with no autocorrelation, if $0 > \rho < 1$ there is a deterministic trend with a degree of autocorrelation, and if $\rho = 1$, regardless of other parameters it contains a unit root. If all other parameters are zero and ρ equals one, then there is no deterministic trend, no offset, and Y_i is a random walk. This formulation is modified and sometimes rearranged in different ways by the three tests used here.

It is important to note that time-series of successive differences of a step-change in an otherwise stationary time series will contain only one difference that would be expected to be unusual. So the DF model is intrinsically insensitive to deterministic step changes. Another important property of a unit root process is that the variance of the process increases over time, whereas the variance of a stationary process is constant. This gives a second strategy for determining unit root like behaviour – testing for diverging variance. The Kwiatkowski-Phillips-Schmidt-Shin test (KPSS), (Kwiatkowski et al., 1992) examines the properties of the variance rather than the fitted parameters, and it primarily focussed on determination of stationarity. As a result it is more sensitive to exogenous changes.

Proposed tests and strategies

The unit root methods used are all coded in R and are, (a) a development of the DF test, the Augmented Dickey-Fuller test (ADF), which takes H_0 of a I(1) unit root against an alternative H_1 of a presumption of no unit root (in this implementation trend and multiply lagged autocorrelation is allowed for), (b) two variants of the KPSS, which takes a H_0 of stationarity (or trend-stationarity) rejecting it in favour of an alternative H_1 of a presumption of unit root, and

(c) the Zivot-Andrews test (ZA) (Zivot and Andrews, 1992), which takes a H_0 of I(1) unit root behaviour with a possible endogenous drift against an alternative H_1 of trend-stationarity with exogenous structural change. A trend change or a step change would constitute an exogenous structural change.

Use of a combination of UR tests is not new. The combination of ADF and of KPSS testing (see below) has been used before in order to add precision to an analysis of monthly inflation expectations (e.g. Fukac, 2005 Appendix B).

Since the tests are being applied to data in which deterministic, exogenous, step-like changes are detected and thus assumed to have occurred, but no such change is allowed for (except in the ZA test) the presumption of unit-root in H_0 or H_1 of the above tests is reinterpreted as evidence of non-stationarity. Evidence of unit-root like behaviour is then sought by examination of the residuals after the removal of the deterministic internal trends and shifts detected in the data.

In general, where evidence of a unit-root is detected, it may be due to undetected deterministic features, and hence will be initially treated as evidence of either deterministic non-stationarity or stochastic non-stationarity (see “Terminology”, Page 88).

ADF

The ADF test is a variation of the DF test for trend stationarity in the possible presence of unit root. It has a null hypothesis of unit root against an alternative of stationarity after compensation for auto-correlation (Elliott et al., 1992, Dickey and Fuller, 1981) (see Chapter 2). The ADF test has relatively low power, and in this type of application a finding of a UR may be because of a single deterministic permanent change (Byrne and Perman, 2006).

Equation 1 is expanded to allow for multiple lags in the case of the Augmented Dickey Fuller (ADF) test, taking advantage of the recursive nature of the formula. This is more explicit below where k multiple lags are included as $\sum_{j=2}^{k} \rho_j \Delta y_{t-j+1}$. The difference series is then …

$$\Delta Y_i = b_0 + b_1 t + (\rho_1 - 1) Y_{i-1} + \sum_{j=2}^{k} \rho_j \Delta y_{t-j+1} + e_t \qquad (2)$$

… and a unit root exists if $\rho_1 = 1$. The equation components are coloured as follows. Green in equation 2 denotes the trend, red denotes the lag autocorrelation that determines unit root, tan denotes the higher order lags, blue denotes the stochastic component. The number of lags can be specified by the user or, as here, selected by using an information criterion.

The ADF test implementation used is programmed in R, available in the package ‘urca’ (Pfaff et al., 2016), and removes auto-correlation followed by a Dickey-Fuller (DF) test. Three variants are available, which test for a unit root, a unit root with drift, and a unit root with drift and a deterministic time trend – which corresponds to the model of Equation 2 (above). I use the latter variant and choose to allow the routine to select suitable autocorrelation lags on the basis of an information criterion, using the call “ur.df(ys, type="trend", lags=7, selectlags="AIC")” following Hacker (2010). These choices are dictated in part by the requirement to automate testing, and because the resulting possible reduction in power in the test (inability to distinguish unit root from near unit root) is compensated by other tests in the suite. The test assumes no exogenous change, and H_0 may be accepted in the presence of one (Kočenda and Černý, 2015 page 76).

KPSS

There are two variant of the KPSS test used here to test for level and trend stationarity. These tests invert the sense of the testing with respect to the ADF test, rejecting an H_0 of stationarity in favour of H_1, a presumption of a unit root. In this case a regime shift may well appear as H_1, with a step change being non level stationary and a trend change being non trend stationary. I use the R package ‘tseries’ (Trapletti et al., 2017) and invoke the two tests as kpss.test(ys), to test for level stationarity (henceforth KPSS-L) and kpss.test(ys,null="Trend") to test for trend stationarity (henceforth KPSS-T). Step-like changes in a time series are by definition non-stationary changes, hence would not be expected to pass tests for level stationarity, but step-like changes in the absence of a trend change will be expected to pass the test for trend-stationarity.

KPSS tests are designed to give weight to stationarity. Assuming that the time-series can be decomposed into the sum of a deterministic trend, a random walk and a stationary error, the model of Equation 1 (above) is re-parameterised as follows with r_t representing the random walk

$$\begin{gathered} Y_t = r_t + \beta t + u_{1t} \\ r_t = r_{t-1} + u_{2t} \end{gathered} \tag{3}$$

Where u_{1t} is a stationary process, and u_{2t} is an i.i.d. process with mean 0 and a variance σ^2.

If $\sigma^2 = 0$ then r_t is constant and the stationary process u_{1t} dominates. If not, then a unit root enters via u_{2t} and r_t is a random walk. Under a random walk, variance increases with time. Therefore this expectation is tested by estimating the variance using the Newey-West estimator (Newey and West, 1994) s^2. To test for trend stationarity, a residual series ($\{e_1 .. e_n\}$)

is given by residuals of an OLS linear regression ($\{e_1 .. e_n\}$).). To test for level stationarity the residual series is replaced by $e_t = y_t - \bar{y}$. Then for both cases, partial sums of residuals are defined as $S_t = \sum_{i=1}^{i} e_i$ and for T samples, the test statistic is given as

$$LM = \frac{\sum_{i=1}^{T} S_i^2}{s^2 T^2} \tag{4}$$

Both the ADF test and the ZA test below, perform by estimating an auto-regression parameter by OLS, whereas the KPSS tests examine the properties of the variance of the time series (KPSS-L) or the difference series (KPSS-T).

Zivot-Andrews test

A drift due to unit root could appear as a trend change, or less likely a step change, either of which would be classified as a deterministic/exogenous change by a shift detection method.

The Zivot-Andrews test (ZA) (Zivot and Andrews, 1992) tests for the presence of a unit root (with possible drift) against an alternative of stationarity with at most one exogenous change. An advantage is that the test also returns a time of possible exogenous change (Glynn et al., 2007) – but note that an exogenous change can be any of step, transient or trend change. The code is in the R package "urca", called as "ur.za(ys, model="both")", which allows for changes in trend or steps. H_0 is UR without exogenous change. H_1 is trend-stationary with a possible exogenous change at an unknown time. The ruling assumption is that there is at most one exogenous structural change, and thus is not often used when more than one such may be present. Also, in a multivariable model, that only one exhibits unit root. In either of these cases other tests are preferred (Liddle and Messinis, 2015). Here, I am testing a single variable with intervals bounded by breaks within which we have already detected exactly one break, whilst others may be below a detectability threshold. It has also been shown that rejection of the null of a unit root could be due to a structural break even in the presence of unit root (Gay-Garcia et al., 2009), whilst the presence of more than one break in the absence of a unit root may lead to the acceptance of the H_0 of UR (Lumsdaine and Papell, 1997).

Acceptance of H_0 does not imply merely UR, but rather, UR without a single break, (Byrne and Perman, 2006), and thus H_1 means not UR or not a single break. Given we know there is a break (detected by MSBV, confirmed by ANCOVA), H_1 means not UR, or more than one break.

The model used here is that documented by Zivot and Andrews (1992) as Model (C). The model follows the ADF approach and its equation contains more complex parameters for: intercept and change of intercept (a step-like change), $\hat{\mu} + \theta DU_t(\hat{\lambda})$; and trend and change of trend, $\hat{\beta}t + \hat{\gamma}DT_t^*(\hat{\lambda})$. The remaining parameters are similar to the ADF; autocorrelation with lags, $\hat{\alpha}y_{t-1} + \sum_{j=1}^{k} \hat{c}_j \Delta\, y_{t-j}$ and the presumed i.i.d. error …

$$y_t = \hat{\mu} + \theta DU_t(\hat{\lambda}) + \hat{\beta}t + \hat{\gamma}DT_t^*(\hat{\lambda}) + \hat{\alpha}y_{t-1} + \sum_{j=1}^{k} \hat{c}_j \Delta\, y_{t-j} + e_t \qquad (5)$$

Circumflexes above represent estimates of parameters. $\hat{\lambda}$ is a value that is minimised during the search for the most likely time of a break, $DU_t(\hat{\lambda}) = 1$ if $t > T\lambda$, the time of change, 0 otherwise , and $DT_t^*(\hat{\lambda}) = t - T\lambda$ if $t > T\lambda$, 0 otherwise. Parameters estimated include the time of change and each of the parameters of the above model. $\hat{\lambda}$ is estimated so as to minimise the one side t-statistic for $\alpha = 1$, which in turn leads to rejection of the null. One should note that in the absence of any deterministic change-point the test functioned as a stationarity test when empirically assessed (next section).

Controlling for false positive and false negatives.

For all of the above tests, the R implementation takes published critical values of the test statistic at the 0.01, 0.05, and 0.1 levels. The KPSS implementation interpolates the test statistic against these values to give probabilities between 0.01 and 0.1, the ADF and ZA implementations simply give the critical values and the test statistic.

None of the tests above consider unit root presence or absence when possible structural breaks (such as shifts or trend changes) exist under both the null and alternate hypotheses. The problem is under active consideration (Kejriwal and Perron, 2010, Harvey et al., 2013, Liddle and Messinis, 2015).

Empirical quantification of false determination rates

All of these tests are posed as null hypothesis tests. As such they only reject the null hypothesis at a particular level once sufficient evidence is found against it, and when the data size is limited, the power (the probability of correctly rejecting the null hypothesis) is similarly reduced. Therefore, the four tests were each tested separately for their false positive and false negative rates using a Monte Carlo method and data segments from length 20 to 100. 1000 iterations were performed at each segment length. On each iteration a random data segment representing deterministic stationarity, and a pure red segment (100% autocorrelation), representing unit-root, were analysed by all four tests, and p-values were collated as Dataset 1 (DS1). These values are used during interpretation, so that for each every test may be interpreted (e.g. a 5% level of significance) as one of three results, H0 supported, H1 supported, N/A; this last meaning that the particular interpretation that would have been given (whether H0 or H1) would only be meaningful given more data.

Further interpretation

The KPSS tests take trend or level stationarity as the null hypothesis and a false positive is the incorrect determination of non-stationarity. This is determined from the stationary data segments. Similarly, false negative rates – the incorrect determinations of stationarity, are calculated from non-stationary segments. The ADF test has a null hypothesis of unit root, so false positives are false determinations of stationarity, estimated from non-stationary data, and false negative rates (false non-stationarity) are estimated from stationary data segments.

The ZA test has a null hypothesis of non-stationarity (with possible drift) against an alternative of deterministic stationarity with a single possible change-point. The false positive and false negative rates are tested the same way as the ADF.

Based on further analysis of the findings reported in Ricketts and Jones (2017), at most 30% of change-points may show evidence of unit-root like behaviour, (59 of 218 ZA tests conducted on segments containing a change-point would have been initially classed as showing unit-root with drift ($p >= 0.05$)). This can be taken into account along with the empirically determined false positive and false negative rates by application of conditional probabilities, since the relevant probabilities depend on the data length. Thus for any particular test it is possible to determine the minimum segment length for which a false positive rate is less than 5%, and separately a length for which a false negative rate is less than 5%. Given an *a-priori* mix of the prevalence of stationary and non-stationary data, one can then determine whether a given result on a given segment is likely to be adequate. Note especially that a finding of stationarity may be accepted as sufficiently accurate whilst a finding of non-stationarity may be deemed unreliable, or vice-versa depending on the test. See Figure Ch4.16 and Table Ch4.4.

For example, consider a segment of length 45 if we assume a 50% *a-priori* likelihood of stationarity. KPSS test findings of stationarity would be accepted at a 5% level of significance, an ADF result would be deemed unreliable and the ZA test would be deemed reliable. On the other hand if the tests all returned non-stationarity/unit root, only the ZA test could be considered reliable.

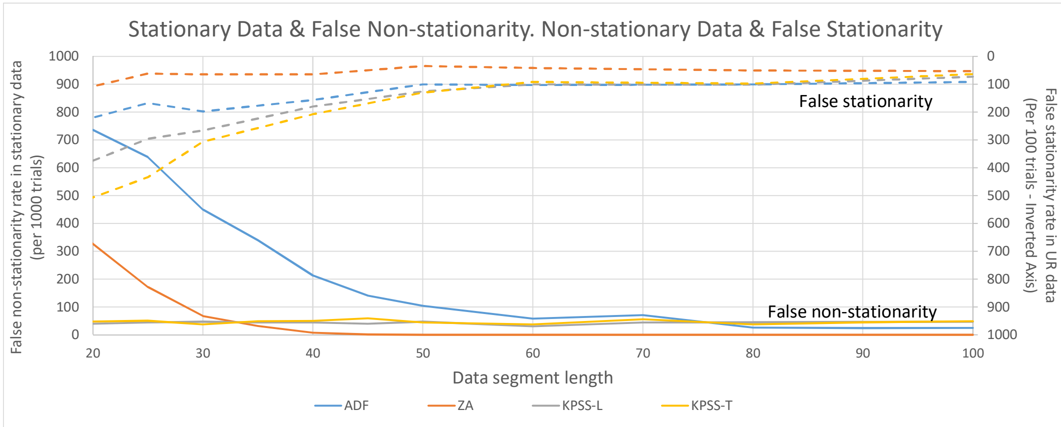


Figure Ch4.16: DS1: False stationarity and false non-stationarity rates (5th percentile) per 1000 shown as a function of segment length for each test. Solid lines represent false non-stationarity rates, referenced to the left axis, and dotted lines represent false stationarity rates referenced to the right axis (inverted).

Table Ch4.4: DS1: Minimum segment lengths to achieve false determination rates of less than 5%. Top, given equal a-priori likelihood of data being unit-root/difference stationary or deterministically stationary. Bottom, Given an assumed 70% a-priori rate of stationarity.

Given equal expectations of stationarity Test	**Segment length to achieve less than 5% false determination of Unit Root/non- stationarity**	**Segment length to achieve less than 5% false determination of deterministic stationarity**
KPSS-L	60	20
KPSS-T	60	20
ADF	50	50
ZA	30	30

Given prior expectation of 70% stationarity Test	**Segment length to achieve less than 5% false determination of Unit Root/non- stationarity**	**Segment length to achieve less than 5% false determination of deterministic stationarity**
KPSS-L	40	20
KPSS-T	50	20
ADF	40	60
ZA	20	30

The two KPSS tests show high rates of false non-stationarity, whereas the ADF shows high rates of false stationarity. As can be seen, based on this analysis, give equally likelihood of non-stationary or stationary data, the ZA test can be separate stationary and unit-root data with a 5% false determination rate with segment sizes of 30, which is highly desirable in this work.

Applying these UR tests

These tests are all applied to the segments of data within which a single change-point has already been provisionally identified. The change-point itself is not otherwise considered.

However, since the climate data being tested provisionally contains a deterministic change and only the ZA test is formulated with this as a ruling assumption, findings of non-stationarity may be caused by the presence of additional deterministic but undetected change-points.

The context here is that (a) an objective change-point method has been used subject to certain assumptions, and (b) a segment of data is delineated between two objectively determined change-points determined similarly. Do the assumptions of the detection method hold for the segment of data and to what extent? Additionally, the tests are only utilised when there is sufficient data to give a desirable level of confidence in their results.

Level stationarity is not just a zero trend. Data with a zero trend as determined by a regression analysis may be either deterministically or stochastically level. In the first case there may be a combination of zero or more change points that can be reliably detected, and the "noise" surrounding the signal is i.i.d. In the second case, a deterministic change-point detection method may return indeterminate change-points and the noise surrounding any extracted signal will not be i.i.d. but retain a UR characteristic. Similarly trend changes: Removal of the trend from continuous, trend stationary data yields level stationary data, as does removal of a shift and trend from discontinuous trend stationary data.

A segment of data with a valid shift point should not be level stationary, it should not be in a segment with unit root behaviour, and if it shows trending behaviour this should not be due to a drifting unit root. It should also have a significant result by ANCOVA. Table Ch4.5 summarises the use of these tests.

How data conditions relate to tests

Level stationarity. Data provisionally containing a step-like change is expected to not be level stationary. Non-stationarity is a property of deterministic trend, and/or deterministic step-change but also of unit root behaviour. The KPSS-L test is used here with an expectation that segments in which a change-point occurs contain a step-like shift but may also contain a change of trend. Hence it is used as a cross check, as the expectation is that no data segment within which a change-point exists will be found to be level stationary. Further, once the deterministic internal shift and trend components are removed the residual should be level and trend stationary. Level non-stationarity in the segment and level stationarity in the residuals supports the existence of a change-point.

Trend stationarity. Data with a provisional change of trend is expected to be non-trend-stationary. Data with constant trend and a step-like change may show as trend stationary depending on the assumptions of the specific test. The KPSS-T test and the ADF test as formulated here may return different results in the presence of a step-like shift and no trend change, with the ADF test showing trend stationarity and the KPSS possibly showing non-stationarity.

Unit root/non-stationarity in the absence of any deterministic change. The presence of a unit root may cause the data to mimic either a step-like change or a change of trend that in either case the MSBV can return as a step-like change. All four unit-root tests are expected to detect this, with the ADF being less powerful, partly due to a potential to overfit autocorrelation lags. Since the detection method has provisionally detected a change-point, tests on the residuals would likely all show non-stationarity, and similarly testing of the segment itself. The ZA test would likely be the most powerful.

Unit root/non-stationarity in the presence of deterministic change. This is a complex issue. The combination of UR and deterministic trend is potentially explosive (Hacker, 2010). On the other hand the climate system is physically bounded and so at worst the combination may appear as step-like. Data in which the tests support unit root in both the segment data and its residuals indicate either a genuine unit root presence or multiple deterministic changes; whilst data with apparent support for a unit root that disappears in the residuals, is consistent with a single deterministic change. However data with multiple change points is misspecified for all tests.

Evidence of misspecification due to data composition methods. Climate data is not homogenous. Data that represents the sum or average of multiple processes such as the GMST may have features which are detectable in individual processes but that remain below detectability thresholds when averaged. For step-like changes occurring at different times in different components, the steps are reduced in size by averaging. Further, if the changes differ slightly in time over a number of components then the deterministic shift-like changes may be confused with either stochastic or deterministic trend. Similarly trend changes: If the step-like or trend change happens simultaneously across all processes then the i.i.d. stochastic noise of each process will partially cancel out, but the signals will reinforce, thus enhancing the signal to noise ratio. A method that imposes or presumes smoothing may identify this as a trend. If autocorrelation is present as part of the signal in the component's data together with trend, the situation is still more complex.

Table Ch4.5: Unit root tests used and their main assumptions. Possibilities not formally considered may deceive these tests by supporting either the null or contrast hypotheses – this is noted in the later rows of the table.

Properties of Test	**ADF (trend and drift)**	**KPSS (level stationarity)**	**KPSS (trend stationarity)**	**ZA**
Ruling assumptions	No exogenous change.	No exogenous change.	No exogenous change.	Not combined exogenous change and unit root. At most one exogenous change (shift or trend change).
Null hypothesis, H0	I(1) Unit Root after allowing for autocorrelation and trend.	Stationarity	Trend stationarity	I(1) Unit Root with drift and no exogenous change
Contrast hypothesis, H1	Presumption of trend stationarity	I(0) Unit Root	I(1) Unit Root	Deterministic with possible exogenous change at a date
When at most a single exogenous change is present				
If exogenous change and UR present	Accept H0, i.e. UR	Prefer H1, i.e. UR	Prefer H1, i.e. UR	May prefer H1, i.e. exogenous change
If exogenous change but UR not present	May accept H0, i.e. UR	If constant trend and step-change then will accept H0, stationarity. Otherwise prefer H1, i.e. UR	If step-change only then will accept H0, trend stationarity. A strong trend change will prefer H1, i.e. UR	Prefer H1, i.e. exogenous change
Unit root but no exogenous change	Accept H0, i.e. UR	Prefer H1, i.e. UR	Prefer H1, i.e. UR	Accept H0, i.e. UR
When multiple exogenous changes are present				
Plus Unit Root	Accept H0, i.e. UR	Prefer H1, i.e. UR	Prefer H1, i.e. UR	Accept H0, i.e. UR
No Unit Root	May accept H0, i.e. UR	Prefer H1, i.e. UR	May prefer H1, i.e. UR if exogenous trend changes present	May accept H0, i.e. UR
After removal of all exogenous change				
If Unit Root	Accept H0, i.e. UR	Prefer H1, i.e. UR	Prefer H1, i.e. UR	Accept H0, i.e. UR
No Unit Root	Prefer H1, trend stationarity	Accept H0, stationarity (unless residual trend remains)	Accept H0, trend stationarity (unless residual trend remains)	Prefer H1, exogenous change (even if there is none)
After removal of main exogenous change but with exogenous change still present				
If Unit Root	Accept H0, i.e. UR	Prefer H1, i.e. UR	Prefer H1, i.e. UR	May prefer H1 if exactly one exogenous change remains. H0 of more than one.
No Unit Root	Prefer H1, trend stationarity (even if residual trend remains)	Accept H0, level stationarity	May prefer H1	Prefer H1, exogenous change

Therefore the finding of unit-root like behaviour in climate records must be interpreted with care, as it may also be evidence of misspecification of data composition. The only valid method of disambiguating the situation is to analyse the component data. This also follows from consideration of the properties of moving average and autoregressive time-series.

The "order" of an AR process is the number of lags, and also the polynomial order required to fit the error terms. The Dickey-Fuller equation (Equation 1) describes an autoregressive single lag, i.e. an AR(1) process. The sum of two AR(1) processes does not conform to an AR(1) process. In general it is most compactly represented as an autoregressive-moving average (ARMA) process of greater order, ARMA(2,1) (Granger and Morris, 1976).

If p and q are the lag order of processes, then two AR processes combine into an ARMA process, where the first parameter of the ARMA is the order of the AR part, and the second is the order of the moving average (MA) part.

$$AR(p) + AR(q) = ARMA(p + q, \max(p, q)) \tag{6}$$

Note that the for the sum of two AR(1) processes, the most compact representation requires three parameters, two different lags for the AR part and one for the MA part.

Treating the result of $AR(1) + AR(1) = ARMA(2,1)$ as an AR(1) process may be deceptive. And yet in many analyses, the issue of the composition of the data is at best brushed off, and autocorrelation is in general approximated as AR(1). In this thesis, for example, I show that apparent unit root-like behaviour in some zonal data sets resolves to deterministic shifts at different times in sub-sectors of those zones, and that this affects the determination of change-points.

Interpreting combinations of tests in the light of ruling assumptions

The UR tests are composed, like all statistical tests, with certain ruling assumptions that bound their applicability. In this work, the impact of specific departures from those assumptions has been considered (see Table Ch4.5), and included in inferences made about the data. KPSS and ADF tests are often framed as NHST tests that select between either implicitly deterministic stationarity of the data segment or purely stochastic (unit root) drift, the latter perhaps mimicking either a trend change or a step-like shift as far as detection methods are concerned. Importantly, the ruling assumptions of these tests include an assumption that no step-like or trend change exists, and the non-stationarity attributable to these features may be taken as presumptive evidence of UR. That is, an exogenous change (a change in structure) in the underlying physical system may result in a deterministic step or trend change in the time series, which may be misinterpreted as non-stationarity. In turn, the presence of these two statistical features, due to the presumption of their absence in the formulation of the tests in question, may be mistaken as evidence of a *stochastic* drift, rather than a deterministic change. This in turn may be taken as evidence *against* an exogenous or structural change in the system and evidence *for* an endogenous one (with the inference that the physical system is unchanged).

This is dealt with by (a) taking into account that a step-change has already been provisionally found, (b) testing the residual of the signal with the internal shift and trends removed, (c) interpreting all of the tests together. The various possibilities are summarised in Table Ch4.7. These tests are tests of the properties of the data, whereas the ZA test is framed around the change in the data and so the presence of a change-point does not conflict with the assumptions of the test.

The ZA test is intended to classify change-points as either endogenous or exogenous in origin. The test assumes a provisional change-point, due to either a unit root process with a stochastic drift (an endogenous change) or a single deterministic (exogenous) change. It removes the effect of this, and performs an ADF-like test on the residual. It is documented to return a finding of endogenous change if there are multiple exogenous change-points, and a finding of exogenous change if there is one exogenous change-point, even in the presence of unit root. Hence if it returns non-stationarity initially, but stationarity in the data, that is evidence of two change-points.

If a change-point is not supported by ANCOVA then this is evidence of components other than pure step-like processes, most notably ongoing deterministic or stochastic trend. However the test introduces extra degrees of freedom, and is less sensitive than the MYBT test in the absence of trend.

If the detection method returns a valid deterministic change-point and the resulting internal shifts and trends are removed, and the residual is tested by the same tests, the expected findings are as follows …

a. If the initial signal conforms to the assumptions of the MSBV or is otherwise stationary apart from step-like changes and/or trend changes:
    - In the presence of a single change-point (i.e. no other than the detected point exists), the KPSS tests and the ADF test will all return stationarity and the ZA will reject a UR, nominally favouring a deterministic change-point.

b. In the presence of a single additional, but undetected change-point, the KPSS and the ADF tests may or may not indicate stationarity, depending on the trend changes, but the ZA may be influenced by the second change-point as discussed above; and in the presence of more than one additional undetected change-point, the ZA detects non-stationarity, i.e. endogenous change in the residuals.
    - The ZA will detect changes of trend without an accompanying step change, thus the implication of additional but undetected change-points does not mean that step-changes have been missed.
c. If the signal contains a possible unit root, findings of UR in the residuals by all tests are signs that the data and the change-point detection are mismatched. However the following should be considered.
    - In the presence of a single step-like change imposed on a unit root sequence, the ZA on the initial data may show exogenous change with endogenous change (UR) in the residuals with other tests simply returning non-stationarity.
    - In the presence of two exogenous changes imposed on a UR progression, the ZA may show endogenous (UR) change with exogenous change in the residual and all other tests are likely to show non-stationarity. In this work, this would simply be considered unclassifiable. The detected change-point would be regarded as weakly supported.

Thus, it is possible to examine the data segment and its residual and to determine whether the apparent change-point is likely to be deterministic or if it is provisionally stochastic/non-stationarity. If it is provisionally stochastic then there may be undetected change-points or the data may contain a unit-root sequence. In either of the latter cases the detection may be suspect. It is also possible, for each of the diagnostic tests, to determine the power of the test, and thus whether the testing has adequately characterised the data (Figure Ch4.16).

By combining these results I produced an automatable classification scheme for the detected change-points, with five broad categories (Table Ch4.9). The scheme reflects possible misspecifications of the data and the detection of change-points characterised by abrupt shifts.

Table Ch4.6: Expected outcomes of the Zivot Andrews test, given data with a presumptive step-like change plus a variety of additional conditions. The first and second columns define results of the tests on the initial data segment and the residual with internal step and trend removed. The last column lists interpretations of the pairs of results. *These possibilities are discriminated on the basis of CP-Index

Initial data with a presumptive step change	Residual with internal step and trends removed	Interpretations
H_0 rejected, accept as Exogenous/Stationary	H_0 rejected, accept as Exogenous/Stationary	*There is a deterministic change with stationary residual.*
	H_0 not rejected, accept as Endogenous/Non stationary	*There is a deterministic change with non-stationary residual*
H_0 not rejected, accept as Endogenous/Non stationary	H_0 rejected, accept as Exogenous/Stationary	*Residual is non-stationary with two deterministic changes*
		Residual is stationary with two deterministic changes
	H_0 not rejected, accept as Endogenous/Non stationary	*Residual is non-stationary with zero exogenous changes: step-change is false positive**
		Residual is stationary apart from two or more undetected change-points
		Residual is non-stationary with more than two deterministic change-points

Table Ch4.7: Expected outcomes of the KPSS-T and ADF tests, given data with a presumptive step-like change plus a variety of different conditions.as per Table Ch4.6

Initial data with a presumptive step change	Residual with internal step and trends removed	Interpretations
KPSS-T H_0 not rejected accept as Stationary. *ADF* H_0 rejected accept as Stationary.	*KPSS-T* H_0 not rejected accept as Stationary. *ADF* H_0 rejected accept as Stationary.	*Residual is stationary, the single change-point did not have a trend change*
	KPSS-T H_0 rejected accept as Non stationary. *ADF* H_0 not rejected accept as Non stationary.	*Location of a single change-point misidentified so that the trend is also miscalculated*
KPSS-T H_0 rejected accept as Non stationary. *ADF* H_0 not rejected accept as Non stationary.	*KPSS-T* H_0 not rejected accept as Stationary. *ADF* H_0 rejected accept as Stationary.	*Residual is stationary and change-point included a trend change*
	KPSS-T H_0 rejected accept as Non stationary. *ADF* H_0 not rejected accept as Non stationary.	*The data segment is non-stationary and the provisional change-point may be a false positive.**
		Residual is non-stationary. The initial segment contained a step and/or trend change.

Classification of data segments

Classification is directed toward assessing reliability of the determination of change-point at a particular time by testing for the presence of undetected deterministic features, and/or non-deterministic behaviour. What is desired is a small set of classes that separately reflect the presence of these sources of confusion. A compact set of five classes, covering these requirements, is shown in Table Ch4.9. Classification takes place in two phases.

Firstly, the validation tests are treated as NHST tests at the 0.05 level. Based on the prior analysis of data segment length requirements (see Figure Ch4.16 and Table Ch4.4), results for which sufficient data exists are assigned values of "Stationary" or "Non-stationary", and other results are assigned "N/A", for not applicable. This means that for each test three results are possible. A classification index is produced as an intermediate step in the further categorisation of data segments. It is intended to compactly combine the KPSS-T, ADF and ZA results with the data length adequacy results. The index consists of a prefix representing the class of the data segment and a suffix representing the class of the residual. For example "26.26", would mean that the data segment was assigned an index of 26 and so was the residual. The values of the prefix and suffix are computed by summing values drawn from Table Ch4.8 for the three chosen tests in each case. They can then be written as a pair. For this example, the tests on the data segment were that KPSS-T, ADF and ZA tests all returned results of "stationary", similarly the residual. Then the index is computed using values from the bottom row; 2+6+18=26 and 2+6 +18=26 to give 26.26. The utility of this intermediate step is mainly that it simplifies further analysis.

Table Ch4.8: Values for the computation of a unique classification index of a segment containing a presumptive step-like change. The same scheme is applied to the initial data segment and to the residual once internal shifts and trends are removed. For each of KPSS-T, ADF and ZA tests, if there are sufficient data for the specific outcome of each test, the Pr value is interpreted as supporting either stationarity or not. Looking up the relevant interpretation (row) and test (column) yields a number for each test. The values are summed. When repeated for the results of the initial data and the residuals, this yields a binomial index number.

Interpretation of test	KPSS-T	ADF	ZA
N/A, Not adequate data	0	0	0
Non-stationary	1	3	9
Stationary	2	6	18

Secondly, remembering that the detection method has already provisionally determined a change-point, a small set of five classes was defined which reflect whether the data segment most likely contains a single change-point, or is accompanied by other points below detectability thresholds or perhaps does not really contain a change-point; and whether, once the change-point is accounted for, the residual data is stationary (which implies higher reliability, and supporting an inference of the presence of a physical regime change) or non-stationary, implying lower reliability (see Table Ch4.9).

There are two major groups of data segments identified here. Those in which the residuals are trend stationary, and those in which residuals are either non-stationary or not demonstrated to be stationary. The ZA test forms the principal basis for classifying these two groups, with the KPSS-T and ADF tests adding nuance where available.

Table Ch4.9: Classifications of data segments. Table A4.1.29, in the Appendix 4.1 gives the translation between index values and segment classifications.

Classification	**Reasoning and interpretation**
Single shift, stationary residuals	The ZA test detects the single shift in the data segment. When the residuals contain no change-point or unit root behaviour, the ZA test will also register a change point. Both KPSS tests for residuals are stationary.
Single shift, non-stationary residuals	The ZA test rejects the step in the data segment because of the presence of behaviour interpreted as unit-root. Removing the step in both the data segment and residuals results in a second rejection because of residual non-stationarity. The KPSS-T test is trend-stationary in both tests and with residuals, KPSS-L is stationary. This is consistent with a step surrounded by internal trends.
Single shift, N/A	The step-change detected by the MSBV is accepted without a valid ZA result because there is insufficient data to probe further (i.e., the segment is too short to provide a reliable result). The KPSS-T tests registers trends stationary in both trials and KPSS-L is stationary with residuals. This is consistent with a short segment of single-shift non-stationary data.
Multiple, stationary	There may be a pair of steps in the data. The ZA detects unit-root behaviour in the data segment, then a step in residuals. Both KPSS tests are stationary in residuals. This indicates the potential presence of a single additional undetected change-point.
Non-stationary	The ZA test detects unit-root behaviour in both the data segment and residuals. Short segments too brief for other tests prevent further insights. Multiple change-points on top of a non-stationary background is too complex a situation to detect with these tests.

Table Ch4.6 and Table Ch4.7 provide a reasoning framework for the use of KPSS-T, ADF, and ZA tests that takes into account that the tests were conducted on the data segment before and after accounting for the provisional change-point, and that the various tests differ in their false determination rates of stationarity and non-stationarity. A careful examination of the cases has been performed and a mapping between the index values above and the classes shown in Table Ch4.9 produced (also see Table A4.1.29 in Appendix 4.1).

Following the discussion on compositional misspecification (above), the two major groups can each be further subdivided into those that are best explained by a single change-point, those taht most likely have additional undetected change-points, ones for which non-stationarity so dominates that the change-point is cast in doubt, and ones from which no further results are available due to data length.

Four examples are expanded in here and summarised in Table Ch4.10.

A. For example, suppose testing of a segment of length 60 returns probabilities as follows. Pr(ANCOVA)=0.01, Pr(KPSS-T)=0.1, Pr(ADF)=0.01, PR(ZA)=0.01, and testing of the residuals returns the same values. The KPSS-T tests have not rejected the null hypothesis of stationarity, ADF tests have rejected the null hypothesis of UR in favour of stationarity, and the ZA test on the segment rejects the null hypothesis of UR in favour of a presumption of exogenous change (supporting the MSBV test). The ZA test on the residuals also rejects a UR, but because an exogenous change has been removed, and because testing has shown that the ZA test selects an exogenous change in the absence of UR, this indicates stationarity in the residuals. So the tests are interpreted thus: KPSS-T and ADF indicate stationarity and ZA indicates stationarity. This yields an index of 26.26 and compactly reflects a chain of reasoning as follows. The detection test (MSBV) has presumptively identified a step-change. KPSS-T and ADF on the segment show that it is trend stationary, and the ZA confirms that the change is exogenous. Tests on the residuals show stationarity. It would be assigned the class “Single,stationary”. Also ANCOVA supports that it is a change-point of some sort, and not likely to be simply a mid-point in a continued trend.
B. If the data segment length was 40, the ADF tests would have the same Pr values but would be ignored and treated as “N/A”. The index would be 20.20, but the class would still be assigned “Single,stationary”.
C. In another case, suppose the segment of length 60 returned Pr(ANCOVA)=0.01, Pr(KPSS-T)=0.01, Pr(ADF)=0.1, PR(ZA)=0.1, and testing of the residuals returns the same values. This time the KPSS-T tests have rejected stationarity in favour of a presumptive unit root, ADF has not rejected a UR in favour of stationarity, and the ZA prefers non-stationarity (formally, UR with possible drift). The index value would be 13.13. The chain of reasoning is that MSBV has detected a change of mean, ANCOVA indicates a presumptive change-

Table Ch4.10. Sample processing sequences for classifying a data segment. Test probabilities are converted into presumptive interpretations, the test result and length of data are used to decide from Table Ch4.4 whether the final result is adequate or should be replaced by "N/A". Index values are then selected from Table Ch4.8. The four sample cases ('A' to 'D') talked about above are summarised here. The two major groups of columns are the Initial data and the residuals and the tests applied to them are shown in the second row. The third row shows the Pr value returned by the test, the fourth shows the meaning of that result. The fifth, "Final Result" shows the consideration of data length being applied so that in B or example the ADF, although returning a finding of stationarity, it is ignored since the test would require a data length of 50 to minimise the false positive rate. The index components are the numerical value assigned to each test, drawn from Table Ch4.8. The bottom row shows the index value and the general class of the change-point (see Table A4.1.29).

A. Length 60	**Initial data**				**Residuals**		
	ANCOVA	KPSS-T	ADF	ZA	KPSS-T	ADF	ZA
Pr	0.01	0.10	0.01	0.01	0.10	0.01	0.01
Tests returns		Stat	Stat	Stat	Stat	Stat	Stat
Final Result		Stat	Stat	Stat	Stat	Stat	Stat
Index Components		2	6	18	2	6	18
Index and Classification	26.26	Single, Stationary					

B. Length 40	**Initial data**				**Residuals**		
	ANCOVA	KPSS-T	ADF	ZA	KPSS-T	ADF	ZA
Pr	0.01	0.10	0.01	0.01	0.10	0.01	0.01
Stationarity		Stat	Stat	Stat	Stat	Stat	Stat
Final Result		Stat	N/A	Stat	Stat	N/A	Stat
Index Components		2	0	18	2	0	18
Index and Classification	20.20	Single, Stationary					

C. Length 60	**Initial data**				**Residuals**		
	ANCOVA	KPSS-T	ADF	ZA	KPSS-T	ADF	ZA
Pr	0.01	0.01	0.10	0.10	0.01	0.10	0.10
Stationarity		Non-Stat	Non-Stat	Non-Stat	Non-Stat	Non-Stat	Non-Stat
Final Result		Non-Stat	Non-Stat	Non-Stat	Non-Stat	Non-Stat	Non-Stat
Index Components		1	3	9	1	3	9
Index and Classification	13.13	Non-stationary					

D. Length 28	**Initial data**				**Residuals**		
	ANCOVA	KPSS-T	ADF	ZA	KPSS-T	ADF	ZA
Pr	0.02	0.10	0.10	0.01	0.10	0.10	0.10
Stationarity		Stat	Stat	Stat	Stat	Stat	Non-Stat
Final Result		Stat	N/A	N/A	Stat	N/A	Non-Stat
Index Components		2	0	0	2	0	9
Index and Classification		Single, non-stationary					

point – but both tests are based on a presumption of stationarity. The KPSS-T and ADF tests on the segment might indicate a trend change or unit root, but if it were a trend change, the tests on the residuals would show stationarity and the ZA test would likely show an exogenous change. The result is consistent with red-drift or multiple processes, and the change-point would be classed as “Non-stationary”.

D. Shorter segments are less easy to analyse with diagnostics tests, and with decreasing data, more tests may be classed as “N/A”. Consider a segment of length 28, and data for which there is *a-priori* belief that 70% of all data has stationary residuals. The data segment returns the following test results. Pr(ANCOVA)=0.02 (and no evidence of a change of trend), Pr(KPSS-T)=0.10, Pr(ADF)=0.1, PR(ZA)=0.01. Testing of the residuals yields Pr(KPSS-T)=0.10, Pr(ADF)=0.1, PR(ZA)=0.1. However once the data length is taken into account we get: for the data segment, KPSS-T supporting stationarity but ADF and ZA being “N/A”. In the residuals, the only test that differs is the ZA which supports an endogenous drift. The latter ZA result supports non-stationarity in the residuals and a step-like shift in the data segment. The ZA test returned in the data was initially exogenous change but this was flagged as a possible false determination. The index is 2.11. There are several possibilities, but due to the short data length this could be classed as “Single, non-stationary”.

Synthetic data experiments

Specific data was used to assess and calibrate each of the tests as now described. The data, DS2, was constructed to test the effect of accelerating trends in combination with step changes of varying size, on average about 50% of these being below the nominal detection threshold. There is no UR behaviour in this second set.

Synthetic climate-like data (DS2)

A standard suite of artificial multi-step time series was constructed in order to assess the performance of the MSBV test with climate like data (hereafter DS2). Each consisted of four transforms. The first element of each of the five data sets (the ‘1’ series) was an artificial 200 year annual temperature set consisting of random data with lag 1 autocorrelation of 25%, lag 7 autocorrelation of 10% and a standard deviation of 0.44; centred about zero. A quadratic trend curve for each set is produced that rises to values between 2.1 and 3.6 degrees. For each of these there is a set of eight shift times and for each shift time, a random shift level. The average size of shift amounts to 1.5 standard deviations, (which is less than the bivariate test would be expected to reliably detect). The second element (the ‘2’ series) consisted of the ‘1’ series plus the associated trend changes, the third (‘3’ series) consisted of the initial data plus the associated shifts and the fourth ('4' series) consisted of the combination of initial data plus both shifts and trend changes, (see Table Ch4.11 which enumerates the construction, and Figure Ch4.17). The data sets are labelled "A" to "E.".

Table Ch4.11 Synthetic Data Timing and extent of Shifts. Total Rise is shown both as anomaly and as standard deviations. Shifts of < 0.5 are not guaranteed to be found by MSBV and are bolded. Note that the years are shown as first year shifted but during analysis we return the year prior.

A		B		C		D		E	
Year	Shifts	Year	Shifts	Year	Shifts	Year	Shifts	Year	Shifts
1955	0.57	1974	0.96	1970	0.80	1951	**0.42**	1955	**0.40**
1983	**0.34**	1975	0.97	1987	0.80	1980	0.83	1981	0.59
1999	0.72	2010	**0.46**	1996	0.68	2010	0.99	2001	0.80
2030	0.85	2027	0.79	2028	**0.41**	2018	0.77	2029	**0.38**
2036	1.00	2032	**0.43**	2036	0.57	2047	0.60	2039	0.59
2055	0.61	2055	1.00	2050	0.94	2058	0.60	2057	0.84
2071	0.94	2074	0.93	2076	0.54	2068	0.87	2071	**0.40**
2097	**0.31**	2085	**0.39**	2095	0.54	2085	**0.42**	2080	**0.42**
Total Rise (as K)	5.34		5.93		5.28		5.5		4.42

These data are constructed so as to provide a higher degree of difficulty than observational data sets, since shifts of less than 2 standard deviations are not easy to locate with precision especially at the ends of the data, there is autocorrelation and a quadratic deterministic trend.

A number of points of potential deception are present by design in this data. The last point in set A and C violate the seven year window rule in the MSBV. Bolded values represent shifts below the theoretical limit of reliable detection. The first two shifts in the 'B' set are a year apart which can certainly happen but cannot be separated. Shifts separated by ten years are detectable but analysis of the MSBV shows that such shifts may still be hard until they are close to three standard deviations (this issue is a domain limit to do with step-changes in general and not the MSBV alone).

It will be noted that this data contains many shifts that are below the nominal detection capability of the bivariate test.

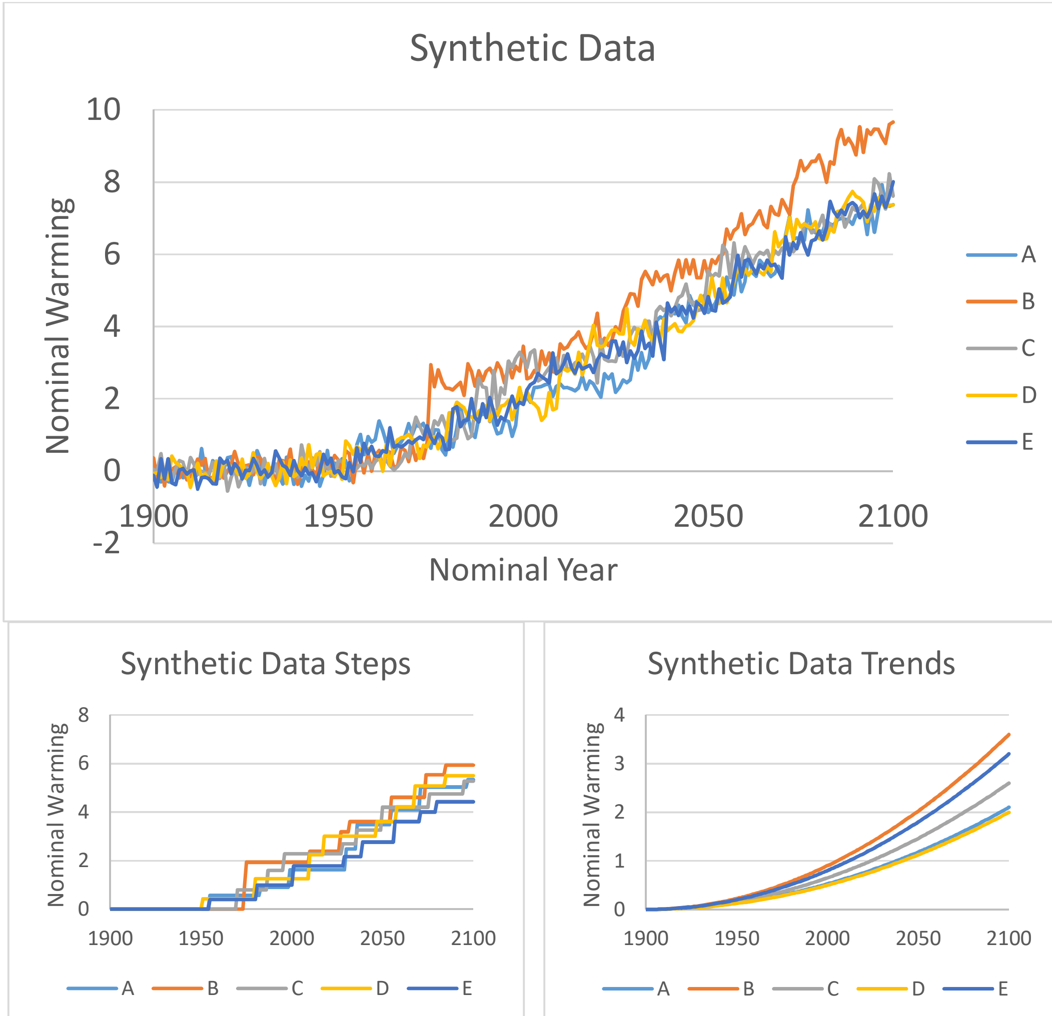


Figure Ch4.17: Synthetic dataset DS2, curvilinear trends and shifts. The top pane shows the sum of the corresponding series in the two lower ones, plus white noise.

Studentized Breusch-Pagan Test for Heteroscedasticity in DS2.

The data set DS2 was constructed with underlying curvilinear trends plus steps plus autocorrelation. The SBP test can be specified with any statistical model, and as explained, I used a linear trend, a quadratic and the disjoint set found by the MSBV test. It is intrinsic to the nature of the test that even formulated thus, it should not be used to discriminate these stochastic models.

Proposed interpretations of SBP tests given this data are: (a) if no change-points are present then any non-linearity must have a quadratic component shown by the quad model approaching homoscedasticity; and (b) if no nonlinearity is present, the linear model ought to approach homoscedasticity. (c)If the break model shows homoscedasticity then the break list is at least plausible, and if neither the quad nor the linear models are homoscedastic then the break model is more supported. If, however, the quad or linear models are homoscedastic then it remains possible that the variability due to change-points cannot be distinguished from random variation. One reason for this would be a large number of such change-points. In other words, the detection process has under-fitted the data.

In the below results (see Table A4.1.28) the '2' series (A2, B2 etc.) is consistent with the composition of random data plus curvature that does not contribute greatly to the variance. The '3' series is composed of collections of steps that will impose a non-normality on the variances and weaken the SBP test, and the'4' series contain curvature in addition to the expected steps. The results essentially support the use of a change-point model, although the possibility that 'C' data might also be simply a quadratic progression cannot be eliminated.

With this in mind, the more detailed examination of data using the tests in Table A4.1.27 can proceed on the assumption that change-point analysis was appropriate, but noting that dataset 'C' may also have an unconsidered auxiliary factor.

Results of summary analysis of DS2

Overall 78 change-points were nominally present in the combination of the '3' and '4' sets (the first two defined steps in the 'B' series are consecutive, and are treated as one). Of these: 8 change-points were potentially detectable but not detected by the MSBV test; 3 of those detected were misplaced or intermediate between closely spaced change-points; and 21 were smaller than the design threshold or within the prohibition period of another shift. These latter points are a potential source of confusion since they are present but not detected and add to noise (hence variance of error) in data segments.

The MSBV detected 50 change-points in total. Of these, one was classified as being "Single,non-stationary", three as "Single,N/A" and the rest as "Single, stationary".

Looking at the change in individual tests from the data segments to the residuals shown in Table Ch4.12, only two of 21 KPSS-T results that were initially non-stationary or N/A did not revert to stationary in residuals. ADF tests also reverted from non-stationarity to stationarity but not *vise-versa*, but the test has a higher data length for reliability of detection of either stationarity or non-stationarity than the KPSS-T test. ZA tests did not revert from stationarity (see also the segment classifications in Figure Ch4.18).

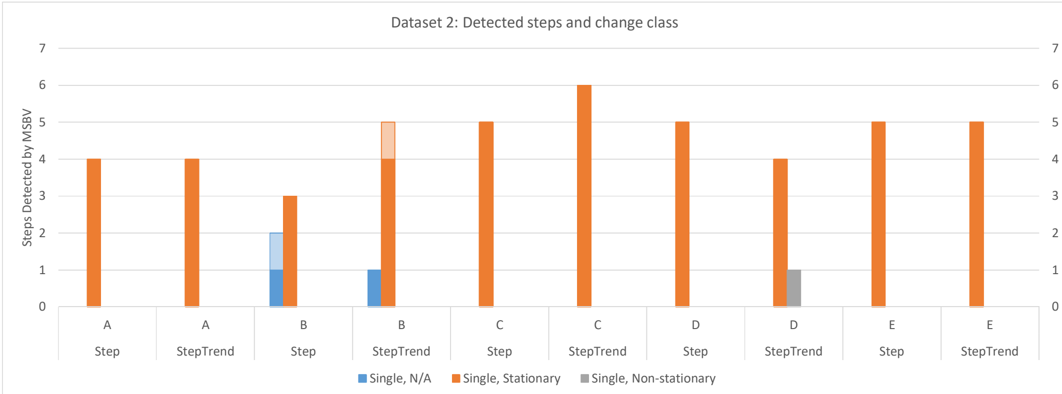


Figure Ch4.18: Data segment classification of change-points detected by MSBV in DS2, Non-saturated colours represent the points which were possible false positives (ANCOVA p>0.05)

These results are consistent with the initial data composition of the data sets, and the interpretations of the testing proposed. The data contain autocorrelation but do not contain unit roots, and the autocorrelation structure is simple. I do not find evidence of compositional misspecification in the data in the form of segments classed as "Multiple" or "Non-stationary". However the MSBV has found two shift-points that ANCOVA does not support, B4:1944 and B3:2081. The first was not defined and is a false positive. The latter is probably an intermediate location due to the presence of a sub-detectable shift in 2085 and classified as "Single,N/A". A corresponding point, B4:2084 is only weakly supported by ANCOVA (Pr = 0.04), is similarly classified, and considered a correct detection by MSBV. B4:1944, however is classified as "Single,Stationary".

The increased support by ANCOVA from B3:2081 to B4:2084 is consistent with a consequential trend change due to the additional quadratic trend in the '4' series.

Table Ch4.12: Classification of segments as stationary or not according to each of the tests. In all cases where a change from the initial data to the residual is seen it is stationary.

	Initial Data			Residuals		
Finding	**KPSS-T**	**ADF**	**ZA**	**KPSS-T**	**ADF**	**ZA**
Stationary	13	10	46	48	29	46
Non-Stationary	21	24	1	2	5	1
N/A	16	16	3	0	16	3

The degree of auto-correlation defined in these datasets does not cause a finding of non-stationarity in the residuals of change-points. This is as expected.

Illustrative tests of individual change-points detected in DS2

Three illustrative cases are now briefly discussed. In the following, a reference such as "B4:2073" refers to the first two columns of Table A4.1.27– data-set: change-time.

A3:2030 (Figure Ch4.19 top pane), is a defined change-point, correctly located by MSBV. However a false-positive change-point forms the early time-bound and three undetected step-changes are present. For interest a runs test and a Kolmogorov-Smirnov test were both performed to assess normality of residuals and both suggest non-normality. The ZA test suggests exogenous change in the data and in the residual and the KPSS-T and ADF revert from non-trend-stationarity in the data segment, to stationarity of the residuals. The ZA test is consistent with that interpretation recorded in the bottom row of Table Ch4.5. The change-point would be classified as "Single, stationary". The difference between stationarity testing and normality testing is to be noted – normality is not required for stationarity (Walpole et al., 1993).

B4:1944 (Figure Ch4.19 centre pane), is an MSBV false positive. Whilst it superficially appears as a level change and as a trend change, KPSS-T and ADF show trend stationarity in the data segment and residuals. It would be classified as "Single, stationary" using the scheme above but, ANCOVA does not support a change-point.

D4:2084 (Figure Ch4.19 bottom pane), is an expected date of small shift, and ANCOVA supports this as a change-point. The segment length is 34 which is less than would allow a conclusion for ADF tests (at this length the false stationarity and false non-stationary rates both exceed 5%), However the KPSS tests show trend stationarity in both the data and residuals while the ZA test shows endogenous change in both. This is in keeping with the presence of multiple undetected step-like changes. In this case a runs test did not support non-normality but a Kolmogorov-Smirnov test did. Its index 11,11 corresponds to a class "Single, non-stationary".

The ANOVA/ANCOVA tests are used here predicated on the assumption that the identified change-point is the most likely by some objective criterion and hence data on either side best represents the prior and posterior states of the climate had the climate shifted. Thus they have a role in the attribution of step and trend change. ANCOVA is a conservative means of assessing whether a change-point based on different criteria stands and hence rejection by this test must be considered with care.

These results suggest that the suite of tests characterises the change-point processes plausibly. The degree of autocorrelation applied is consistent with the findings of (Allen and Smith, 1994), and did not result in the incorrect classification of the data as non-stationary.

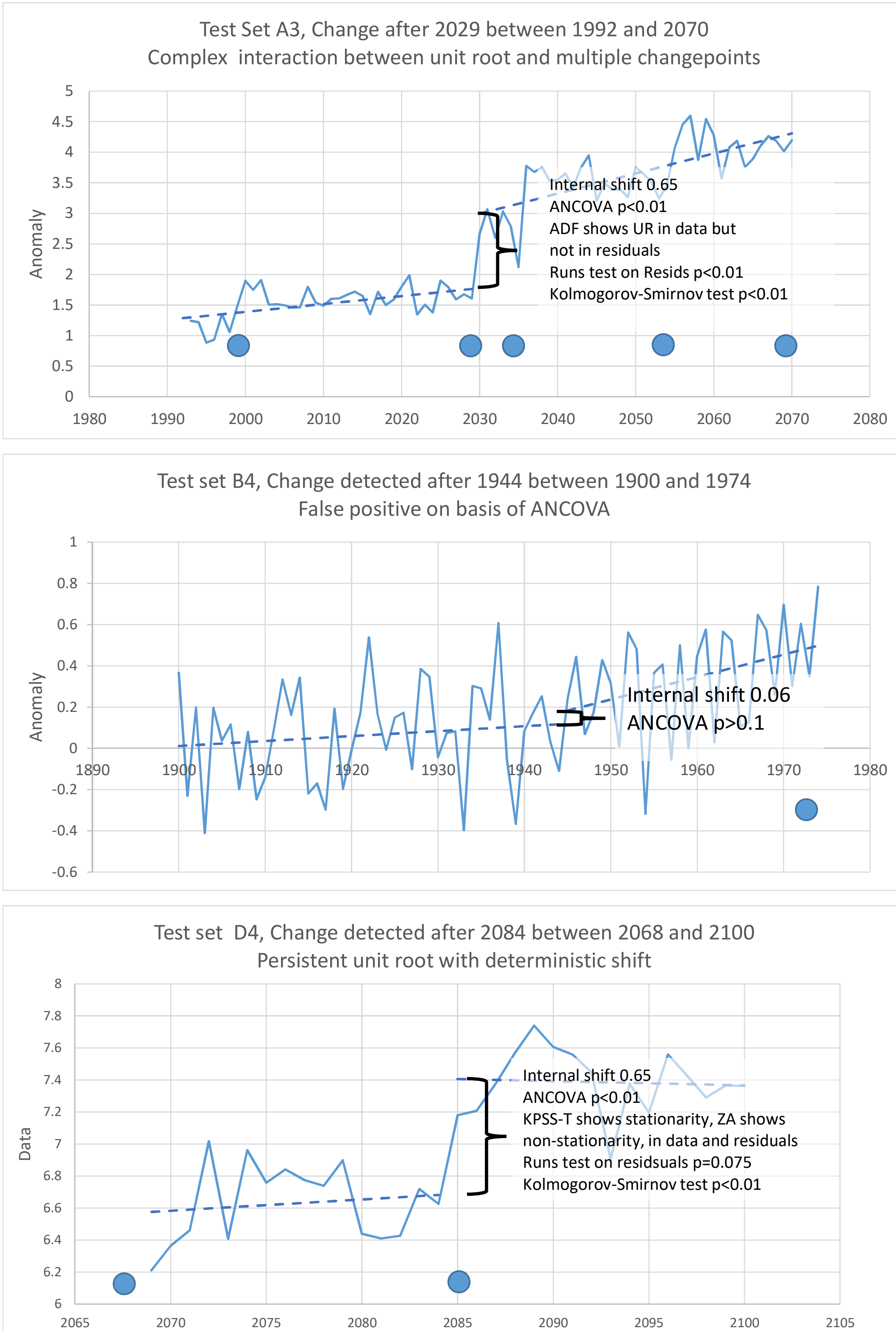


Figure Ch4.19: Illustrative analyse for dataset SD2. Blue dots represent expected change points. Top: A3:2029, illustrating data confounding the detection method. Middle: B4:1944. A false positive with ANCOVA rejecting the change-point. Bottom: D4:2084. A short sequence with a significant shift, correctly detected by MSBV but with apparent unit root behaviour as well.

The interaction of data composition and autocorrelation.

This section provides a basis for a major finding in Chapter 5, to wit that apparent non-stationarity in some zonal records is resolved by analysis at finer scale.

The role of autocorrelation in shaping climate signals was investigated by Allen and Smith (1994), and has recently been raised again (Beaulieu and Killick, 2018). When dealing with auto-correlated data Beaulieu and Killick (2018) applied pre-whitening (Rodionov, 2006b) but used more complex methods for computing ρ (the autocorrelation coefficient), whilst retaining an AR(1) structure. Importantly, they noted that autocorrelation is overestimated if there is an un-treated shift in the mean. In fact any deterministic structural break will have this effect. Since climate data is often obtained by some form of spatial composition, and autocorrelation is often assumed to be adequately represented as AR(1), there is a potential to misattribute to autocorrelation some of the signal which is due to either shifts or trend-changes. Consequently, attempts to remove such autocorrelation, based on a simple AR(1) model, may remove components of interest in the signal. As seen, the stationarity tests I have outlined are themselves not proof against these misattributions and, unless this is accounted for, may potentially register non-stationarity when in fact the signal has a deterministic shift. This is the key to my reasoning about the southern mid-latitudes in particular, seen towards the end of Chapter 5, and a prime motivator for the classification scheme.

In Ricketts and Jones (2017) we showed that zonal averages include more regional regimes, and that these cause segments containing change-points in the zonal averages to be taken as non-stationary. The following analysis relates the degree of autocorrelation actually injected into a signal to the degree of autocorrelation found after trend and shifts are added in and after the data is composited in a similar manner to that used in climatology, and then analysed for change-points

Composite signals were produced by averaging component signals where sometimes shifts occur as clusters at different times. Each component signal represents the signal from a particular location, and shifts are delayed by a series of two year intervals and then by an extra ten year interval. The components were averaged to simulate the production of a mean climate, and then the MSBV was run on the composite signal to test the limits of the MSBV under varying combinations of autocorrelation, shift and trend change (see Box Ch4.7, below). The validation suite was run over each component, over the composite, and the segments found by the MSBV. Empirical autocorrelation was assessed by linear regression of the signal and its lag. All combinations of: a step of 0 or 1.5 degrees; autocorrelation factors of 0, 0.33, 0.66, and 1.0; change of trend of 0, 1, 2,3, and 4 nominal degrees/century; were used to create composites. It would be expected that the MSBV would locate up three deterministic change-points in the composite signal due to shifts in the components. The sequence of five shifts in components at two year intervals after year 30 was expected to simulate a slow developing large shift. The shifts at 50 and 70 years of 1.5 standard deviations, each in one component of seven, was expected to be difficult for the MSBV as it is below the design thresholds.

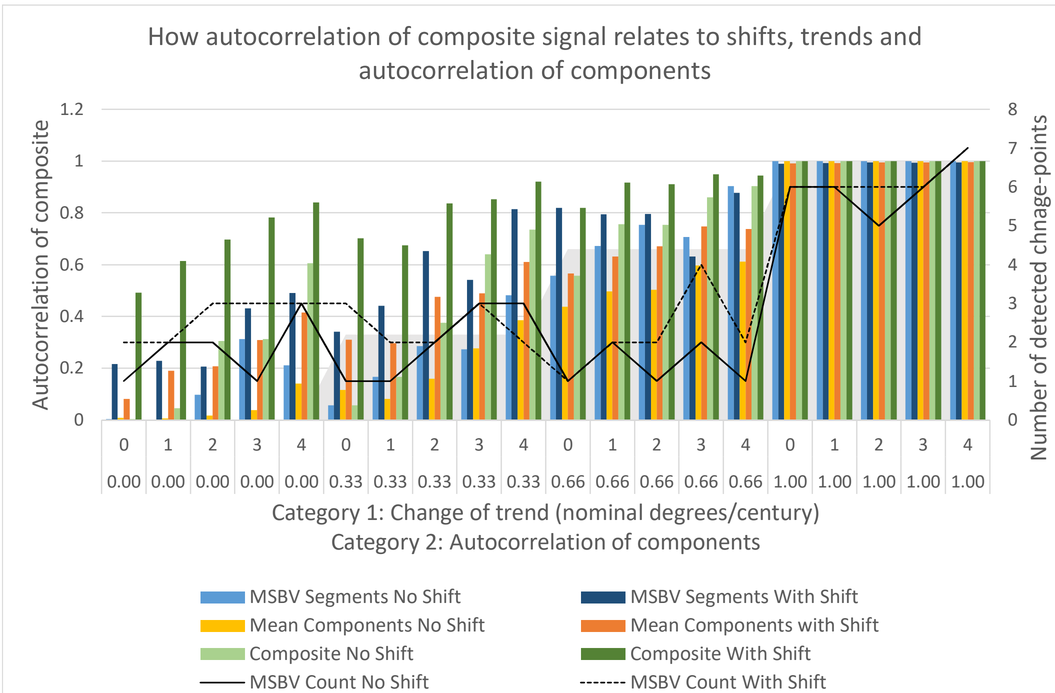


Figure Ch4.20. Autocorrelation in a composite signal and its components are compared. There are 4 levels of autocorrelation applied to components, indicated by the grey shading. There are five levels of trend-change indicated by the numbers on the x-axis (in nominal degrees/century). Dark colours indicate that a step change of 1.5 nominal degrees occurs in addition to a trend change lighter colours indicate trend change only. The numbers of change-points detected by MSBV in the composites is shown for the case of composites without shifts (solid line) and with added shifts (dotted line).

How apparent autocorrelation relates to composition, trends and shifts.

This part of the testing examines how the internally defined autocorrelation relates to each component of the composite signal. Since the composite is obtained by averaging, this is expected to reduce the variance attributable to random noise, and also to reduce the detectability of shifts since each shift occurs in only one component. Since the simple test for autocorrelation does not compensate for trend or shifts (in line with methods often used in climatology), it is expected that both trend and shifts may affect the empirical autocorrelation parameters. Figure Ch4.20 summarises the findings with respect to composition.

Pseudocode for data generation and testing

```
def genBreak(length=100, pos=30, shift=1.0,autocorrelation=0., trend1=0.0,
trend2=0.0):
#
#Generate sample data with specified shift, trends and autocorrelation
#
  data= np.random.rand(length) #random data
  if autocorrelation != 0.: #carry forward specified portion of signal
      for i in range(len(data)-1):
        data[i+1] += data[i] * autocorrelation
  std=np.std(data)
  data[pos:] += shift*std
  if trend1 != 0.:
    add in trend1 up to pos
  if trend2 != 0.:
    add in trend 2 after pos
return data

def generateTestData(length=100, pos, offsets, shift, autocorrelation, trend1,
trend2):
#
#Generate a bunch of components and average them
#Returns both the components and their average
#
  for offset in offsets[:]:
    data2,_=genBreak(length, pos+off, shift, autocorrelation,trend1, trend2)
  result.append(np.array(data2))
return data/len(offsets), Years, np.array(result)


firstBreak=30
for autocorrelation in [0., 0.33,0.66,1.]:
  for offsets in [[0, 2, 4, 6, 8, 20, 40] or other test cases]:
    for trend2 in [trend1, 1., 2., 3., 4]:
      generateTestData giving composites and components
      perform UR and autocorrelation tests on composite and components
      perform MSBV on composite
      for each break detected perform UR and autocorrelation tests on segment
```

Box Ch4.7: Pseudocode for the simulation of climate-like data averaging in the presence of multiple processes.

There are four sets of five clusters. The first five clusters of bars represent composites with no intrinsic autocorrelation but five levels of trend change. Then follow similar data with autocorrelations of 0.33, 0.66 and 1.0. The last is explicitly composed of random walks with and without an added shift, and the added shift seems to fully dominate the behaviour.

Examination of the first five groups of bars (all composed without intrinsic autocorrelation) shows that the degree of imputed autocorrelation in composites (without shift light green; with a shift, olive green) rises with the amount of trend change, but much less so in the absence of a shift. Comparison with the components (no shift, yellow; with shift, orange) shows that this is the case for the components as well, but since the composites show much greater imputed auto-correlation, the composition has itself had an effect. When the MSBV is used to break the signals into segments (no-shift, light blue, with shift, dark blue), the mean autocorrelation of the segments is intermediate, showing that the detection and removal of deterministic changes also affects imputed autocorrelation. As the intrinsic autocorrelation of the components increases the effect of shifts and trend changes is still present but less clear. Finally the solid and dotted lines indicate that not until the signals are highly dominated by red-noise is the number of change-points affected.

The main findings of this part are:

(a) Composite data sets with a shift register as having high autocorrelation even when the components do not (olive green bars in Figure Ch4.20).
(b) When no autocorrelation is present in component signals, it is still observed in composites (Figure Ch4.20, light and olive green) especially if there are shifts. This can be attributed to the lagged timing of changes.
(c) Increased autocorrelation in components leads to increased detected autocorrelation in the composite and in the segments delineated by the MSBV. In all cases adding shifts in increases the autocorrelation detected.
(d) Applying the MSBV produces segments that in general have more apparent autocorrelation than the components but less than the composite. This is consistent with the composite signal having sequences of steps that the MSBV cannot separate.

Case study, analysis of previously published data

Five mean annual surface temperature datasets were analysed for date and extent of internal shift, and trend at global, hemispheric and zonal scale, and the results published by Jones and Ricketts (2017b), and will be referred to as Obs2017. These change-points have now been further characterised by the tests proposed in this chapter. All datasets provided estimates of global averaged annual temperatures. Three sets (NCDC, HadCRUT4, and GISSTEMP3), provided estimates of hemispheric averaged annual temperatures. Two sets, (NCDC, and GISSTEMP3) included estimates of annual zonal average temperatures, although these used different latitude bounds. Both also include estimates of Tropics and composites that overlap zones. Two datasets, (Cowtan and Way, a.k.a. Had4-Kriged, and Berkeley) contained only estimates of global annually averaged temperatures. HadCRUT4 also provides estimates for the zone 30S-30N which overlaps the tropics. The NCDC dataset also included separate estimates of mean land and mean ocean temperatures.

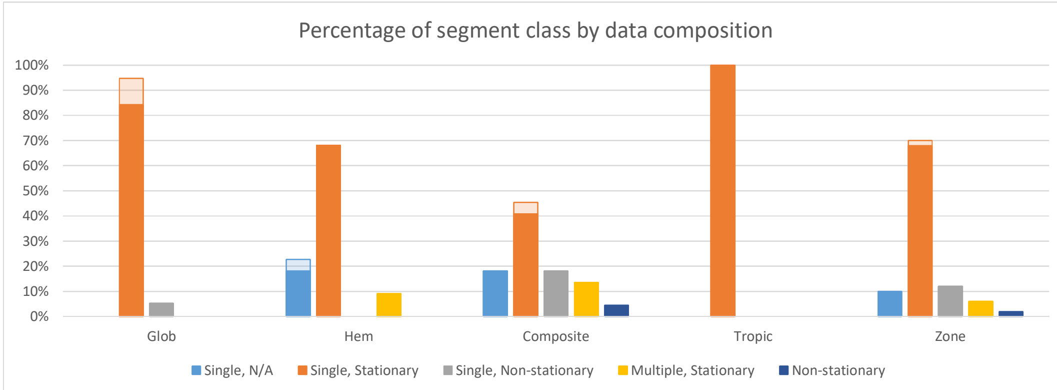


Figure Ch4.21: Classification of the land-ocean data-points published by Jones and Ricketts (2017b) from five separately sourced estimates of mean annual temperatures at global and sub-global scale. Unsaturated colours represent change-points which may be false positives on the basis of a CP-index.

This gave 218 change-points detected by MSBV of which all but 9 would be accepted on the basis of a CP-index (28 meeting the assumptions of the MSBV but ANCOVA $p>0.05$).

A major point of interest is that clearly, the composition of the data has an impact on the class that change-points are assigned to. Tropical change-points are always "Single-stationary", global change-points are mostly "Single-stationary"; whereas zonal, composite and hemispheric change-points may include indicators of compositional misspecification: evidence of undetected change-points and non-stationary residuals.

Summary and conclusions

I have outlined a program of assessing presumptive shift dates in climate data. In doing so I have attempted specify testing that fits within a formal framework of severe testing (Mayo and Spanos, 2006). I have also strongly leaned on the idea of model misspecification (Mayo and Spanos, 2004). It is telling that the MSBV, based on maximized likelihood of step changes rather than minimised residual in a segmented model, none the less compares favourably to other methods such as structural-change. The tests outlined here assist a probative analysis, firstly by adding nuance to the findings, and secondly by providing the basis of a change-point classification, they assist strong reasoning. They have been selected because they are individually automatable and complementary, and the utility of this has been indicated in the case study. The chain of reasoning involved in the use of multiple tests is complex but the final classification scheme is compact and as seen, informative.

The tests for heteroscedasticity are conventional. Homoscedasticity of residuals is, however, generally considered insufficient evidence for the adequacy of a statistical model. It is, none the less, an assumption of OLS regression. Hence it is possible that if variance changes as part of a regime change (see Killick et al., 2010), heteroscedasticity would be expected. The use of three separate statistical models over the whole of a time-series allows one to determine whether the data suggest that a step-and trend model is adequate. Homoscedasticity around a linear trend would suggest a higher order model over fitted. Around a quadratic trend, it suggests that the results of a change-point analysis be considered carefully since fitting a piecewise linear model to trend dominated data is fraught (and especially when the trend is continuously varying) due to there being no preferred starting change point (Jarušková, 1997, Jarušková, 1996). This is part of a model misspecification test. Applying the same test to the derived disjoint segmented model deduced by change-point analysis is a sensible cross-check. It is not a test for optimal parameterisation, and in fact this is likely to be harder to determine in climate data than generally recognised.

Using unit root/stationarity tests to locate an important potentially confounding condition in the data – transient or predominating unit root – likewise assists with building a more complete picture of events. The meaning of transient unit roots in climate data is an open question, the more so since the methodologies are econometric and the economic use is very different. Here, I use the methods to firstly identify violations in the data of basic assumptions made in model selection and regression. Transience is a form of non-stationarity against which very few analytic methods in climatology have much defence, and yet it is of substantial interest. The MSBV has shown itself robust to transient temperature fluctuations (e.g. the El Chichón eruption). Transience of this sort may have either a deterministic or a stochastic origin, it may be signal or noise.

Beaulieu and Killick (2018) address some of these issues from the point of view of autocorrelation (or system memory), but their paper was published too late for a proper incorporation into this thesis. The paper proceeds by producing eight statistical models with combinations of autocorrelation (or not) and the presence or absence of either shifts or trend changes, and performing statistical model selection based on information criteria to select preferred statistical models – however they do not include one with both shifts and trend changes, despite having cited authors who do (Seidel and Lanzante, 2004), and argue against the circa 1996 change, the so-called hiatus. This chapter demonstrates that autocorrelation can be introduced into climate data by virtue of the data preparation. Later chapters use this to assist with finding a "natural scale" for regional regimes.

However even more importantly, it is the combination of these tests and their various contrast hypotheses that enables one to be sure that where a step-like change-point is provisionally detected at a particular time, it (a) provides a better explanation of the linear underlying processes than its absence (ANCOVA), (b) the underlying stationarity assumptions are not violated (ADF and KPSS), and (c) is unlikely to be due to random drift, fast or slow, (ZA). The window and shuffle tests in JR2017 address similar issues to the ZA.

The ANCOVA approach is quite conventional and serves in attribution of change to two types of change, step-like (or level change) and trend-change (or innovation as it is sometimes called (Tsay, 1988)). One should note that rejection of a change-point by the test is a conservative result given that some detection method has presumably established it to a high confidence. The CP-index allows a researcher to apply less conservative, but by no means less stringent rules to testing false positives. Accepting points that are ANCOVA $p>0.05$ but which meet the ruling assumptions of the MYBT – no underlying trend (i.e. CP-index of zero) is acceptable, the more so after non-stationarity is eliminated.

The various test all differ in their propensity to produce false positive or false negative results and this has been addressed, initially by sensitivity testing for false positives and false negatives separately and then because the KPSS and ADF tests invert the null hypotheses with respect to each other, synthesising these into stationarity/non-stationarity determination rates given resumed prior rates of stationarity. This might seem over complex, but it minimises unsupported presumption, in order to get maximum value from small data sets. The end result is that a taxonomy or typology of data segments/change-points is available that can be used to separate change-points into groups reflecting the presence of a previously unconsidered source of deception, red-noise or unit-root behaviour.

A basis has been established for potentially detecting signatures of a data composition mis-specification whereby features emerge or submerge in composited data due to averaging signals (especially ones moving in time and space). The signature is a reduction in evidence of non-stationarity when signals are decomposed or segmented using the MSBV.

This typology has been demonstrated in sample data (DS2), and has shown that data composition is a source of variability in type and a source of apparent autocorrelation. It has demonstrated that the MSBV is fit for use, although with the caveat that post hoc testing should be performed. The case study supports previous work published using these data, but also underlines the complexity of the subject. The typology is used in Chapter 5 to demonstrate that land and ocean respond differently to regime changes, and function at different spatial scales.

Chapter 5: Abrupt decadal shifts in observed and modelled mean annual zonal land and ocean temperature records.

Introduction

Recent work has shown that abrupt regime changes in Earth's climate either follow or include step-like pathways (Jones and Ricketts, 2017b, Bartsev et al., 2017, Yan et al., 2016, Yan et al., 2015, Reid et al., 2015). And that while these step-like changes can be detected in global averages signals they likely originate in decadal and regional processes (Rodionov and Overland, 2005, Overland et al., 2008, Reid et al., 2015, McCarthy et al., 2015, Alheit et al., 2005, Freitas et al., 2015, Trenberth, 2015, Trenberth and Fasullo, 2013).

Two findings in JR2017 were that three major events post World War 2 have affected both the global temperature progressions and the biosphere simultaneously, these being circa 1976, 1986 and 1997. In each case a major ocean basin has been involved.

We have also published a conference paper (Ricketts and Jones, 2017) (henceforth RJ2017) which examines in detail the step-like structure in ocean temperature records in the zone 60S-30S. Three findings in the paper are of note. (1) The Southern mid-latitude zone is very far from homogeneous. (2) But a sector scale of 45 degrees by 30 degrees approximates a natural scale of self-organisation in this region. (3) Compositing of temperature records at greater scale creates much of the reported redness, and obscures the detail of decadal scale processes.

This chapter utilises the multistep bivariate test (MSBV) and the validation suite from the previous two chapters. It extends aspects of the analysis of observed global and zonal annual temperature records, and the model based projected global annual temperature records published previously (JR2017). Included in RJ2017 was an examination in detail of the step-like structure in ocean temperature records in the zone 60S-30S, the Southern mid-latitudes. A chain of reasoning based on the validation suite showed that sub-dividing the zone into 45 degree sectors and comparing the results with the zonal record itself gave evidence that the temperature response structure of the zone is very far from homogeneous – with the area south of the Indian Ocean showing more shifts than the south of the South Pacific. Separate sectors had differing change-point records from each other and the zonal average. Accounting for these increased the homogeneity if residuals for sectors compared to the mean zonal record. This is evidence of a form of misspecification where an assumption made during aggregation of data creates deceptive features or hides actual features – a “data composition misspecification” (see Chapter 4).

This paper extends the analysis used in RJ2017 to the remaining zonal records of observed climate; and also extends JR2017 to zonal analyses of a large subset of the temperature records from the global climate models submitted for the IPCC fifth climate assessment report (AR5), as part of the Climate Model Inter-comparison Project Five (CMIP5). The validation suite is used to provide more nuanced analysis than has hitherto been possible. Equally this chapter also serves to allow a more detailed introduction to the validation methods themselves.

Three major questions are addressed below.

1. Does the composition of area averaged climate data bias results? RJ2017 suggests that abrupt decadal scale variability is a function, at least in part, of persistent but regional structures with state-like transitions. Further, as also shown in Chapter 4, it is the data composition misspecification by simple area averaging that leads to artefacts such as excessive autocorrelation or variation of autocorrelation, failure of homogeneity tests and even identification of data as showing a unit-root (red-noise) behaviour. If this is so, then a decomposition to regions smaller than the regional structures would be expected to give data in which the various artefacts were reduced. This is in fact what is seen, suggesting that amongst other things few conclusions about climate regimes can be drawn from global temperatures alone, due to averaging of the effects of large scale regionalised regimes.
2. Does the analysis of climate models at zonal scale support the presence of regime shifts? As seen below, yes it does. JR2017, and to a lesser extent RJ2017 had also shown that three well documented post WWII bio-physical regime shifts circa 1976, 1986 and 1997 were signalled as step-like regime shifts in temperature records at various spatial scales e.g. Jones and Ricketts (2017b, Figure 2). JR2017 had also shown that analysis of the ensembles of global climate models, covering the observed period, analysed by MSBV had clustering of step-change data that aligned with those obtained from observations. It had also shown broad clustering in the annual averaged CMIP5 projection runs (Jones and Ricketts, 2017b, Figure 7). If the conclusion implied in

RJ2017 of data composition misspecification stands (reduction of non-stationarity with reduction of scale), it can be expected, and is in fact shown, that the broad clustering will be "sharpened" by finer scale aggregation, and that this will also show in the validation tests.

3. What can global climate models tell us about the hypothesis H2 of JR2017 – of interaction between climate variability and forced warming? Climate models are the only tool by which the relationship of these step-like changes and forced warming can be explored into the pre-industrial/pre-instrumental past or into the future. Since H2 suggests interaction between natural variability and warming, a detection and attribution study can be performed. Three aspects of regimes can be tested: zonality, duration of regimes, and intensity of shifts. Increases in intensity above a detectability threshold will show as an emergence of an apparently novel sequence of events. If this latter condition occurs then it may well appear as a change of zonality in this type of analysis. H2 is supported by changes in duration, or zonality of regimes, and most likely both the size of shifts and of the associated trend changes.

As an example of the possible uses to which this approach may be put, special attention is paid to the first clustering after the end of the model switches from historical to model forcing, which suggests that the recent (at the time of writing) heat release event marks a new regime of higher temperatures and possibly faster warming.

The rest of this chapter is structured as follows. Data sources are provided. Following this the general methods and three analyses are described. These are (a) an extension of the previously published zonal analyses, based on supplied GISTEMP3 zonal land/ocean averages (JR2017, RJ2017), to separate land, ocean and land/ocean averages reconstructed from GISSTEPM3 gridded data. This is compared to the supplied NCDC land, ocean and land/ocean analysed in the case example of Chapter 4 (Validation). This establishes a baseline for the further use of this gridded dataset. (b) The extension of the sectoring approach foreshadowed in RJ2017, to the above data, again using the validation suite to further probe compositional misspecification with the validation tests. (c) A zonal analysis of a selection of global climate models forced by the lowest and highest responding future greenhouse gas emissions scenarios (RCP2.6 and RCP8.5). Finally the results are summarised and discussed. Most tables are moved to an appendix to the chapter, Appendix 5.1.

Data sources

The following data was used in this chapter, as detailed in Appendix 1.1.

- NCDC zonal data version v3.5.4.201504. Anomalies are based on a 1971-2000 mean. Combined land and ocean area averaged times series were provided for 60S.30S, 30S.00N, 00N.30N, 30N.60N, 60N.90N, 90S.00N (S hemisphere), 00N.90N (N hemisphere), 90S.90N (global), 20S.20N, 90S.20S, 20N.90N, 60S.60N. The zone 90S.60S was deemed not adequate for use.
- GISTEMP3 gridded land-ocean anomaly data Anomalies are based on a1951-1980 mean on a 2°x2° grid.
- Global climate model data. Gridded monthly surface temperature data from a selection of global climate models was analysed similarly to GISTEMP3 by combining matched historical (1850-2005) and RCP2.6 and RCP8.5 runs to (2006-2100). Data were used at the supplied resolution and area averaged to produce representations of the same zones as GISTEMP3. They were also further subdivided into land and ocean using the model land/ocean masks supplied by each modelling group. Not all models had data for all of the pre-industrial controls, RCP2.6 and RCP8.5, but all available realisations with the RIP code "r1i1p1" for each were used (see Table Ch5.13). The meaning of the term "RIP code" is explained in section "Global Climate Models.on p224.
- Land Ocean Masks The supplied WOA09 1°x1° data was regridded to 2°x2° to match the GISTEMP3 data using CDO operators.
- Models are treated here as independent samplings of possible climates, and four atmospheric prescriptions are applied within each model. Models are complex and initialisation requires a period of "spin-up" (running against a stable atmosphere) before data collection commences. The set of runs chosen here are related within each model as follows. Following spin-up, the model is run with a prescribed pre-industrial atmosphere representing a hypothetical stable atmosphere unperturbed by human activity – the piControl. The model is then seamlessly switched to run with a prescribed atmosphere that varies with time representing the observed atmosphere between 1850 and the end of 2005 – the historical run. Then the model is seamlessly switched to one of a series of hypothetical future atmospheres and run out to at least 2100. Here I have selected two extreme prescriptions, RCP2.6 and RCP8.5. The first represents a world in which greenhouse warming is controlled to the point that warming ceases by 2050, with a net increase of 2.6 W/m^2; the second represents a world where greenhouse warming follows a business as usual trajectory so that at 2100 the net increase in warming is $8.5W/m^2$. Thus with piControls, historical, and

RCP2.6 and RCP8.5 runs there are estimates of a purely hypothetical stable past with no greenhouse forcing, a representation of how the model evolves given observed atmosphere and two hypothetical futures.

Table Ch5.13: CMIP5 Global Climate Models selected for analysis

CMIP5 Model	Pre-Industrial Controls	RCP2.6	RCP8.5
ACCESS1-0	X		X
ACCESS1-3	X		X
bcc-csm1-1	X	X	X
bcc-csm1-1-m		X	X
BNU-ESM	X	X	X
CanESM2	X	X	X
CCSM4		X	X
CESM1-CAM5			X
CNRM-CM5	X	X	X
CSIRO-Mk3-6-0		X	X
EC-EARTH			X
FGOALS-g2	X	X	X
FGOALS-s2	X	X	X
GFDL-CM3	X	X	X
GFDL-ESM2G	X	X	X
GFDL-ESM2M		X	
GISS-E2-H	X	X	X
GISS-E2-R	X	X	X
HadGEM2-ES	X	X	X
IPSL-CM5A-LR	X	X	X
IPSL-CM5A-MR	X	X	X
IPSL-CM5B-LR	X		X
MIROC5	X	X	X
MIROC-ESM	X	X	
MIROC-ESM-CHEM	X	X	X
MPI-ESM-LR	X	X	X
MPI-ESM-MR	X	X	X
MRI-CGCM3		X	X
NorESM1-M	X	X	X
NorESM1-ME	X	X	X

Results

Data preparation

In order to provide continuity of data under different compositions, the GISTEMP3 gridded data and global climate model data were averaged out zonally by area weighted averaging using the same zones as were used for NCDC zonal data. For climate model data, land/ocean masks supplied with climate models were used, and a land/ocean mask derived from the WOA09 ocean map was used in conjunction with GISTEMP3 data.

A trial run of the MSBV was used to produce change-point dates for these zones, and the results were cross checked against the results for the NCDC zonal data previously analysed in JR2017 to ensure consistency.

Data for a sectorial analysis matching the above was performed by utilising the GISTEMP3 2°x2° gridded data, matching the NCDC zones and extracting area weighted sectors where each sector covered 45 degrees of longitude within the zones. The grid is chosen is shown in Figure Ch5.22 and matches the sectors used to analyses the Southern mid-latitudes (SML) in (Ricketts and Jones, 2017). It was chosen to select as many ocean only sectors as possible in the SML zone).

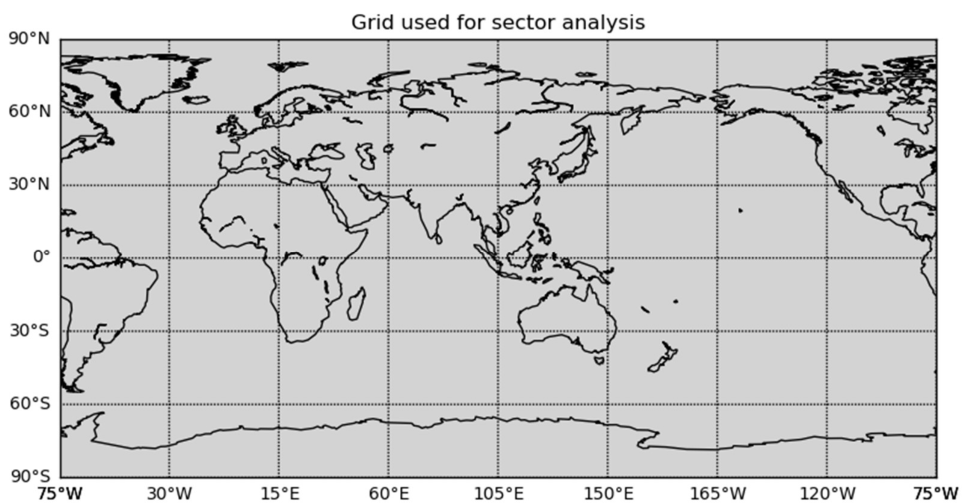


Figure Ch5.22: Grid used for sector analysis.

Autocorrelation

The Dickey-Fuller equation shown in Chapter 4 as

$$Y_t = \mu + \beta t + \rho Y_{t-1} + e_t \qquad (1)$$

was adapted to assess separate the contribution of trend and stochastic autocorrelation in segments containing a change-point, and the residuals once internal trend and shifts are removed. b represents the trend and c is the AR(1) autocorrelation coefficient.

$$Y_t = a + bt + c(Y_t - Y_{t-1}) \qquad (2)$$

Section 1: Zones in observational data

The primary purpose of this analysis is to extend, and add diagnostics to, the previous results included in JR2017, and lay a comparative basis for the analysis by sector in the next section. Previously published results for zonally averaged data obtained from NCDC are compared to the same zonal averages for the GISTEMP3 data. Since the zones provided for GISTEMP3 differ, the GISTEMP3 gridded dataset was processed to provide identical zones. It was area averaged over the same zones as the NCDC data reported in JR2017 (henceforth NCDC-z) using the 2°x2° Land Ocean Mask to provide land, ocean and combined land-ocean splits, a total of 36 time series with the Zone 90S.60S omitted (GISS-g). The same dataset was regridded to 5°x5°, and zonally averaged as before (GISS-g*). The dates for GISS-g* were compared to those of GISS-g. Where comparable, the GISTEMP land/ocean data reported in JR2017 (henceforth GISS-z) are also compared with the other datasets. All were analysed using the MSBV to find change-points and then the validation suite applied to the results. Differences between diagnostics for land and ocean changes are noted. There are five 30 degree zones (90S.60S omitted as previously noted) and seven zones which are various overlapped composites including global and hemispheric averages.

Detection of events

Data from different providers is pre-processed by the provider preferred methods of choice. This includes infilling of missing data, homogenisation of observations where instrumental effects are suspected, issues arising from the conventions that air temperatures are sampled at 2m above ground, but ocean temperatures are samples at the surface.

The effect of data preparation can be explored to a limited extent by inter-comparison of the datasets. Dates as well as the change-point types, and the multi-test index can be sensitive to smoothing and homogenisation of data.

If composition is a significant source of deception this can be judged by comparison of the zonal data with composites, and with sectors. Compositional misspecification will show as a greater tendency in the composites of apparent non-stationarity once change-points are accounted for. Similarly the composite of land-ocean may have an admixture of the characteristics and timings of step-like changes relating to land or to ocean. Similarly sub-zonal sectors are expected to be simpler and more stationary.

Change-point dates for the globe, hemisphere and five zones, aggregated by decade are shown below in Table Ch5.14. The detailed analysis using the validation suite (unit root and ANCOVA tests etc.) is provided in Table A5.1.32 in Appendix 5.1 with the most pertinent findings summarised in Figure Ch5.23,below.

Table Ch5.14: Change dates computed from NCDC zonal data as per JR2017 (labelled NCDC-z) are compared to GISTEMP3 gridded data re-aggregated into zones (GISS-g), and where comparable, GISTEMP3 zonally averaged data as provided (GISS-z). Year of change is shown (add one for first year of regime) for global, hemispheric and the five analysed zones. Land is shown in orange, combined land/ocean in green and ocean in blue. GISS-g and GISS-z can only be compared for global and hemispheric land/ocean data. Superscripts of 1,2, or 3 mean that for that analysis the ANCOVA test does not reach statistical significance. 1 means that the slope of the segment prior to the change is significant at 5% and the posterior part is not. 2 is vise-versa. 3 means that the trends are both statistically significant, which combined with ANCOVA is evidence of a false positive. Values of 1 or 2 indicate weak support for a change-point. GISS-g is shown only where the year differs from GISS-g*

Dataset	Zone	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
	Global									
GISS-g	90S.90N	1920					1979		1997[2]	
NCDC-z		1924					1979		1996	
GISS-g	90S.90N		1936				1978		1996[3]	
GISS-z		(1919)	1936				1979		1996	
NCDC-z		1929					1978		1996	
GISS-g	90S.90N	(1913)	1936				1976		1996[3]	
NCDC-z		1929					1976	1986	1996	
	Northern Hemisphere									
GISS-g	00N.90N	1920						1986	1997	
NCDC-z		1921					1979		1996	
GISS-g	00N.90N	1924						1986	1996[2]	
GISS-z		1923						1986	1996	
NCDC-z		1924						1986	1996	
GISS-g	00N.90N	1925						1986		2000[2]
GISS-g*		**1920*	**1935*					**1986*		*2000[2]*
NCDC-z		1925						1986	1996	
	Northern Hemisphere Zones									
GISS-g	60N.90N	1919							1994	
GISS-g*		**1922*							**1994*	
NCDC-z		1919						1987		
GISS-g	60N.90N	1919							1994	2004
NCDC-z		1919						1987		2001
GISS-g	60N.90N	1919							1994	2004
NCDC-z		1926								2000
GISS-g	30N.60N	1920						1985	1996	
NCDC-z		1920					1980		1996	
GISS-g	30N.60N	1920						1987	1997	
NCDC-z		1920						1987	1997	
GISS-g	30N.60N	(1914)	1931						1997	
NCDC-z		1929				1963		1988	1997	
GISS-g	00N.30N	1923					1978		1997	
NCDC-z		1925					1978		1997	
GISS-g	00N.30N		1935				1978		1996	
NCDC-z		1925					1978		1996	
GISS-g	00N.30N		1935				1978			2000
NCDC-z		1925						1986		

	Southern Hemisphere									
GISS-g	**90S.00N**		1939				1979			2001
NCDC-z			1936		1956		1978			2001^{1}
GISS-g	**90S.00N**		1938				1976		1995^{2}	
GISS-z			1936			1968	1978		1995	
NCDC-z			1936			1968	1978^{2}		1996	
GISS-g	**90S.00N**		1938				1976		1995^{2}	
GISS-g*			*1938				*1976		*1995^{3}	
NCDC-z			1936			1968	1978^{2}		1996	
	Southern Hemisphere Zones									
GISS-g	**30S.00N**		1939				1976		1994^{1}	
NCDC-z			1936		1956		1978			2001^{1}
GISS-g	**30S.00N**		1939				1978		1996	
NCDC-z			1936				1978		1996	
GISS-g	**30S.00N**		1939				1978		1996	
NCDC-z			1936				1978		1996	
GISS-g	**60S.30S**		1931				1976			2002
GISS-g*			*1936				*1976		*1996	
NCDC-z			1937				1976			2002
GISS-g	**60S.30S**		1936			1968	1976			
GISS-g*			*1936			*1968	*1976		*1995^{2}	
NCDC-z			1936			1967	1976		1995	
GISS-g	**60S.30S**		1936			1986	1976			
NCDC-z			1936			1968	1976		1995	

Several things are demonstrated by the analysis in Table Ch5.14.

1. There is strong similarity between the datasets where they can be compared.
2. GISS-z (used in JR2017) and GISS-g (to be used in this work) are very similar especially later in the record. GIZZ-z and NCDC-z support a change after 1968 in the SH land-ocean data and GISS-g does not. This could be due to minor variation in composition because GISS-g and NCDC-z both return this date in the Southern mid-latitudes (SML).
3. Three post war shifts, all previously reported in the literature, circa 1976, 1986, and 1996, all show in multiple zones over land and ocean. 1968 shows as a presumably ocean based shift in the SML.
4. Superscripted dates indicating either only weak evidence of a change (1 or 2) or that a continuing trend has possibly led to a false determination of a step (3), are mostly associated with the time of the so-called hiatus, circa 1996. For most zonal analyses where this date shows, there is no reason at all to reject the date.
5. Differences in the post WW1 dates are of some interest. In the Northern Hemisphere (NH), NCDC-z shows changes in early to mid 1920s, whereas GISS sets all show changes in the mid 1930s. However the SH shows a strong consensus of changes in the 1930s.

The specific conclusion from this comparison is that the GISS-g data provide similar dates to the NCDC-z and GISS-z data, both at the resolution provided (2°x2°) and the resolution to be used for a spatial analysis (5°x5°).

Looking at GISS-g (see Figure Ch5.23 below, details in Table A5.1.32), both Land and Ocean points appear to be marginally more likely to be in segments classified as Single,Stationary (i.e. a single deterministic change within otherwise stationary data) than Land-Ocean points. This is consistent with composition of data by across land and ocean leading to more complexity in the resulting signal, a cause of compositional misspecification. Additionally, as illustrated in the bottom pane of Figure Ch5.23, analysis at zonal scale increases the number of points classified as Single,Stationary, and in the case of zonal land data, all change-points are so classified.

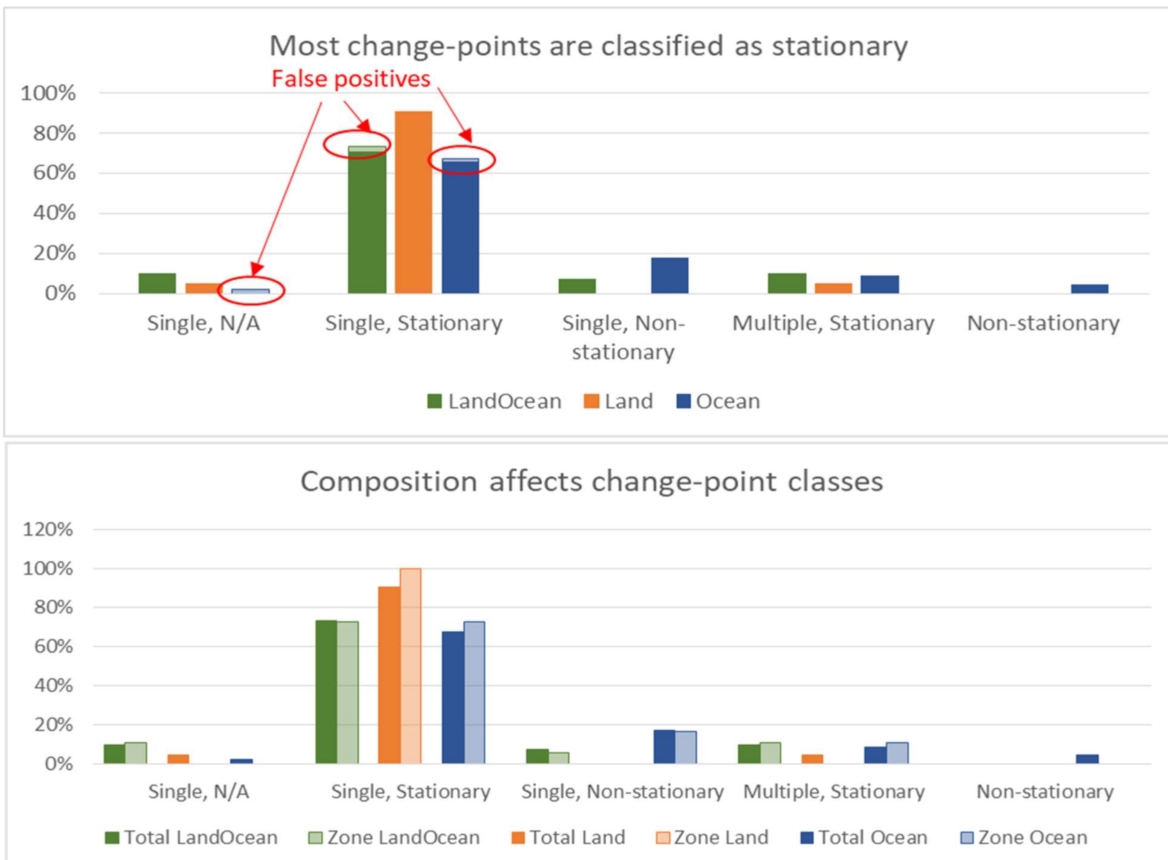


Figure Ch5.23: Change-point classes, GISS-g. Top pane shows the percentage of Land-Ocean, Land, and Ocean change points of each diagnostic class, with unsaturated colours representing points that are deemed false-positives on the basis of a likely continuing trend and rejection by ANCOVA. Bottom pane shows the same classes but this time saturated colours represent the same percentages as the top pane including false positives, and the unsaturated colours represent the percentages when only zonal change-points are selected. Land points are always stationary, and if zonal, always unambiguously single shift-like change-points.

Table A5.1.32 derived from GISS-g, is separated into three parts, land, combined land-ocean and ocean, and lists each year of change detected with diagnostics. Of the three well documented post-WW2 bio-physical events, circa 1976/8, 1986/8 and 1996/8, the first is seen more strongly in the SH, the second mainly on the Northern mid-latitudes and the third is very wide spread (See also Table Ch5.14). In addition an event identified in the late 1960s in the SH (Vives and Jones, 2005, Jones, 2012, Hope et al., 2010) is indicated in the Southern mid-latitudes. Table A5.1.36, an analysis of NCDC-z with matching zones is also provided for interest. The change-points for both datasets are plotted with the data in Figure Ch5.24 below.

There is a clear interhemispheric difference shown here in Table Ch5.14, Table A5.1.32, and Figure Ch5.24. But the great Pacific reorganisation, circa 1976, has been attributed as a NH event (Trenberth and Hurrell, 1994, Mantua et al., 1997, Minobe, 1997), and thus the non-detection in the Northern mid-latitude zonal records, especially given detection in the Southern zones is of some interest. It is possible this is simply methodological and a shift was masked by the circa 1986 event which was confined to the NH (Reid et al., 2015).

The results of the SBP test, applied to both NCDC-z and GISS-g datasets and all 36 time-series from each are tabulated in the Appendix 5.1, Table A5.1.31. Heteroscedasticity tests indicate that the step-and-trend model derived from MSBV analysis, when considered overall, is viable. However the difference in the ocean analysis between NCDC-z and GISS-g is striking, especially with the composite zones. It is notable that the GISS-g versions of the ocean based change-points show more heteroscedasticity that their NCDC-z counterparts. This may well be due to differences in part of the record. When GISS-g global ocean data (90S.90N) was tested by SBP against the change-points derived from the equivalent NCDC-z data (1889, 1929, 1976, 1986, 1996), the result is more probable (Pr=0.047) than the Pr=0.003 returned when using those derived by MSBV and GISS-g data (1901, 1938, 1976, 1995), and commensurate with Pr=0.069 given NCDC-z data and change-points. GISS-g has a larger excursion in sea surface temperatures in the early 1940s than other data sets. The analysis by sector (next section) suggests that this anomaly is relatively localised since it is strongest in the sector centred on the middle Indian Ocean (see Figure Ch5.25).

This aside, these results show that for land zones, the break-points residuals would be accepted as homoscedastic but the composite 90S.20S would not (NCDC-z: $0.01<p<0.05$, GISS-g: $p < 0.01$). For the most part regime shifts over land are well explained by the shifts detected. Land-ocean in both data sets are also homoscedastic ($p <=0.1$). Whereas the ocean shows differences between the datasets. For NCDC-z, 60N.90N (the Arctic) as well as 60S.30S may have had ($0.01<p<0.05$) but for GISS-g, 30N.60N, 20N.90N, 90S.00N (the SH) and 90S.90N (globe) all register as having residual heteroscedasticity, indicating that undetected features may be present.

The presence of heteroscedasticity in the residual time-series after removal of the break model may indicate (a) the presence of other types of variation than linear trends or abrupt shifts – including artefacts, or (b) an undetected abrupt shift.

The great Pacific reorganisation

There are 24 separate change-point determinations between 1976 and 1981, which correspond to the well documented biophysical changes and a change of the PDO to posisitive phase. For eight land based change-points, the ANCOVA results support a change-point, but only the global land change after 1979 has good support for a trend change by ANOVA[2] ($0.01<p<0.05$) (and the change-point preferred by the ZA is 1963, the year of the Mt Agung volcano (Labitzke and Naujokat, 1984)). Six of the eight are classed as Flat (trends not present and shift confirmed by ANCOVA) by the CP-index. All are classed as Single,Stationary.

However, for ocean points the situation is a little more complicated. ANCOVA gives highly supportive *p*-values, the 90S.20S composite is in a segment classed as Non-stationary but without significant trend while the 90S.00N (the SH) is classed as Single-stationary. This is a mild anomaly, but noting that 30S.00N has a large internal shift ascribed to 1978 in a Single-stationary segment, and that 90S.60S is generally not suited to processing the 90S.20N result may reflect poor data quality.

[2] The change after 1976 in 30S.00N has a statistically significant trend and is the only of these eight detections to have so. Oddly it does not have a significant change, and the prior trend is not statistically significant. In fact the issue to significance of trend changes is more vexed that one might assume from the literature.

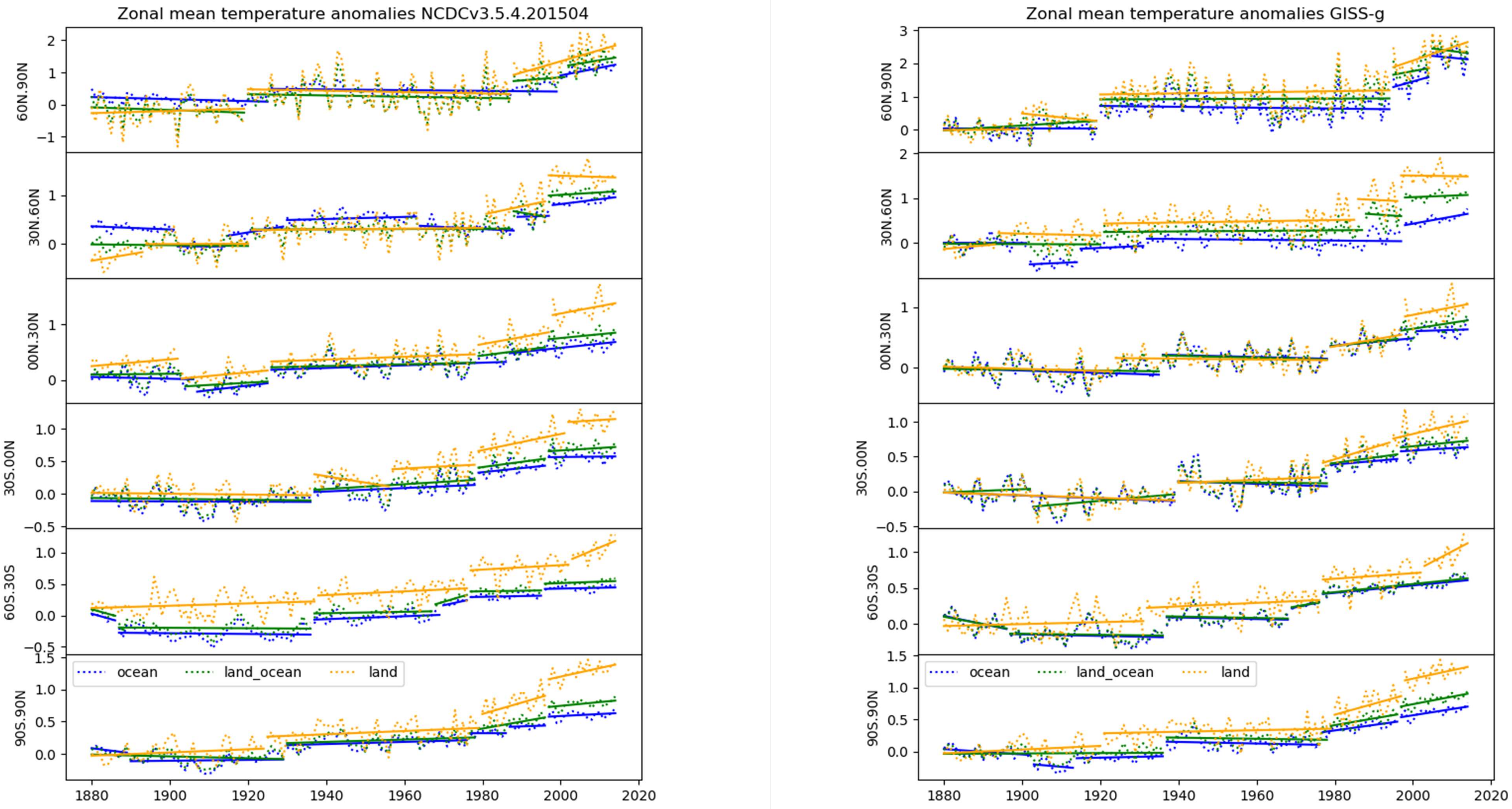


Figure Ch5.24: Left: NCDC v 3.5.4 (April 2015). Zonal change-points computed by MSBC for five of six zones, plus a global average, broken down by land and/or ocean extent, corresponding to the detailed analysis in Table A5.1.32. Right: GISS-g: The bottom pane of each is the global mean annual average for comparison. Of interest: land temperatures rise over a greater extent than ocean or land & ocean ones, the so-called hiatus, circa 1996-8 is present in all zones and all except land in the Arctic and ocean in the Northern tropical oceans.

Step-like change-points detected and reported in JR2017 are further substantiated.

Different versions of the same data produced by two different agencies give the same overall story post 1950, but differ somewhat – mostly in the northern tropical oceans – prior to that.

There is a little evidence that the composition of the data by averaging induces artefacts of the sort that *any* detection method will be sensitive to: lowering of events below detection thresholds; and apparent increases in autocorrelation and non-stationary progression.

Therefore given that the zonal data samples multiple ocean basins and that these potentially change independently, the tests were repeated for eight sectors of 45 degrees longitude in each zone, applying the reasoning in RJ2017.

Section 2: Sectors of zones in the observational data

The primary aim of this section is to explore the relationship between shifts in zonal records and in sub-zonal to regional records from which the zones are constructed. If regimes are more regional and at least partially independent within zones, averaging multiple regions before analysis is a compositional misspecification which is anticipated to cause the effects listed at the end of the prior section (also see discussion in Chapter 4 (Validation)). Consequently sectors should show simpler structure than the zones from which they are drawn.

Sectors consisted of the same six 30° NCDC-z zones sectored into 30°x45° blocks, with the Western extent of Drake's Passage (75°W-65°W) as an origin so that if forms the Western side of the first sector in zone 60S-30S. Sectors in zone 90S.60S, were omitted from further analysis as previously, but shown for illustrative completeness. They are then masked to produce ocean, land, and combined land-ocean area weighted averages; the same scheme as published in RJ2017. The MYBV was run for each resulting time series (see Figure Ch5.25 below), and the validation suite applied to the results as per Section 1. This gives 280 change-points. The resulting temperature change trajectories show quite a deal of variation. Change-points and analysis are tabulated in Appendix 5.1,

Table A5.1.33**,** and summarised in Figure Ch5.26 and Figure Ch5.27, below. The data segments containing change-points are less complex and the proportion of Single,Stationary segments increases when sectored data is used (Figure Ch5.26, top pane), and the number of probable false positives as shown by the CP-index (Figure Ch5.26, bottom pane) decreases. That is, the data are better specified for the MSBV. Additionally the proportion of change attributable to shifts increases at finer scale, and autocorrelation reduces (Figure Ch5.26).

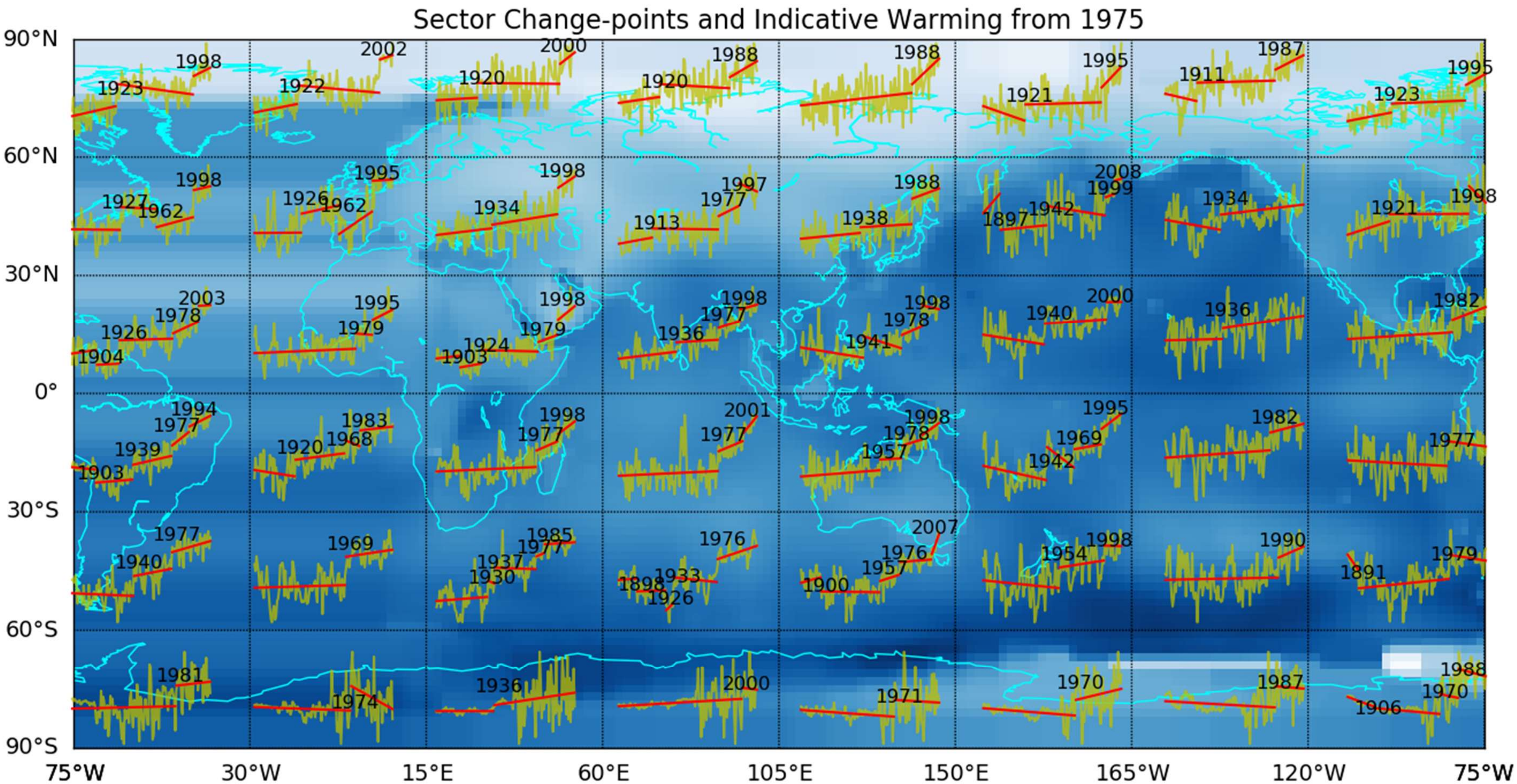


Figure Ch5.25: Zonal/sub-sector, combined atmosphere-ocean area weighted annual mean temperatures. Change-points identified by MSBV are used to delineate a disjointed segment statistical model, and shown within the grids. The background is warming from the 1950-1981 average anomalies as estimated by the 2007-2016 mean. Of interest the Eastern and Western sides of the Pacific show differentiation, and the Northern Atlantic shows features in the mid-20th Century that are suggestive of a slow rise and fall. Dates shown are the first changed-year, the start of a new regime.

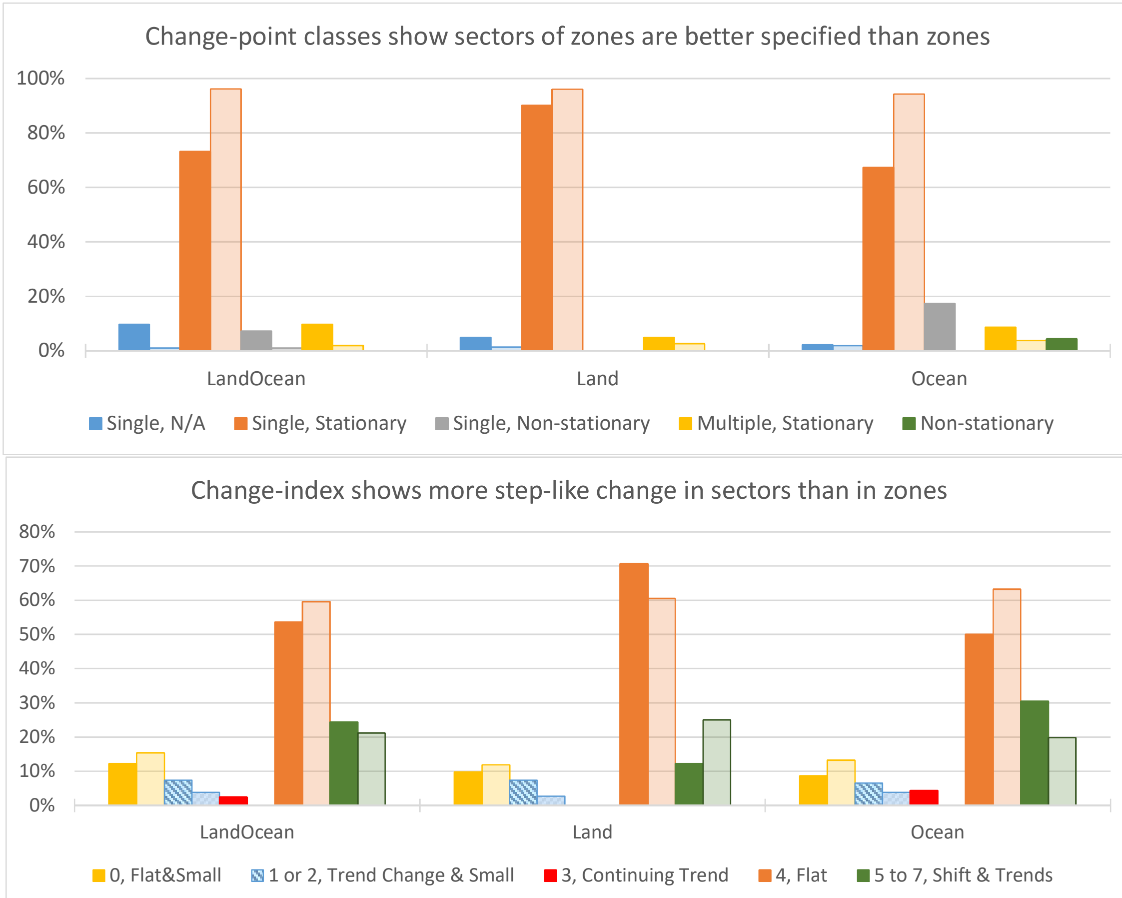


Figure Ch5.26: The two different assessment schemes for change-points both support the contention that data becomes less complex and less misspecified for the MSBV at finer scale. Top pane (extending Figure Ch5.23): in all three splits of the data, the change-point class of change-points in sectors of zones are more likely to be classified as stationary than from the zones themselves. Bottom pane: shows the change-index results for the same splits. Again, unsaturated colours represent sectors. Of particular note, none of the sector change-points had an index of 3 (corresponding to a continued trend), and the proportion of index values 1 or 2 reduce to less than 5%. Index values of 0 (MSBV assumptions not violated and thus may be preferred to ANCOVA) increases. Over land, the number indicating an indisputable shift plus a degree of trend increases.

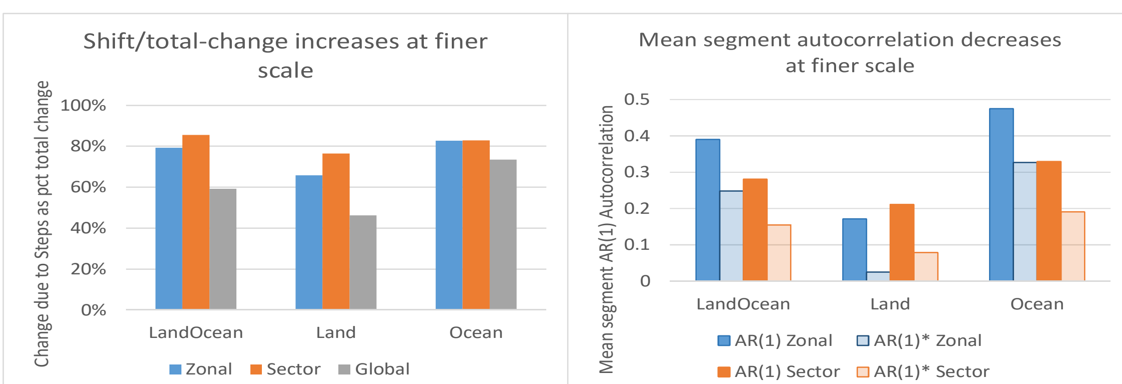


Figure Ch5.27: Left pane: Attribution of temperature change between shifts, and trend-like progression, shows that overall most of the change is attributable to shifts, the more so at finer scale. Ocean records are much more step-like (see Table Ch5.17 below). Right pane: Auto-correlation coefficients after attribution of autocorrelation between trend and non-trend processes (see Table Ch5.16). Analysis of data segments containing change-points shown in saturated colours; residuals unsaturated and indicated with an asterisk; Blue is zonal, orange is sectors.

Analysis with the SBP test , however shows that some of the sectors remain more complex and less well specified to the MSBV. The SBP was run for each sector of the combined land-ocean data and compared to the same test for zonal data (Table Ch5.15 below). In fact, only the northern mid-latitudes (30N.60N) show no heteroscedasticity, whilst the sectors 30S.00N/150E.120W (mid-South Pacific) and 60S.30S/30W.15E (southern Indian) show strong signs that there are other processes at play and this is apparent in Figure Ch5.25.

Table Ch5.15: Land-Ocean data: Studentized Breusch-Pagan tests of the implicit step and trend model of the full zone and the sectors or those zones. P-values less than 0.01 are shown in red, otherwise those less than 0.05, are shown in green. Only the Northern mid-latitude zone shows no evidence of heteroscedasticity even at sector scale. Visual comparison with Figure Ch5.25 suggest that in some cases a slow oscillation may be present.

	Full Zone	**75W.30W**	**30W.15E**	**15E.60E**	**60E.105E**	**105E.150E**	**150E.165W**	**165W.120W**	**120W.75W**
60N.90N	**0.678**	0.043	0.019	0.382	0.490	0.324	0.548	0.619	0.245
30N.60N	**0.126**	0.908	0.352	0.074	0.210	0.894	0.115	0.070	0.553
00N.30N	**0.563**	0.874	0.723	0.476	0.928	0.258	0.021	0.830	0.395
30S.00N	**0.302**	0.097	0.711	0.671	0.300	0.277	<0.01	0.039	0.343
60S.30S	**0.035**	0.482	<0.01	0.283	0.647	0.154	0.252	0.023	0.052

Table Ch5.16: Tracing sources of apparent autocorrelation Average of coefficients of equation (2) for each of combined Land-Ocean, Land, and Ocean splits when data are aggregated into Zones, and Sectors of Zones. Trends don't vary between Zones and Sectors, but the autocorrelation coefficient always reduces in residuals compared to initial data, and reduces from zone to sector over land-ocean and ocean splits.

	Initial segments		**Residuals**	
	autocorrelation	Trend (°C/yr)	autocorrelation	Trend (°C/yr)
Zones				
Land-Ocean	0.394	0.00620	0.249	0.000141
Land	0.187	0.00807	0.046	0.000136
Ocean	0.432	0.00537	0.275	0.000114
Sectors				
Land-Ocean	0.281	0.00582	0.155	0.0000571
Land	0.211	0.00741	0.0786	0.000054
Ocean	0.329	0.00526	0.191	0.0000926

Table Ch5.17: Ratios of change attributable to shifts to sum of shifts and change attributable to trends. Shown are the results when global averaged data are analysed, zonally averaged data are analysed, and when data is analysed at sector scale. (See also Figure Ch5.27 above)

Shift/Total Ratio	**Land-Ocean**	**Land**	**Ocean**
Global	0.59	0.46	0.74
Zonal	0.79	0.66	0.83
Sector	0.86	0.77	0.83

The reduction in the mean segment autocorrelation from initial data segments to residuals is consistent with deterministic change-points being removed, and in fact the virtual disappearance of it for land based analyses is highly suggestive of autocorrelation not being a factor at all over land. This would be consistent in turn with regimes over land Reductions moving from zonal to sector averaging are consistent with additional autocorrelation induced (not reduced) by averaging of the data.

The move to sub-zonal scale also shows that the major regime-like shifts post WW2 are more complex in their distribution than hitherto recognised. Analysis of the GMST record shows three major events in data up to 2014 and there is agreement within a year on the timing of these.

However the sectorial analysis tends to suggest a mix of localised and more generalised events clustering around the major dates. Figure Ch5.28 gives two views of the same data – the years of change in land, ocean and combined land-ocean sectors. Post WW2, the mid-latitude North Atlantic shows a shift circa 1961/2, corresponding to the change of phase of the AMO from warm to cool, and the South Pacific shows a more extensive shift in 1968. It is only later that three wide spread, more recognised events occur. The spread might potentially be due to data quality or processing issues, methodological artefacts, or have a physical basis. Chapter 6 (Spatial patterns) suggests there is a real physical basis with multiple ocean and atmospheric systems involved.

After the two post WW2 ocean events (the second of which is visible in land-ocean data), a slightly complex series of events, constituting the Great Pacific reorganisation circa 1976/8, shows with shifts predominant in the SH oceans and with land shifts occurring in two distinct events two years apart. The circa 1986 Atlantic event shows mainly as land based changes. The so-called hiatus circa 1996 also resolves into a North Atlantic regime change around 1994 and an extensive coordinated event commencing two years later.

Interhemispheric contrasts were highlighted by concentrating on those events which occurred in more than one sector. A simple partitioning of the change-dates into NH and SH, selecting only the more extensive events, is illustrated in Figure Ch5.29, and shows clustering at nearly decadal intervals. The circa 1968 event occurs in eight SH sectors of the sixteen available and is this quite extensive, the circa 1976 event is again seen to be more widespread in the SH, the circa 1986 event shows over NH and the circa 1996 event is more widespread.

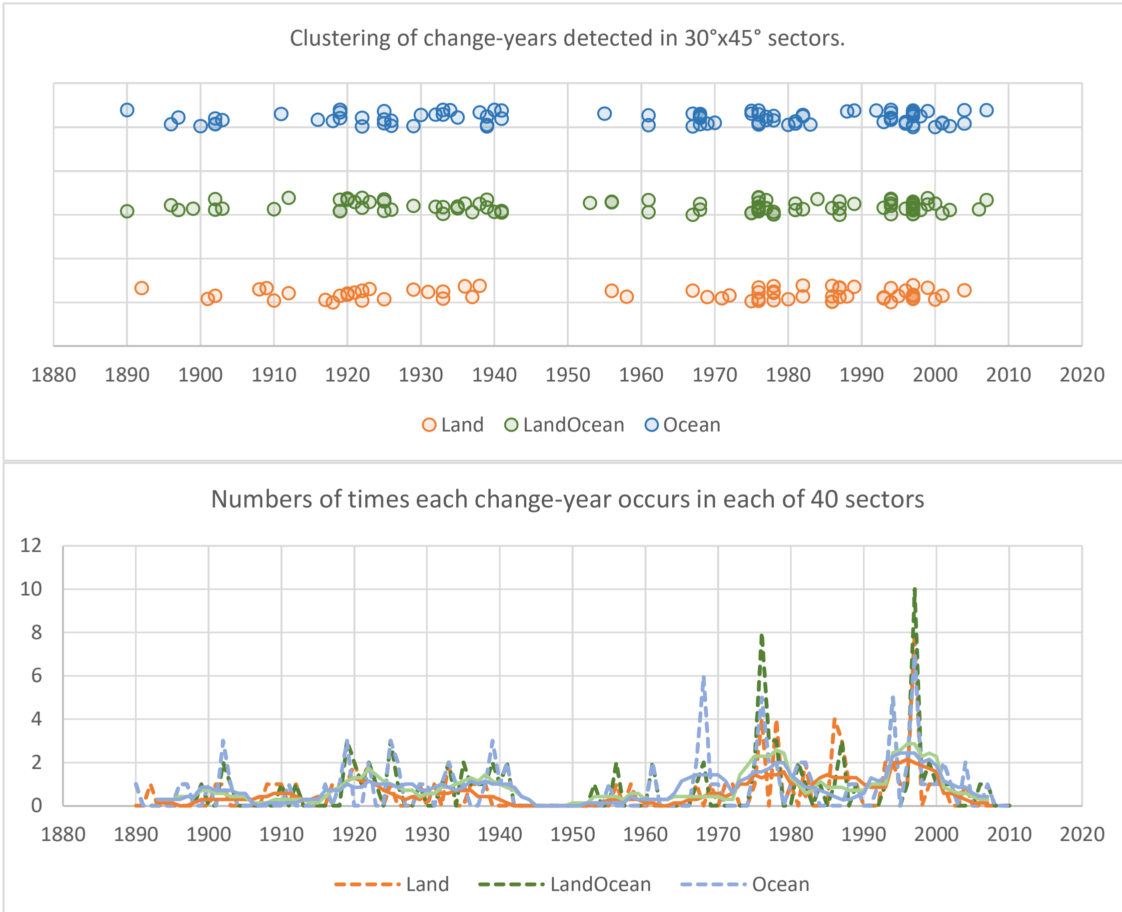


Figure Ch5.28: Change-years detected by MSBV in forty sectors, covering all of the globe except for 90S.60S: Top-pane shows the years, spread vertically by random jitter. Bottom pane shows the number of sectors at each year (dashed lines) and seven year running means (solid lines). Post WW2 shows increasing numbers of changes commencing in the 1950s, with observable clustering and culminating with a step-like shift involving 25% of sectors circa 1996. The SH ocean regime change circa 1968 is quite wide spread in the hemisphere oceans, the circa 1976 great Pacific reorganization shows with a distinct double peak over land, the circa 1986 event which affected the NH shows here mostly in the land record, and the so-called hiatus circa 1996 shows as the most extensive event, but preceded by discernible changes round 1994.

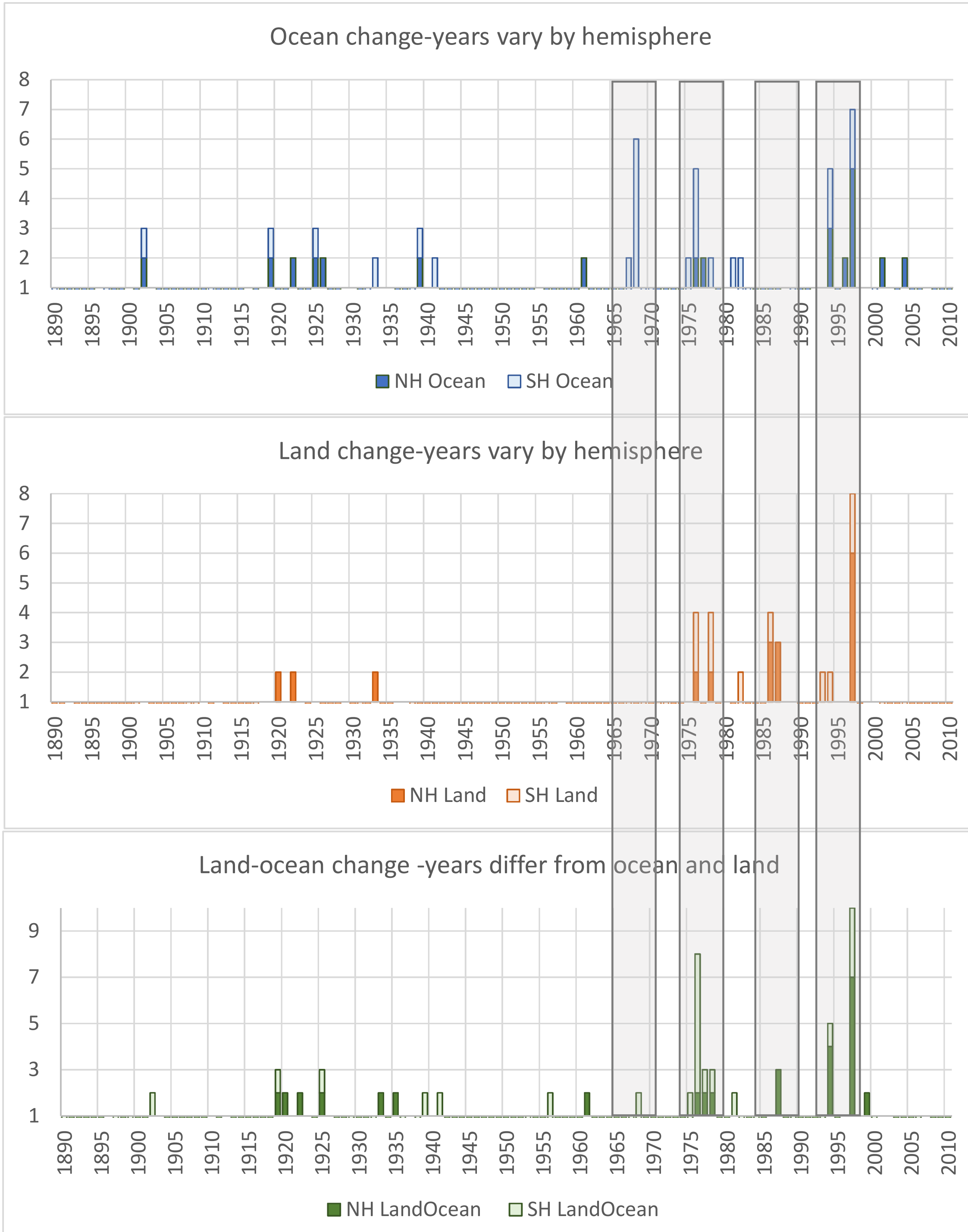


Figure Ch5.29: Numbers of occurrences of specified years of change for NH and SH, land, ocean and combined laand-ocean sectors, showing only those years which occur in two or more sectors.

To summarise this section. There is further evidence that the area averaging of climate data produces artefacts of the sort that obscure real signals in the record. Moving from zonal to sector scale decreases metrics of both non-stationary progression and autocorrelation, whilst also essentially eliminating suspicion of false positives in the MSBV due to continued trend being identified as a step-like change.

The analysis also supports the somewhat unexpected results seen in Table Ch5.14, whereby the step-like changes seen in the global record and corresponding to the great Pacific reorganisation of 1976/7 are seen more strongly in the Southern Hemisphere (SH) than the Northern (NH) despite being documented first in the NH.

The so-called hiatus is shown to be a complex of events. The 1994 events correspond to Northern Atlantic changes and align with a change of AMO from cool to warm phase in the same general areas as the corresponding 1961 change of AMO phase from warm to cool (Veres and Hu, 2013). Following this is an extensive set of shifts with dates that align with the PDO change of phase from warm to cool.

Section 3: Changes in zonality as evidence for H2 (interaction)

Global climate models are run under a variety of assumptions so as to probe differing aspects of climate. In this study, there are two principal sources of variation: (a) the selection of model and (b) the model atmospheric chemistry prescriptions.

The detection part of the study is performed using the MSBV. The change-point index metric is used to determine whether a change-point is supported in the presence of trend (CP-index values 0 or 4-7), ambiguous (index values 1 or 2) or not-supported (index 3).

The mean temperature records (a.k.a. “tas”, temperature at surface) for the global climate models and runs listed in Table Ch5.13 were selected. For each model the same zones as supplied for NCSD were produced by area weighted averaging, using the supplied land/ocean masks to yield land, ocean, and combined land ocean runs.

The final analysis consisted of more than 35,000 rows in a table representing all change-points detected for 29 climate models, pre-industrial controls, and two principal future climate trajectories. The tests from the validation suite were run for all change-points as described in Chapter 4.

The purpose in analysing the zonal data from models was to address the following questions, raised in JR2017 and also above.

1. Do GCMs reveal patterns of zonality consistent with observations?
2. Do patterns of zonality change with warming? Such a relationship has not been established in the literature although much work identifies state based organisation of the oceans. How does this zonality relate to the hypothetical unperturbed Earth, and then to the imposition

of forced anthropogenic change and the projected zonality under differing regimes of forced warming?

3. JR2017's hypothesis H2, that forced warming interacts with natural variability, requires interaction in either or both of two ways. Is there evidence to support either of the following?
    a. It may alter (and probably shortens) the periodicity of quasi-oscillatory ocean systems such as the PDO (McGregor et al., 2010, Kirby et al., 2010).
    b. It may intensify the extents of transitions from phase to phase of variability modes. This could be observed as step-like shifts rising above the detection threshold and first appearing as novel – new patterns of change may emerge and these may be coordinated with each other, "recruited", across zones.

Analytic approach

For this analysis all models were treated as equally representative samples of possible past, historical, and future Earth climates, which allows for an initial assessment based on ensembles. Addressing the issues above in order:

Altogether 35,296 individual change-points were generated by MSBV including six non-overlapped zones, global, hemispheric and other zones. The six non over-lapped zones were selected and tested from the pre-industrial control a.k.a. piControl (2762 points), RCP2.6 (5588 points) and RCP8.5 (6597 points) runs. This covered land and ocean and combined land-ocean splits and six non-overlapped zones. It was immediately obvious (see Table Ch5.18, below) that for the piControls the duration of regimes between step-like changes was far longer than observed and longer than the observational periods in the same models. Pre-industrial controls runs varied in duration and in nominal dates. Therefore they were analysed against a normalised base interval of 1000 years, and additionally analysed after bootstrapping. The question of zonality in the pre-industrial Earth and during changes of industrialisation, then on into projected futures was explored by computing the duration of regimes in each zone during suitably chosen epochs. Apart from the piControls, these were the years until post WW2, 1850-1950; the years from then until observed onset of rapid warming, 1951-1975; the years until RCP2.6 and RCP8.5 suddenly diverge, 1976-2040; and from then until the end of the 21st century 2041-2100.

The RCP8.5 data in particular after 2005, appears to more often resemble noise about an increasingly curvilinear progression. This tendency was described as steps progressing to an escalator under warming (Jones and Ricketts, 2016a). Increasing trend also increasingly pushes the MSBV out of its range of assumptions, the data is increasingly misspecified for the detection method. Therefore it is important to assess the effect of this misspecification. For this purpose the CP-Index (Chapter 4, Validation) was used and false positives defined as those with a CP-index of one, two, or three (ANCOVA p-value > 0.05 pre-change or post-change trend non-zero (p> 0.05)). This is a broader definition than the rejection of change-points with a CP-index of three.

The difference between the two RCPs is quite striking (see Figure Ch5.30). Even though RCP2.6 and RCP8.5 do not diverge greatly until after 2040 the results for two intervals, 2016-2040 and 2041-2100, differ greatly. For RCP2.6 overall the false positive rate is quite acceptable, 4.7% over all zones between 1850 and 2100. However for RCP8.5 they are much higher post 2005. Over 1850-2100 false-positives are 24%, for 1850-2005 they are 3.3%, and post 2005, 35%. The impact of this on individual statistics has not been fully assessed. This is possibly to be due to the RCP8.5 change-points in the 21st Century being not well aligned to step-like change.

The ensembles used for the two scenarios are not identical, but this does not explain Figure Ch5.30. As a result detailed analysis of RCP8.5 is treated as questionable after 2020.

Zonality in observations and models during observational period.

Changes in zonality are judged by changes in the duration of regimes. To provide a suitable measure of duration of regimes in zones, a mean return period is defined. This is simply the average interval between change-points between the years in question for all of the models in the ensemble.

The performance of the model ensemble was compared to observations in the period 1880 to 2014.

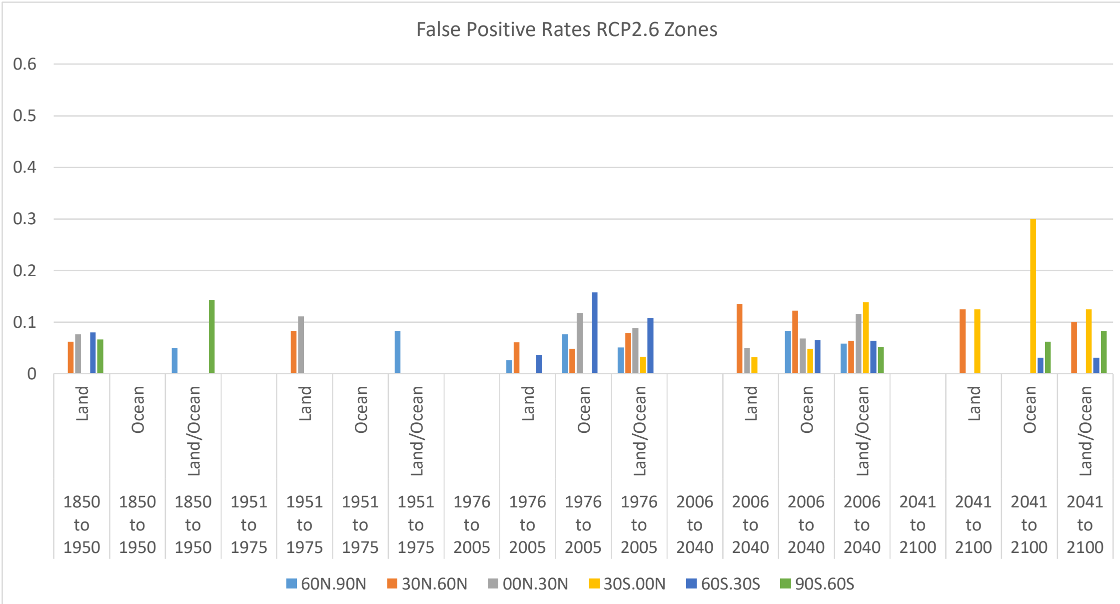
Table Ch5.18: Ensemble mean/median return intervals for significant (ANCOVA p<=0.05) regime like changes. Results are shown with North-most in the left column to South-most in the right hand column. A. Pre-Industrial Controls (ensemble of 23 global climate models). B. Observed. C. RCP 2.6 (ensemble of 25 climate models). D. RCP8.5 (ensemble of 28 climate models). Boot-strapping was used where possible due to relative sparsity of data in some cases, in particular for Pre-Industrial controls. Return intervals returned by bootstrapping are shown to the left of the slash and raw means on the right. (See also Table A5.1.34 and Table A5.1.35 in the appendix)

A. Pre-Industrial	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	342/341	387/387	830/850	808/828	422/424	409/414
Ocean	190/192	114/114	438/445	400/410	58/59	108/109
Land/Ocean	215/216	176/177	445/450	454/460	55/56	154/154

B. Observed 1880 to 2014	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	67	34	45	34	45	N/A
Ocean	67	22	45	45	27	N/A
Land/Ocean	45	45	34	45	27	N/A

C. RCP2.6 1880 to 2014	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	54	57	49	53	52	78
Ocean	47	39	46	44	38	68
Land/Ocean	48	44	46	49	39	74

D. RCP8.5 1880 to 2014	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	48	45	42	47	47	73
Ocean	38	29	36	38	28	51
Land/Ocean	35	36	37	40	30	57



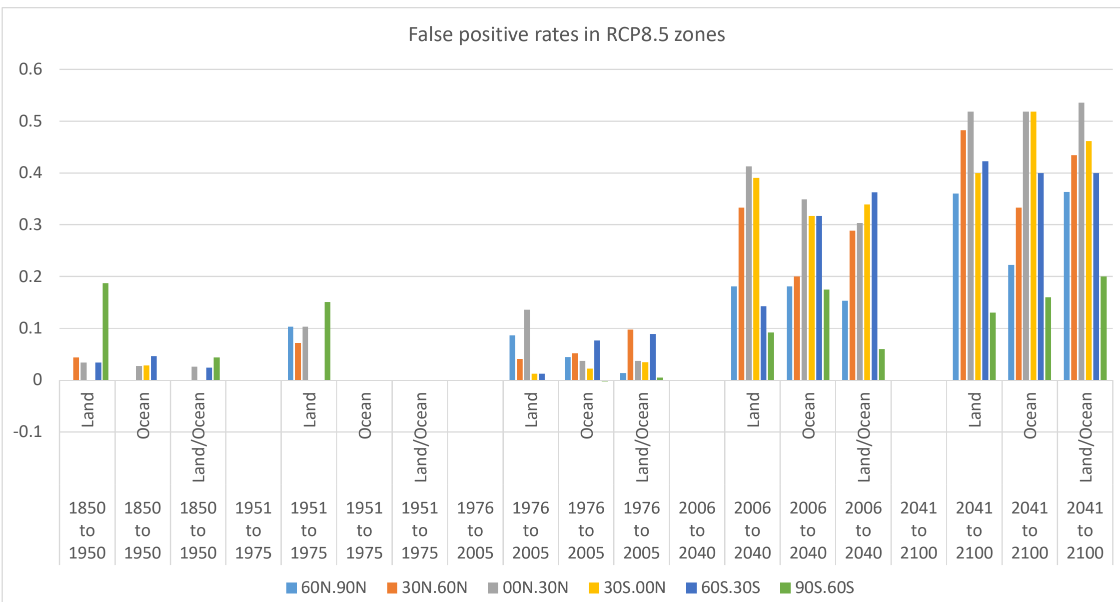


Figure Ch5.30: False positive determination rates for RCP2.6 and RCP8.5 model runs given MSBV as a detection method. Note especially the difference between the post 2006 groups, where false positive rates average out at 35%

Figure Ch5.31 and Table Ch5.18 show the computed return periods for step-like regime changes over the observational period (1880-2014), for observations and models broken into zones. The southern and northern mid-latitude oceans show shorter return periods than other ocean zones or land zones. In general model land periods are longer than observed, but for observed and modelled data, models correctly show land returns greater than oceans.

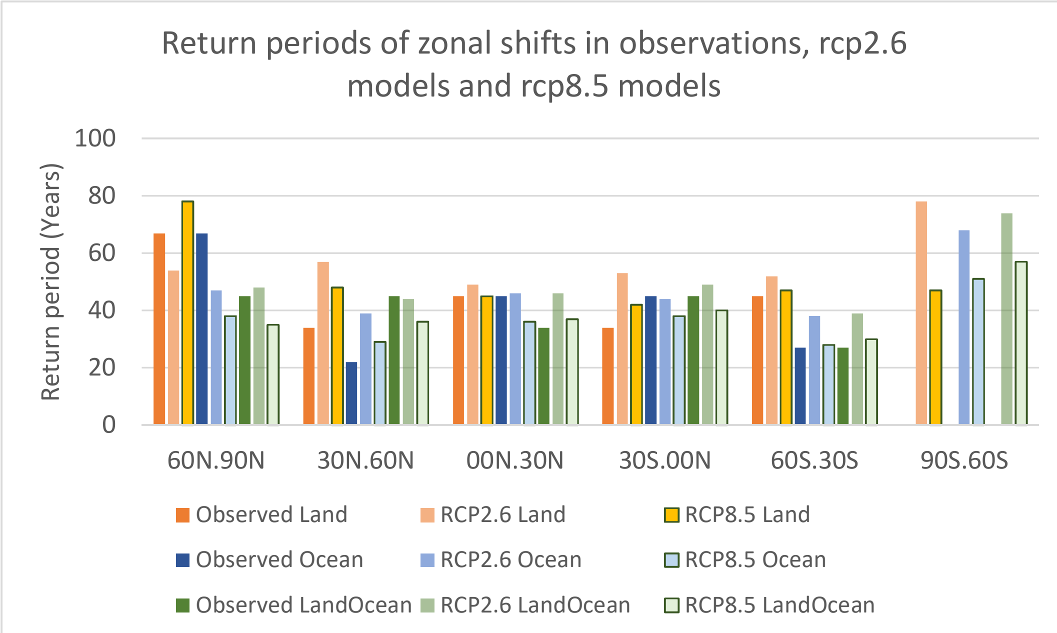


Figure Ch5.31: Return periods of step-like regime shifts in the six defined zones shown for land, ocean and combined land-ocean data (see Table Ch5.18). The model behaviour mostly mirrors observations in that ocean return periods are generally shorter than land ones, more so in the northern mid-latitudes (30N.60N) and especially in the southern mid-latitudes (60S.30S).

Pre-industrial controls do not have an equivalent to an observational period. While the return periods are very much longer, the oceans of the northern mid- latitude are noticeably shorter than the rest and the southern mid-latitude even more so. The last point is consistent with findings previously reported in the SML (Ricketts and Jones, 2017, see their Figure 2).

This indicates that the models tell a similar story as observations as far as present day zonality is concerned. This gives a basis for examination of changes within ensembles of RCP2.6 and the early part of RCP8.5 model runs.

Changes in patterns of zonality

The PI-Control periods are very different from the RCP runs but are consistent with the proposition that in an undisturbed Earth with a stable atmosphere (and in the absence of singular events, for example volcanoes) the SML is the principal area where localised step-like regime shifts occur. Within the historical period, observations and models both show regime changes becoming more frequent as warming continues, with the land based regime changes post 2040 in particular becoming less frequent and ocean ones reverting to something similar to pre-1975 rates.

The analysis of return periods of regime shows major features as follows

1. The Pre-Industrial climate as modelled shows very rare shifts over land and more over the oceans.

2. The SH mid-latitudes show much more frequent shifts than the tropics with the Northern mid-latitudes also more active.
3. Throughout, the Southern mid-latitudes shows a tendency to be more active than other zones, especially once greenhouse gas levels and mean global temperatures rise above pre-industrial levels.
4. The early historical parts of the models (RCP2.6 and RCP8.5 share common historical data within models) still seem to quite rapidly converge on observations and the Northern and Southern mid-latitudes are more active than the tropics. At the same time the tropics seem to rapidly become about five times more active than in the pre-industrial era.
5. The RCP2.6 and RCP8.5 ensembles contain a slightly different mix of models and the two "eras" from 1850-1950 and 1950-1975, differ a little but once observed temperatures commenced a rapid rise after 1976 – which is mirrored by the models – the two RCP sets converge.
6. Once the two RCP ensembles diverge around 2040, RCP8.5 continues even more actively with regime shifts in the Southern mid-latitudes occurring at rapidly as every 17 years. RCP2.6 on the other hand rapidly reverts to activity more reminiscent of early industrialisation with land zones showing even more reduction in activity.

Figure Ch5.32 shows the changes in return periods of regime shifts for both RCP2.6 and RCP8.5 (values are tabulated in the appendix, Table A5.1.34 and Table A5.1.35, and the changes of clustering across zones can be seen in Appendix 5.2 also, Figure A5.1.54, Figure A5.1.55, and Figure A5.1.56, for interest). To estimate error bounds the model data were pooled for each of land, ocean or combine land-ocean in each zone, and then bootstrapped to give an estimate of the distribution of return periods. As can be seen, for all zones, and for both RCPs the period from 1976 to 2040 shows much shorter return periods than the previous periods. These difference are very significant. Again, it can also be seen that the mid-latitude oceans appear as more active than elsewhere. The changes of zonality are consistent with a pre-industrial climate with abrupt shifts that are predominantly oceanic and mid-latitudinal in origin, where the southern mid-latitudes are especially dominant. Once anthropogenic warming commences the system rapidly increases the frequency and zonality of step-like changes, and as shown in RCP2.6, once warming is reduced or becomes negative regimes shifts rapidly reduce in frequency with the normally most active area, the Southern mid-latitudes remaining more active. In the presence of continued warming, as in RCP8.5 no reversion is observed.

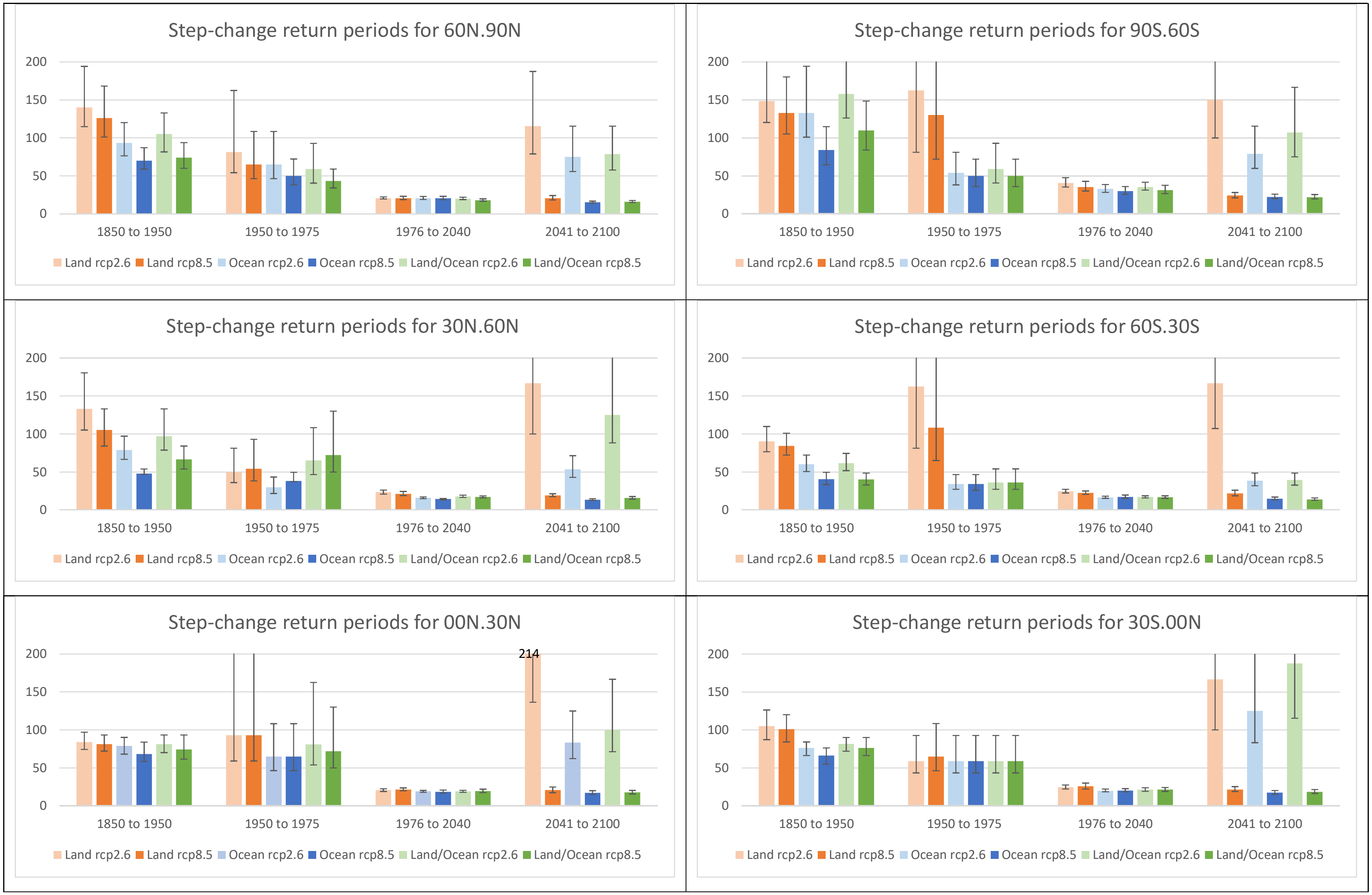


Figure Ch5.32: Comparison of median of bootstrapped return periods for significant step-like changes in RCP2.6 and RCP8.5 forcings. Error bars are 5%-95%. Each zone is plotted separately with the most poleward zones (Arctic and Antarctic) on the first row.

Intensification and "recruitment" of zones.

Not only does the frequency of step-like shifts change within zones, events also seem to involve more zones at once. From the above analysis and as seen in Table Ch5.14 earlier, it appears that whereas in the pre-industrial times most regime shifts identifiable by the MSBV will have been localised ocean based events, this is not the case in industrial times. Rather, it appears that a number of events have shown as coordinated regime like changes over multiple zones. Additionally, as seen in JR2017, a very comprehensive ensemble of RCP4.5 climate model runs shows clustering around the three documented post WW2 change dates, and suggest another change may occur in the second decade of the 21st Century. Events may appear in different zones at slightly different times. Whilst it is possible that this is due to imprecision in detection, it may also be physical, for instance involving multiple systems. In either case it is reasonable to treat all events in adjacent zones within +/- two years as a single event. Therefore a simple analysis was conducted. A year by year count of the number of models in each RCP ensemble for which the window round that year of +/- two years contained significant (ANCOVA p <=0.05 or CP-Index = 0) change-points for three or more zones. These are plotted and a five year running mean of the counts is plotted as well (see Figure Ch5.33). Land, Ocean and Land-Ocean are plotted separately.

In the historical period a land peak occurs about 1969, and ocean peaks around 1976 and 1985. RCP2.6 shows a consensus of land based shifts circa 1970, and ocean shifts circa 1976 and 1986 (with some land shifts in 1986). RCP8.5 based on a slightly different mix of models, shows land and land-ocean peaks circa 1971, and ocean and land-ocean peaks circa 1976 (this may correspond to the earlier SH change and the later Pacific changes observed in Section 1. By far the strongest consensus peaks are in land, ocean, and land-ocean between 1995 and 1998. In RCP2.6 there is a pair of peaks, round 2008 and 2019, which are possibly separate estimates of the same phenomenon from differing models rather than being two distinct events (due to the MSBV's seven year prohibition) with land and land-ocean showing peaks between 2019 and 2021. In RCP8.5 there is an ocean and a land-ocean peak, circa 2015 and an ocean peak circa 2021, which should be treated as two separate estimates of the same event from differing models for the same reason. However, given the evidence that the MSBV is operating beyond its design criteria the RCP2.6 results are preferred.

If the consensus approach to the models is valid, then the well documented bio-physical regime changes post WW2 are represented in the model consensus. The so-called hiatus, circa 1996/8 is strikingly strongly represented. There is a strong indication that another such regime change is to be expected in the second half of the second decade of the 21st century. There are early indication that this may have occurred. The PDO showed signs in 2013 of changing phase (see for example

(Thomson and Emery, 2014) where STARS was used), and recently a large jump has been detected in Global Mean Temperature and associated with ocean heat release (Yin et al., 2018).[3]

The analysis of model zones gives strong evidence that regime changes interact with warming in line with H2. They become more frequent, and as shown by RCP2.6, once warming reduces so regimes become less frequent again. As shown by the transition from pre-industrial to observed forcings (Table Ch5.18) and the clustering in Figure Ch5.33, especially the decline post 2040 in RCP2.6, the coupling between forcing and interaction is quite rapid, and interaction also shows in the spatio-temporal clustering with events occurring across zones at similar times together with the numbers of zones increasing.

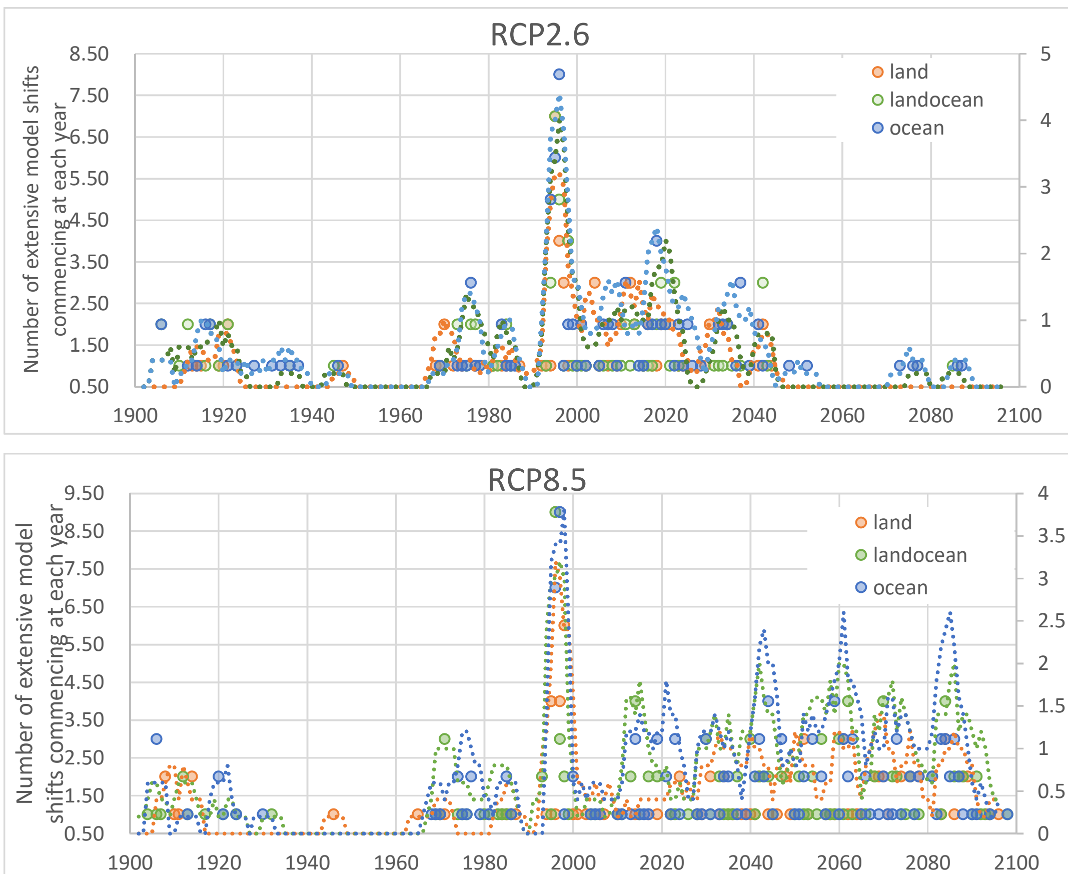


Figure Ch5.33: RCP2.6 (top) and RCP8.5 (bottom). For each split the number of models which predict shifts in three or more zones for each year (dots). The dotted line is a five year running average on the secondary axis.

[3] It is now highly likely that a regime shift occurred circa 2014

Discussion

Observational records

In the observational series, there are distinct differences between land and ocean change-points. And there are differences between zones (see also Table A5.1.32 in the Appendix).

Composition misspecification issues also occur, and they show not just by use of stationarity tests, but also by cross comparisons of events by date. For instance, in the zonal analysis of NCDC zones (Appendix Table A5.1.36), an event circa 1957. Over land a shift after 1957 seems to be associated with the southern tropics, but does not show in oceans. It shows as an insignificant negative trend change in the tropics 20S.20N with a positive shift, with a CP-index of zero (ANCOVA has p-value > 0.05 as do pre and post trends). However the test on 30S.00S zone assigns a positive trend change plus a positive shift, confirmed by ANCOVA and the same date is assigned by the ZA test as an exogenous change. This is consistent with the event being localised in the SH, and 20S.20N zone carrying a mix of signals.

Land records and Ocean ones then mostly align with events from 1976/8, 1986/8 and 1996/8 although it is not clear whether statistically insignificant land shift in 2002/3 in the southern mid-latitudes should be assigned a meaning.

In general change-points over land, change-points are almost entirely classed as “single-stationary”, which I interpret to mean the events are wide spread and coordinated.

The ocean signal is more complex than the land one and almost certainly reflects the results of complex internal states. As already noted the tendency for smaller sectors to be more likely to be stationary strongly supports this and provides a first approximation of the dimensions of co-varying areas in that they are bigger than the sectors and smaller than sub-basins or zones.

The exogenous change year, as determined by the ZA, corresponds to the most likely year of change with any or all of a step, a transient, or a trend change. Hence the year given is not based on the same criteria as the MSBV. Thus, when the MSBV and the ZA agree on the year of change, this can be taken as evidence that within the whole data segment step-change is the most likely type change, *and* that random drift is not the explanation. Several dates show up persistently in the ZA tests which seem to be indicative of a strong but transient responses. In particular dates after 1936 up to 1945 over land and after 1936 up to 1962 over oceans are frequently found by ZA and not by MSBV. This suggests that a strong but relatively transient heat release was responsible. Further support to this is shown by the spatial analysis in the next chapter. 1963 shows up over land commonly by ZA tests and only rarely otherwise. This is consistent with the Mount Agung volcano and a transient reduction in atmospheric temperatures and subsequent recovery. The series of step changes between 1919 and 1925 is prominently a land feature, while there are no ZA identified exogenous changes between 1914 and 1935.

The three most identifiable and best characterised global regime changes are circa 1976/8, 1986/88 and 1996/8. Closer examination of the ocean results shows that step-like and transient responses are not easily separated prior to 1950.

1963 is very likely the transient response to the Mount Agung volcanic eruption.

1945 and surrounding years may be a complex heat release event. It is predominantly an extra-polar event, likely originating in the tropics, and involving the Pacific and Indian through-flow. This year is also the year often assigned as the change of sign of the PDO from warm to cool phase (Minobe, 1997).

The ocean records in the southern mid-latitudes could likely represent records within which there was a strong effect of compositional misspecification. This was tested by examining sub-sectors of the segment, and reported above in Section 2, and also by Ricketts and Jones (2017).

Based on these results it seems plausible that there are several components to be considered in deciding what constitutes a regime change at decadal scale.

1. The appearance of step-like regime shifts in temperature records at any scale is not an artefact of random drift.
2. The simple composition of data from regions which show any form of independent but sequentially organised regime changes is likely to be deceptive and may initially give the appearance of random drift, but detailed analysis shows otherwise.
3. Non-stationarity, when detected by the tests used here, almost entirely resolves to exogenous step-like changes when the data is subdivided spatially. This suggests that such extreme autocorrelation is a result of systemic complexity, rather than simple behaviour.
4. Land and ocean responses seem to be differently organised. This is entirely in line with the idea that heat transports over land are the result of atmospheric states and these states are in turn influenced by ocean temperature states with which they are partially mutually causal. It’s also a sign of rapid adjustments in radiative equilibrium, and is a feature of the circa-1987 shifts (Jones, personal communication).

Analysis of climate models

Regime changes behave differently by zone.

It is clear that the majority of events detected in observed and in all modelled time series are relatively local to regional in scale. It is very clear (Figure Ch5.32) that in the pre-industrial control runs, regime changes are quite rare in the tropics, especially over land, whereas they are more common in the mid-latitudes. The mid-latitude oceans and 60S.30S in particular, show rates of regime change much higher than the tropics. The majority of these shifts however have the characteristics of incoherent composition – being averages of processes which are under non-shared influences.

Observed and modelled historical data show less of a differential across zones and a higher frequency of shift over-all. The observed mid-latitude oceans still have more changes than the tropics as do the modelled oceans.

Regime changes behave differently within zones over time

Analogues for the major post-WW2 regime changes are present in the climate model consensus analyses. The time-wise relationship is almost certainly a reflection of the rate of warming. During times of little warming (modelled and observed pre-1950, PiControls, and RCP2.6 post 2040) regime changes are relatively rarer, and less tropically concentrated than during high rates of warming.

Given that the PDO again seems to have changed phase, it is likely to be followed by onset of more rapid warming (Henley, 2017) and also a substantial shift upward. RN Jones has extensively analysed recent data and shows just such a shift (personal communication).

Regime changes behave differently between land and ocean.

Zonally analysed ocean shifts tend to be more likely than land to be measured as having being non-stationary, especially in the SH. In the SMLs there is evidence that this remains after the internal trend and shift are removed, indicating that there is a great deal of structure in the signal. The simplest explanation is the ocean signal is regionalised with shifts at differing locations and times. Land shifts by the same logic would be predominantly zonal. This is seen in Section 2.

The zonality changes under modelled anthropogenic warming.

The climate models tell a story of shift-like regime change under steady state conditions being distinctly absent from the tropics and more common in the mid latitudes, and also being mostly an ocean phenomenon. Coherent shifts involving three or more zones, are almost absent, but become common in the RCP runs.

The earlier part of the historical runs up until 1950, show more activity and some more tropical involvement. The first sign of multiple zone shifts are post WW2, circa 1976, 1986 and 1997 and the tropics are involved in the first and last of these. These are both coincident with a PDO phase change. RCP2.6 and RCP8.5 are both consistent with continued tropical involvement under strong warming with the RCP2.6 scenario showing a rapid reversion to near pre-industrial frequencies toward the second half of the 21st Century.

Conclusions

The temperature record is complex, but not so much so that features cannot be extracted. The stationarity tests are interpreted as indicators of misspecification of data composition, rather than necessarily as indicators of underlying uncorrelated climate processes, or as reasons to accept or discard an analysis.

This chapter has shown that regime shifts interact with forced global warming. They (a) vary in intensity, (b) become more global and coherent with forced warming, (c) become more frequent, all consistent with the hypothesis of warming/variability interaction.

It has also shown that regime shifts (d) are roughly decadal, (e) regional on what appears to be sub-ocean basin to ocean basin scale, (f) possibly more zonal over land.

Some evidence has been given than ensembles of global climate models suggest that a step-change and regime change is due in the second decade of the 21st Century, and I have noted that some publications suggest this has already happened at the time of writing.

The validation suite has proved useful in giving nuance to the work, and further refinement of this approach is warranted.

The MSBV seems to return many possible false positives in strong warming scenarios when judged by the ANCOVA test on a point by point basis (see Appendix 5.1, Figure A5.1.54, Figure A5.1.55, and Figure A5.1.56). This is an issue which will need to be addressed in future work.

There is no guarantee that structures responsible for decadal scale variability are aligned with observations, or GCMs or even within the set of individual realisations of any one. The ensemble of CGMs analysed here, none-the-less, produces a consensus, especially post 1950, that produces peaks of regime changes circa 1976, smaller in 1986, and very strongly circa 1997, each corresponding to an observed and well documented event. The future evolution, post 2006, is "blurrier" but the most prominent peak in RCP2.6 is centred on 2018. RCP8.5 also shows a matching peak, with increasing less coherence going forward in time. The prior considerations temper this to an extent as a prediction – at the very least however it suggests that GCMs form internal decadal variability modes that can be entrained by external forcing and that this entrainment is coherent for a while.

Therefore there is a possibility that a regime change has very recently occurred in the model ensemble, and this suggests that the next generation of models, which will have incorporated the observed conditions up until 2015 should be studied for evidence of the next change after this one[4].

[4] And (see Chapter 7) there are early indications that a regime change occurred circa 2014, coincidental with a change of phase of the PDO from cool to warm.

Chapter 6: Spatial distributions of abrupt climate shifts

Introduction

Zonal analysis of observed and modelled climates (Chapter 5, Zonal Analyses) shows that there are distinct variations in the zonal distribution of abrupt shifts. At least two climate variability modes, the Pacific Decadal Oscillation (PDO) (Bjerknes, 1969, Meehl et al., 2009, Trenberth and Hurrell, 1994), and the Atlantic Multi-decadal Oscillation (AMO) (Knudsen et al., 2011, Schlesinger and Ramankutty, 1994) have been previously linked to tele-connected changes at global scale. Signatures of abrupt regime change corresponding to both have been found in zonal analyses (JR2017) and were further elucidated in Chapter 5.

This chapter extends the previous chapter. The principal finding is that coherent spatial patterns of step-like shifts in the surface temperature record correspond to phase changes in the PDO, and the AMO; and that these align with changes in deeper ocean circulation. These new patterns are quite different in meaning from the signature patterns of these indices computed by empirical orthogonal functions (EOF see for example (Newman et al., 2016 Fig. 1)). Spatial features that align with the indices are delineated to create a type of ontology. The last part of the chapter makes use of this to lay out some challenges for the use of global climate models. In doing so, it becomes clear that climate models are not initialised to, and do not evolve so as to, reflect the coordination of the large scale structures found. The finding of the last chapter, that climate model ensembles appear to predict a shift in surface temperatures between 2015 and 2020, suggests further work well out of scope for this thesis.

Of the six requirements for a program of investigation to severely test the relationship between step-like and trend-like processes listed in JR2017, this chapter addresses four of them.

- Test 1 What patterns of step changes can be detected in temperature observations? Do particular dates and locations line up with known events or processes?
- Test 2 Do models forced by historical emissions reproduce the patterns of steps changes shown in observations?
- Test 5 Do other climate variables also undergo step changes?
- Test 6 Are temperature time series more step-like or trend-like?

As with the previous chapter, data are examined using the MSBV, this time at 5°x5° grid scale, internal shifts and internal trends are computed, the numbers of change-points are counted.

A strong result is that shift-like temperature progression becomes more predominant over trend-like progression when data are analysed at finer scale. This is shown by comparisons of global, zonal and spatial observed temperatures, and using global and spatial modelled temperatures.

The previous chapter does suggest that abrupt and decadal regime changes are dominated by ocean changes with land shifts following, possibly after a persistent regime like change of atmospheric circulation. Peyser et al. (2016) have demonstrated a potential switching mechanism involving the West Pacific warm pool, which Jones interprets as a component of a trigger mechanism for regime changes (personal communication), and Henley (2017) concludes that Pacific Decadal Variability is the result of continuous tropical-extratropical interactions at decadal timescales, warning that too little attention has been paid to alternating attenuation and amplification modes of variability.

The rest of this chapter is organised as follows. Section 1: The multi-step bivariate test (MSBV) is used to delineate shift-like features in the surface temperature record (addressing Tests 1 and 6). Section 2: A variation is used to examine changes in the regression relationships of the ocean mixed layer and the deeper layers (addressing Test 5). Section 3: A sample of global climate models is examined at spatial scale, and some features are identified similar to those present in observations (addressing Test 2). Section 4: This draws together results from previous sections. Warming attributed to shift-like behaviour is shown to increase when data is analysed at finer than regional spatial scale. This can only happen if shifts are present in the data *and* they occur regionally (Test 6).

Data sources and methods

Unless otherwise noted, all spatial data used in this chapter are as per Appendix 1.1, and were regridded to a common 5°x5.

The GISTEMP3 surface temperatures, ocean temperatures at 100m and 700m, and CMIP5 surface temperature data were analysed for the presence of step like changes in annual averages using the MSBV with a Rapid Assessment Stopping rule (Chapter 3), for every grid point where there were no missing data. Once step-like changes were located, the internal trends and internal shifts were computed. An ANCOVA test is performed, but other tests from the validation suite were not performed routinely due to computational constraints, however selected data were tested during development as required. Change years reported here are the first changed year – the first year of the new regime, in keeping with (Vives and Jones, 2005, Jones, 2012).

Spatial analyses of observed temperatures.

Section 1: Patterns in gridded annual observed surface temperatures.

The GISTEMP3 data were analysed on a 5°x5° grid as above. Variables collated include the times of change for each grid point with their step-change value, probability, and T_{i0} value. For each step-change point, linear regression of the data on either side was used to produce estimates of internal trend, and the internal shift as the change-point. The probability associated with the change-point by an ANCOVA test was obtained using purpose built Python code at the same time. For every grid-point, I computed year by year the progressive change corresponding to the combined internal trends and internal shifts. I also computed the progressive changes imputed to just the shifts, and to just the trends, and the number of changes. This produces separate spatial files in NetCDF format for each variable, representing the annual values of each variable at each grid point over time.

Summary plots

The "step and trend" statistical model is composed of abrupt internal shifts and ongoing trends between these shifts (imputed from the step and trend model and the changes detected by MSBV), which when combined give a close approximation to the mean temperature. This enables an attribution of warming between abrupt internal shifts and ongoing internal trends spatially. Figure Ch6.34 and Figure Ch6.35 show the spatial patterns of overall temperature change attributable to abrupt internal shifts and internal trends. Figure Ch6.36 shows the number of step-like changes detected during the same period (see also Figure Ch6.39 below which focusses on the Pacific and Atlantic Oceans). Each figure shows the spatial map in the left pane, and a time-latitude Hovmöller plot on the right. These Hovmöller plots show the evolution over time of a spatial signal by averaging all but one spatial dimension and plotting the time-series of the result in colour. In the first three figures, the signals are zonally averaged (all values along each latitude).

Figure Ch6.37 compares the statistical step-and-trend model on a spatial grid to the observations, both being treated as anomalies from the mean of the last 20 years of the nineteenth century. It shows on the left pane, the sum of the internal shifts and trends 1880 until 2008. On the right pane, the difference between that and an estimate of observed temperatures (the 15 year mean about 2008). This year was chosen because the detection method will not detect changes after then due to a seven year prohibition rule. The step-and-trend statistical model captures much of the longer term variability. One would expect that, simply because more segments enable closer tracking of a signal, regions with more step counts also have closer agreement between the step-and-trend estimate and the observations (hence reduced spread of errors). This is the case. All of the locations on the 32x72 grid of the difference (illustrated in Figure Ch6.37, right pane), were grouped according to the number of corresponding change-points (as seen in Figure Ch6.36), and tested for a dependency between the number of change-points and the spread of differences from observations. There were too few with more than five changes to be meaningful. Bartlett's test for homogeneity of variance was applied to the groups and this strongly supports a dependency between all groups (p=8.88x10^{-15}), and between two, three, and four steps (p=0.01) (see also Figure Ch6.38, below).

Total shifts are overwhelmingly positive, over land especially. Negatives values are confined mostly to: an area in the Western tropical Pacific North contained in the region 120°E to 150°E, 15°S to 15°N; much of the South Pacific bounded by 120°E to 270°E, 45°S to 60°S; much of the Northern Pacific from 45°N to 60°N. As can be seen in the right pane of Figure Ch6.34, since the late 1980s the zonal average temperature change attributed to shifts everywhere has been positive.

The nett contribution of trends however is more complex, but shows that much of the extra-tropical Pacific, extra-tropical North Atlantic and much of the Southern Ocean show nett negative temperature trends when internal shifts are factored out. As can be seen in the Hovmöller plot in Figure Ch6.35 below, two zones approximately 30°N and 60°S have shown negative trends throughout. The Northern polar and later the higher mid-latitudes have shown general warming due to trends. Since the 2nd World War the Southern Hemisphere also shows warm trends.

The Southern Hemisphere, particularly in the poleward mid-latitudes, shows more of a tendency for local regime shifts to occur than the Northern mid-latitudes (Figure Ch6.36). Looking closer, "hot spots" of step change in the oceans appear to be more common to the West of most ocean basins, except the North Atlantic; The are also more common off the coast of Chile, and over Western Australia and Southern India. Only five locations, all oceanic, have not experienced a step-like change. Of these

Shift-like changes occurred first in the late 1890s in the higher and mid-latitudes, then in the Northern tropics and more generally in the 1900s. In the 1930s shifts in the Tropics and extra-Tropics occurred. More substantial changes occurred first in the Southern Hemisphere in the 1960s and quite rapidly over the rest of the planet in ensuing decades.

The difference between land and ocean is interesting. Land, in contrast to the ocean, shows mostly positive cumulative step changes and with the exception of the Eastern part of North America only positive contributions due to trend. The East of North America showed a very strong upward shift in temperatures circa 1997 followed by cooling trends to the point that as of now the nett contribution to warming due of trends is nil, and the total temperature rise results from a series of shifts, with most coming post 1997. The West of North America has warmed in a series of smaller shifts with trend predominant only North of 60°N.

Ocean changes show quite a different pattern. Often, and mostly extra-tropically, total shifts and cumulated trends are of opposite sign. For example: the Western Pacific around 30°N, the region just north of PNG, around 55°S in western Indian Ocean, the mid Pacific 160°E to 225°E from the equator to 45°S and much of the Eastern side of the North Atlantic Ocean. The tropics however, show continued residual trends.

The distribution of "hotspots" of regime changes is shown in Figure Ch6.36 below, and a more detailed examination of the Pacific and Atlantic Oceans is shown in Figure Ch6.39 below. Several things are highlighted. Firstly, in both the Pacific and Atlantic the rate of occurrences of regime changes marked by shifts increases over time. Secondly the North Pacific shows increases in such regime shifts to the West, but the South Pacific clearly has a zone of action off Chile that connects with the Southern Ocean directly South. This region has been associated with decadal scale variation in density compensated temperature and salinity (spiciness) anomalies that seem to relate to ENSO and the PDO (O'Kane et al., 2013, Risbey et al., 2014). The same region is associated with strongly negative overall internal trends against relatively minor overall total change. The more Eastern (closer to Chile) area has very high shift to total change ratios. Thirdly, the Tropical Pacific shows more activity to the Western side than the Eastern, and the Eastern extent of the North Pacific does not have a "hotspot" that corresponds to the South Pacific. The Western tropical Pacific activity against Eastern tropical quiescence has been shown to be quite meaningful and is highly consistent with a self-regulating heat engine in the tropical Pacific that governs the process of step-like change in climate (JR2019).

The Atlantic shows overall higher rates of regime shifts with the Northern tropical area being a somewhat more active that the Southern tropical area. The regions of lower activity correspond spatially to the extra-tropical gyres. Activity in the West tropical Atlantic does not appear to co-vary with the Eastern tropical Pacific. There are four regions where no shifts are found. A small region in the tropical mid-Pacific, consistent with the heat pump proposed in JR2019, the Northern and Southern Pacific mid-latitudes, and the middle of the North Atlantic mid-latitudes.

Finally the relative contributions of internal shift and internal trends are computed by using area weighted averaging to obtain values for land, ocean, and combined land-ocean. Global and hemispheric values are in Table Ch6.19 below, where clearly, the majority of warming appears to have occurred in periods of rapid or abrupt change. It can be seen that overall, warming over land shows about 62% of change due to shifts and oceans 79%. Land/Ocean is composed of separately analysed land and ocean grids and serves here to provide contrast to previous zonal analyses conducted on blended averages.

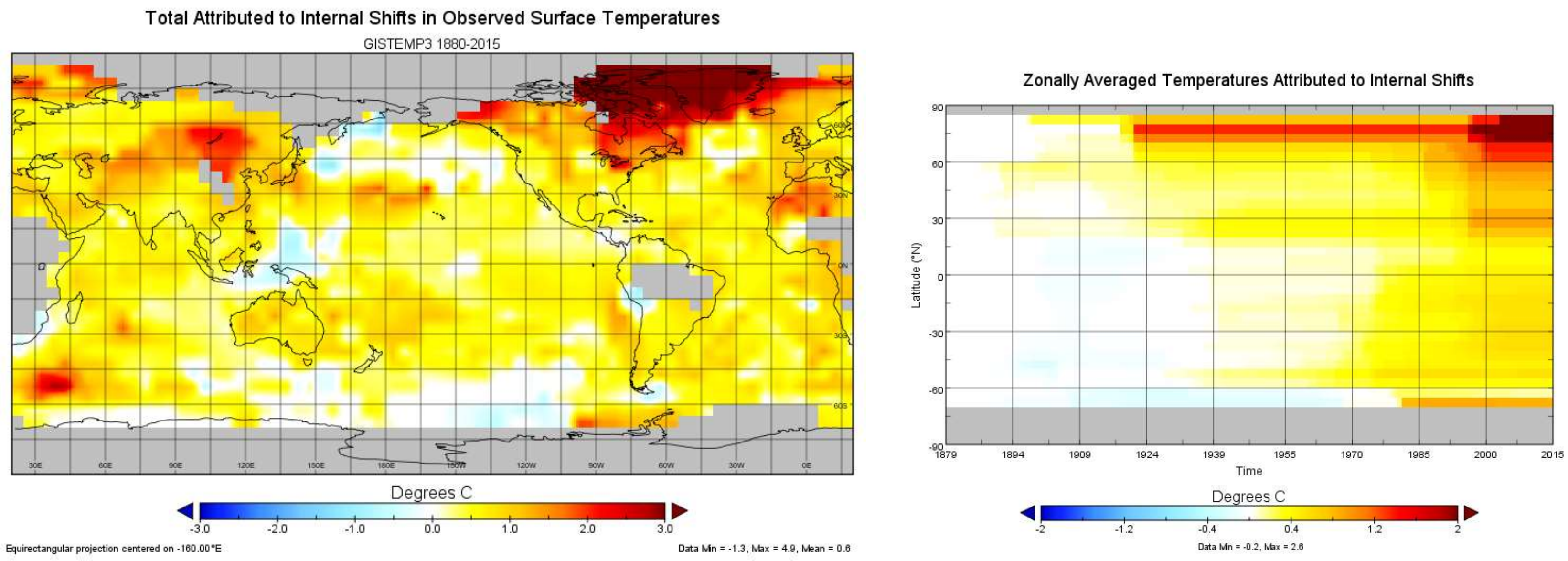


Figure Ch6.34:GISTEMP3 surface temperature changes. Left pane. Total temperature change attributed to internal shifts from 1880 to 2015. Right pane. Time-latitude Hovmöller plot of the same data showing zonally averaged changes evolving over time.

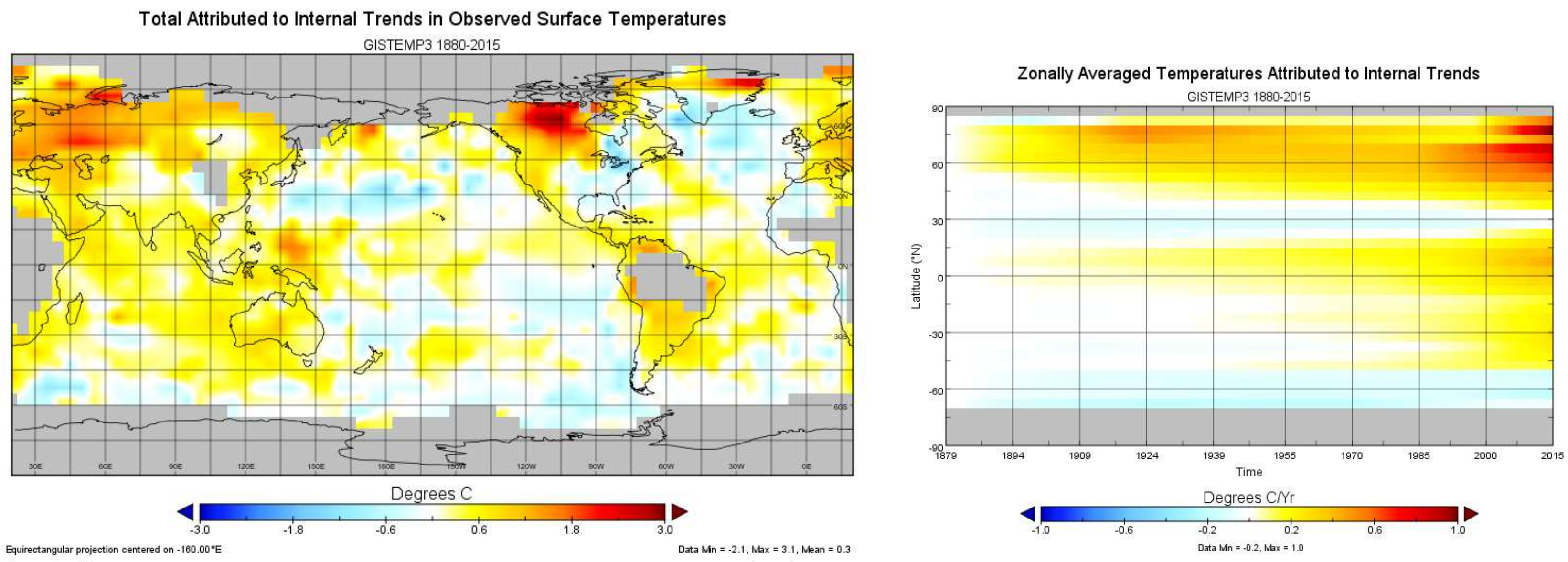


Figure Ch6.35: GISTEMP3 surface temperature changes. Left pane. Cumulated internal trends of surface temperature from 1880 to 2015. Right pane. Time-latitude Hovmöller plot of the same data showing zonally averaged changes evolving over time. Note that progressive warming due to trends shown this way will be smooth due to the removal of shifts.

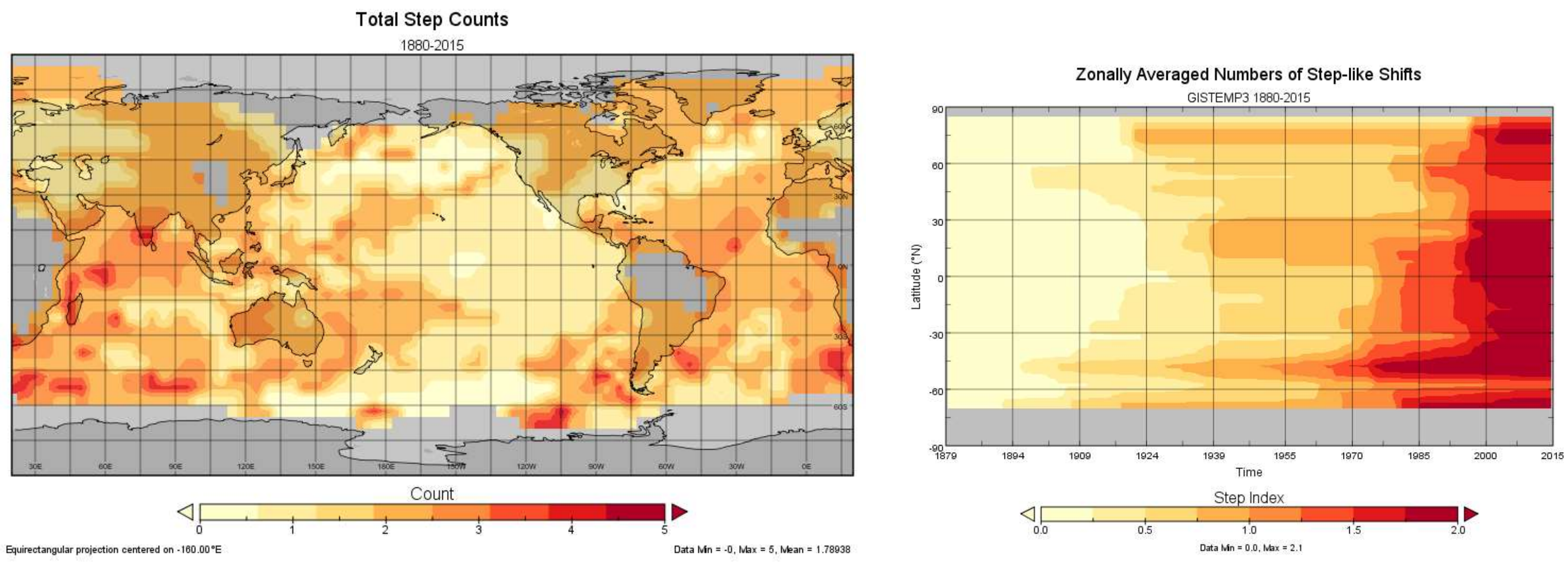


Figure Ch6.36: Numbers of detected step-like changes in surface temperature corresponding to the above. Left pane. Spatial distribution. Right pane. Time-latitude Hovmöller plot of the same data showing zonally averaged counts.

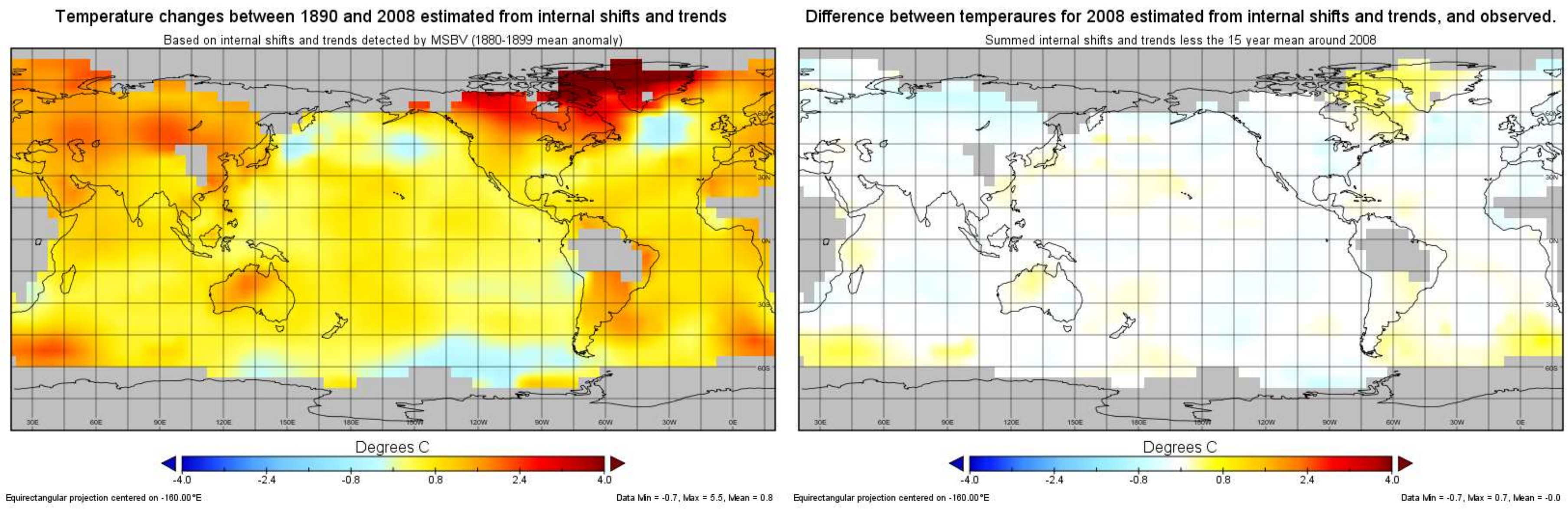


Figure Ch6.37: Left pane: Sum of changes due to internal shifts and due to internal trends (Figure Ch6.34 and Figure Ch6.35). These are based on anomalies from 1880-1899. Right pane: The difference between the data shown on the left and the fifteen year GISTEMP3 mean observed temperature centred on 2008 (Positive where the former exceeds the latter).

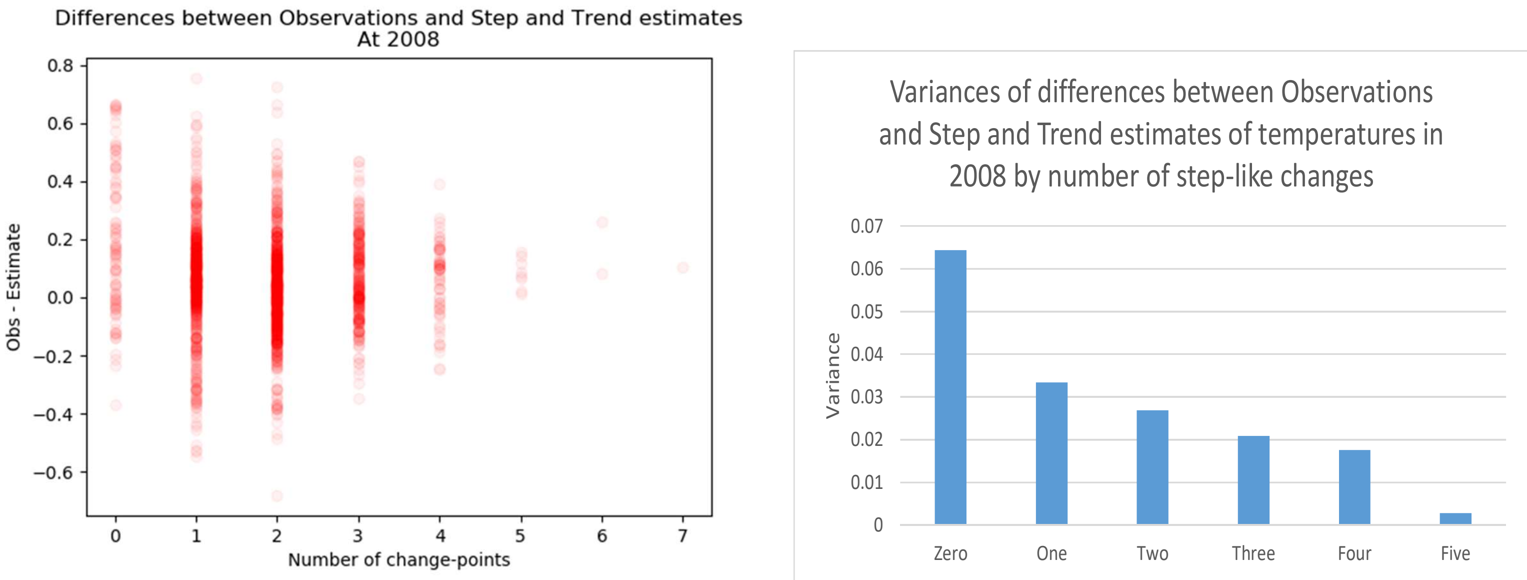


Figure Ch6.38: Where step changes are more frequent (see Figure Ch6.36), the difference of the observations and the estimate from internal trends and shifts (Figure Ch6.37) becomes closer as shown by the spread (left pane) and the variance (right pane).

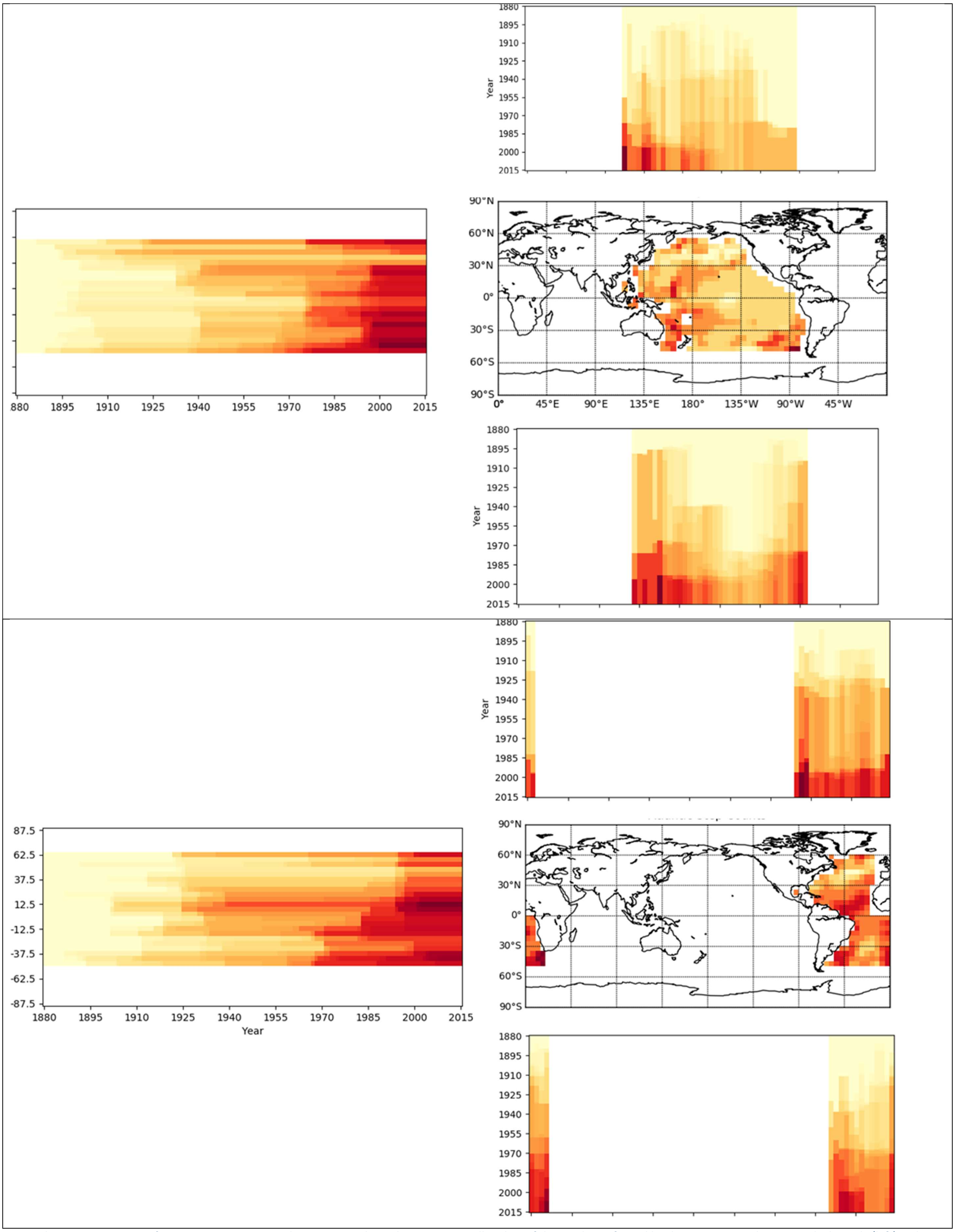


Figure Ch6.39: Top pane, Pacific Ocean: Bottom pane Atlantic Ocean: Shown here are Hovmöller plots of the numbers of step-like changes averaged over time by latitude (left) and longitude (Northern Hemisphere above, and Southern Hemisphere below spatial maps), of the number of step-like changes (or shifts) up until 2015.

Table Ch6.19: Differential warming attributed to internal shifts and internal trends – and for land, ocean and combined land/ocean. The trend proportion is simply the temperature change attributed to trend divided by the total change.

Location	Hemisphere	Shift °C	Trend °C	Total °C	Shift Proportion	Trend Proportion
Land	Globe	0.70	0.44	1.13	0.62	0.38
Land	NH	0.79	0.47	1.27	0.63	0.37
Land	SH	0.42	0.32	0.74	0.57	0.43
Ocean	Globe	0.50	0.13	0.63	0.79	0.21
Ocean	NH	0.52	0.15	0.66	0.78	0.22
Ocean	SH	0.49	0.12	0.61	0.80	0.20
Land/Ocean	Globe	0.56	0.22	0.78	0.72	0.28
Land/Ocean	NH	0.64	0.29	0.93	0.69	0.31
Land/Ocean	SH	0.48	0.15	0.63	0.76	0.24

Annual patterns

The MSBV reports years of change, that is, the year after which a change is clearly established. This is not necessarily a universal usage, and so the notation used in this discussions is YYYY/Y (e.g. 1996/7) to mean the year of change/first changed year. Where single dates are mentioned in the figures and titles this will be the first changed year.

The previous chapter produced three main events post WW2, all corresponding to well documented wide ranging biophysical changes, that lead to apparent step-like regime changes in the mean global surface temperature (see also Jones and Ricketts, 2017b) an event circa 1968/9 is included here because it was detected in the Southern Hemisphere and has been previously documented (Jones, 2012, Kirono et al., 2009). The zonal analysis of the previous chapter shows that none of these are truly global, although the so-called hiatus event which shows in most zones between1996/7-1998/9 comes closest (and in zones where it is not present zonal changes are found between 2001/2 and 2003/4, leaving an open question as to whether these are the same event).

The years of change in the GISTEMP3 mean global surface record are 1929/30, 1978/9, 1996/7. Additionally examination of the zonal shifts Table Ch5.14 shows that in the 20th Century, shifts may have occurred within a couple of years in more than one zone 1920/1-1921/2, 1926/7, 1937/8, 1968/-1970/1, 1976/7-1978/9, 1987/8-1988/9, and 1996/7-1998/9. It is possible that the 1929/30 change in the global analysis date is intermediate due to averaging artefacts. The deceptive effect of averaging was examined in some detail in Chapters 4 and 5, and is a finding of this thesis.

Accordingly, patterns of change points were examined annually concentrating on the six groups of dates above, with 1921/2 to 1926/7 as one group (see Figure Ch6.40 to Figure Ch6.45, all below). It is clear that changes detected in global and zonal records are spatially complex, tend to be focussed on specific regions and evolve over several years. The spatial evolution is shown in these figures, each with four panes. From to top left to bottom right the panes are as follows: (a) the first year change at each grid point is shown with the earliest in light orange and the latest in red; (b) the internal shift at each point; (c) the internal trend prior to the change; and (d) the internal trend after the change. Only points which when tested by ANCOVA have p<=0.05 are shown. Note that this is highly conservative, in that ANCOVA is less sensitive that MSBV where all of the assumptions of the MSBV are met.

1922-1928: (See Figure Ch6.40). An organised shift-like response in the East of the North Atlantic in 1925 corresponds to a change in sign of the AMO from negative to positive (see figure 1 in Knudsen et al., 2011). The trends on the more southerly arc of this feature can be seen to have changed from a slight cooling to a slight warming, whilst the in the northerly part the reverse is the case. A similar feature present in 1994/5 (Figure Ch6.45 below) also corresponds also to a change of the same sign in the AMO (McCarthy et al., 2015). In 1922/3 there is also a step-like response over western Greenland which has been previously noted (Mosley-Thompson et al., 2005). The next time such a shift is detected is in 1997 when there is also a much more wide spread response, but this feature is less obvious. Links between the AMO and Greenland regional temperatures have been explored (Trenberth and Caron, 2001, Mann et al., 2014) but not to the best of my knowledge from the angle of regime changes in surface temperatures.

1937-1943: (See Figure Ch6.41). A step-like shift in 1937 over Northern India, and then from 1939 a step-like rise in the Western extra-tropical Pacific in the vicinity of West Pacific Warm pool, which propagates a little to the North and strongly to the South by 1941. Changes in ongoing trend are quite small. Although the PDO enters a cooling phase after 1945 as given by the PDO index (e.g. Trenberth, 2015), by this analysis most shifts had happened earlier. The ocean based step-like shifts are not seen again in this shape until the more wide spread event of 1994-1997.

1967-1973: (See Figure Ch6.42). A predominantly Southern Hemisphere event that was detected as a change in zonal mean air temperatures (24°S-44°S) circa 1968 along with contemporaneous step-like changes in minimum temperatures over SE Australia (Jones, 2012). The year also coincides with a step-like change in mean sea-level air pressures in Southern Australia and in rainfalls over SW Western Australia (Hope et al., 2010). A change in rainfall over South Africa about the same time may relate to teleconnections with the Indian Ocean (Richard et al., 2000), or to the SE Atlantic via atmospheric teleconnections (Reason et al., 2006).

1974 -1979: (See Figure Ch6.43). This is the period of the well documented Great Pacific reorganisation, nominally 1976, (Trenberth, 1990, Trenberth and Hurrell, 1994, Minobe, 2002, Overland et al., 2008) after which the temperatures globally show a still continuing upward trend.

The Indian Ocean shows a rather diffuse but upward step with increased trends whereas the mid-latitude, Eastern Pacific shows a general negative trend. Land changes show after 1978 in Eastern Australia, the Indonesian archipelago and, the sub-polar area of central North America and around the Ural Mountains, all as up-steps with intensification of prior warming.

1983-1988: (See Figure Ch6.44). Previously, associated shifts in the North Pacific have been documented (Hare and Mantua, 2000), and an extensive bio-physical shift has been shown in the North Atlantic (Reid et al., 2015, Beaugrand, 2004). During 1983 there is a distinct equatorial up-shift of the tropical East Atlantic with an increase of a prior slight warming trend. In 1985, the US North West, 35°N-55°N, shows an up-shift of which the central portion is not strongly supported by ANCOVA, with the outer portion continuing to trend upward. During 1987and 1988, Europe and much of Asia from 90°E at the same latitudes shows up shifts, with continued warming which is less prominent over Asia.

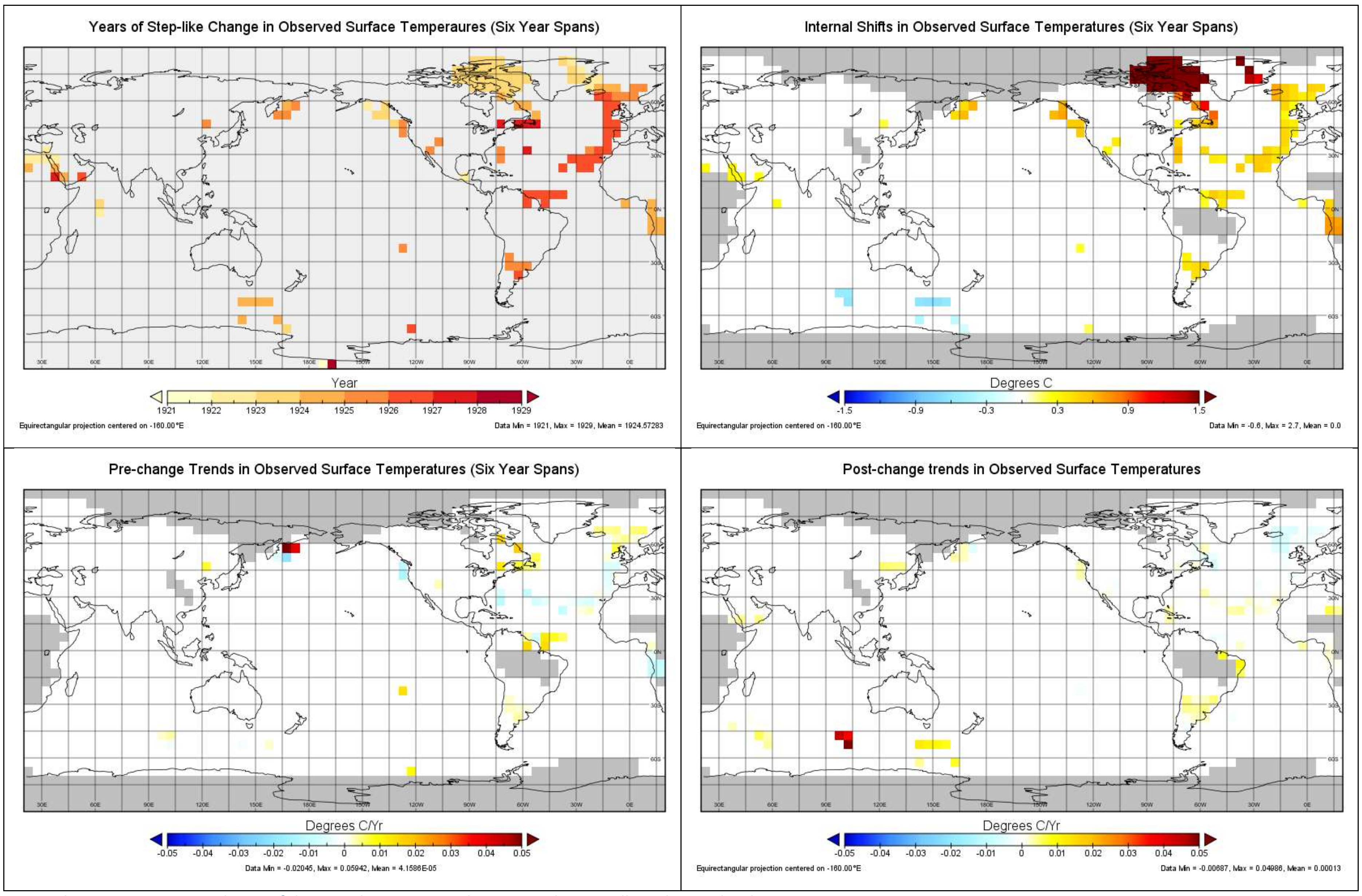


Figure Ch6.40: 1922-1928. Evolution of the step-change events that occurred principally in the North Atlantic circa 1925.

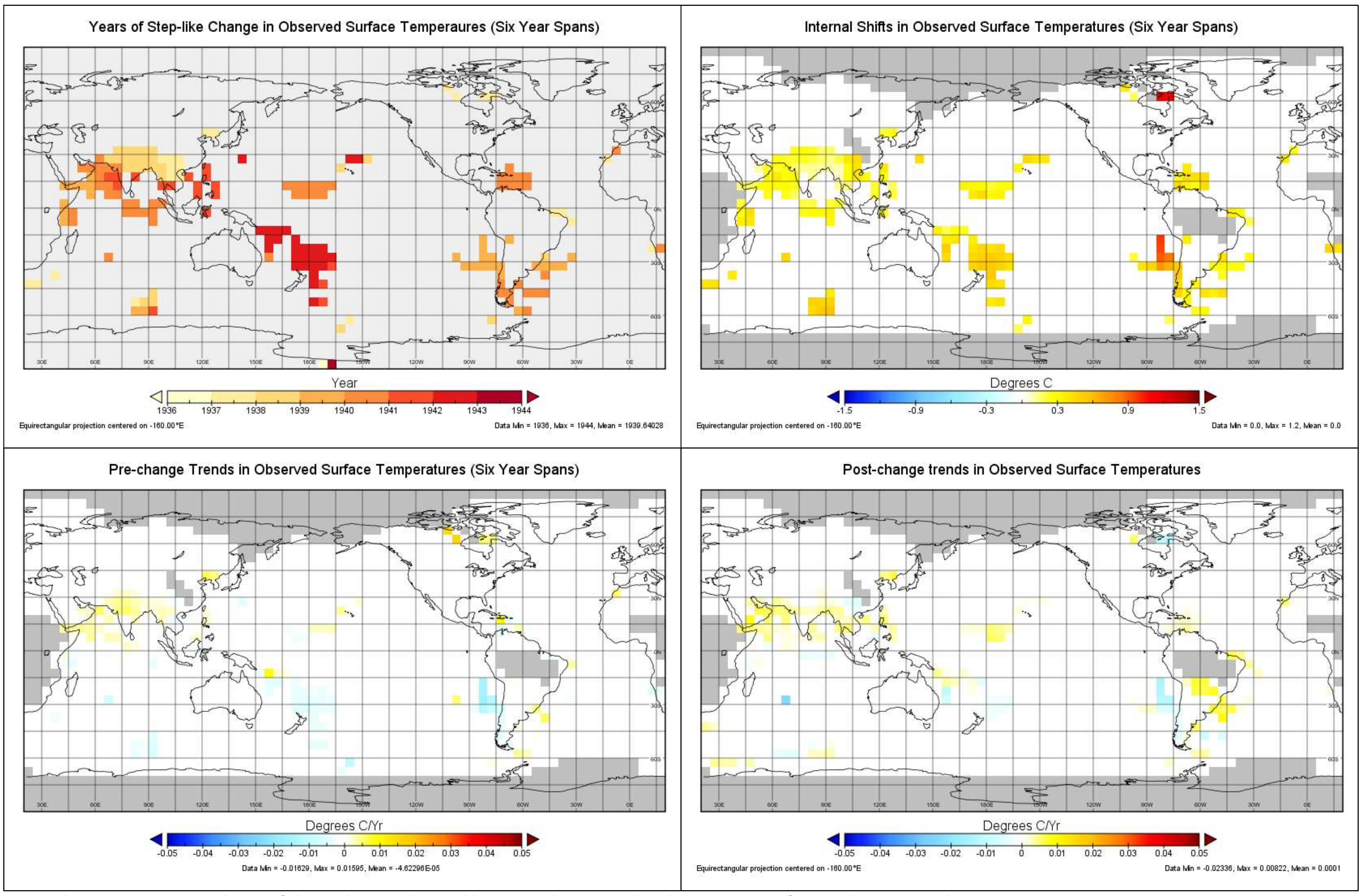


Figure Ch6.41: 1937-1943. Evolution of the predominantly step-like events that occurred in the Western Pacific, mainly in 1939 and 1941.

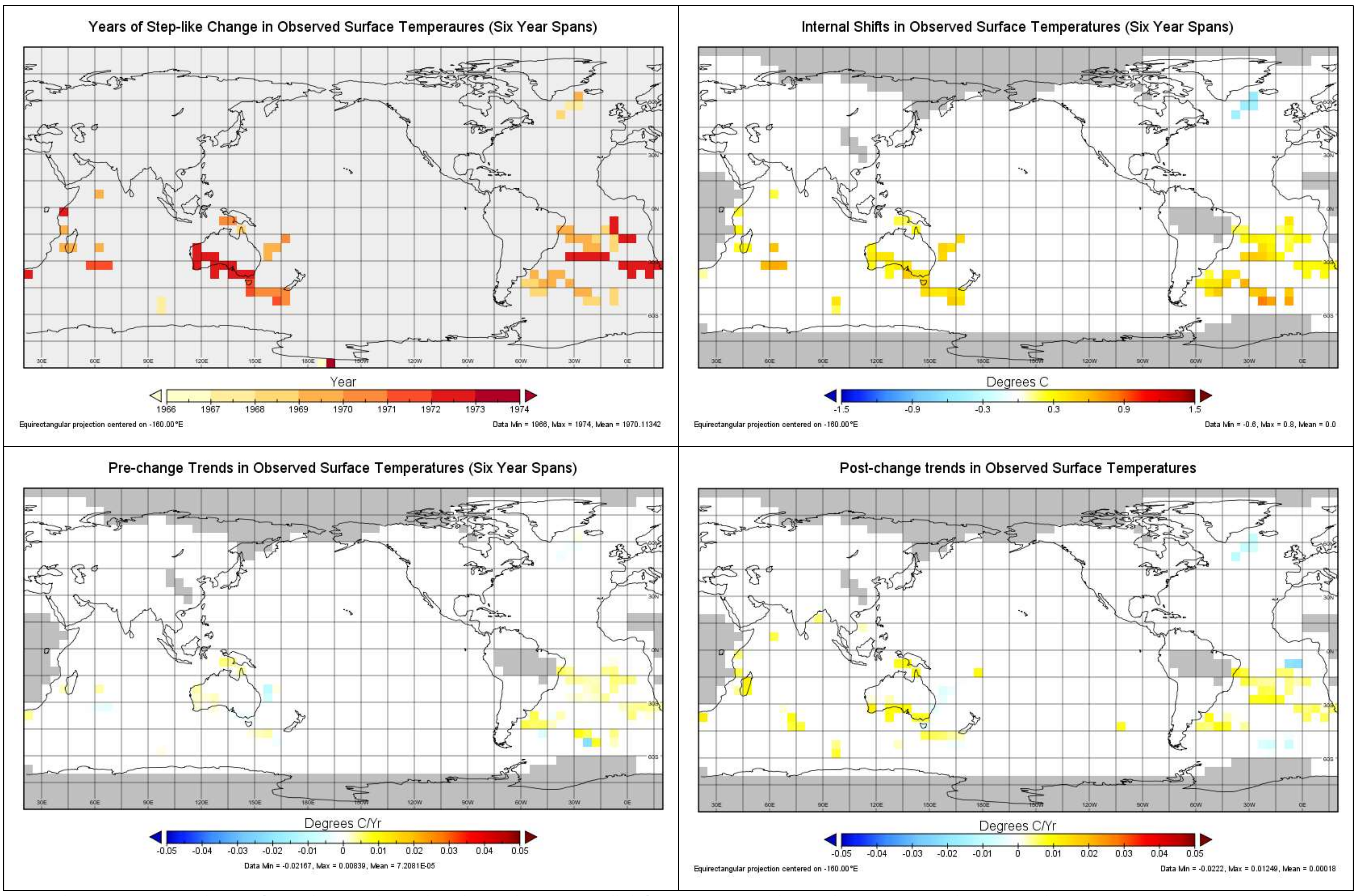


Figure Ch6.42: 1967-73. Evolution of a predominantly Southern hemisphere event. All shifts are positive and most show an increased warm trend.

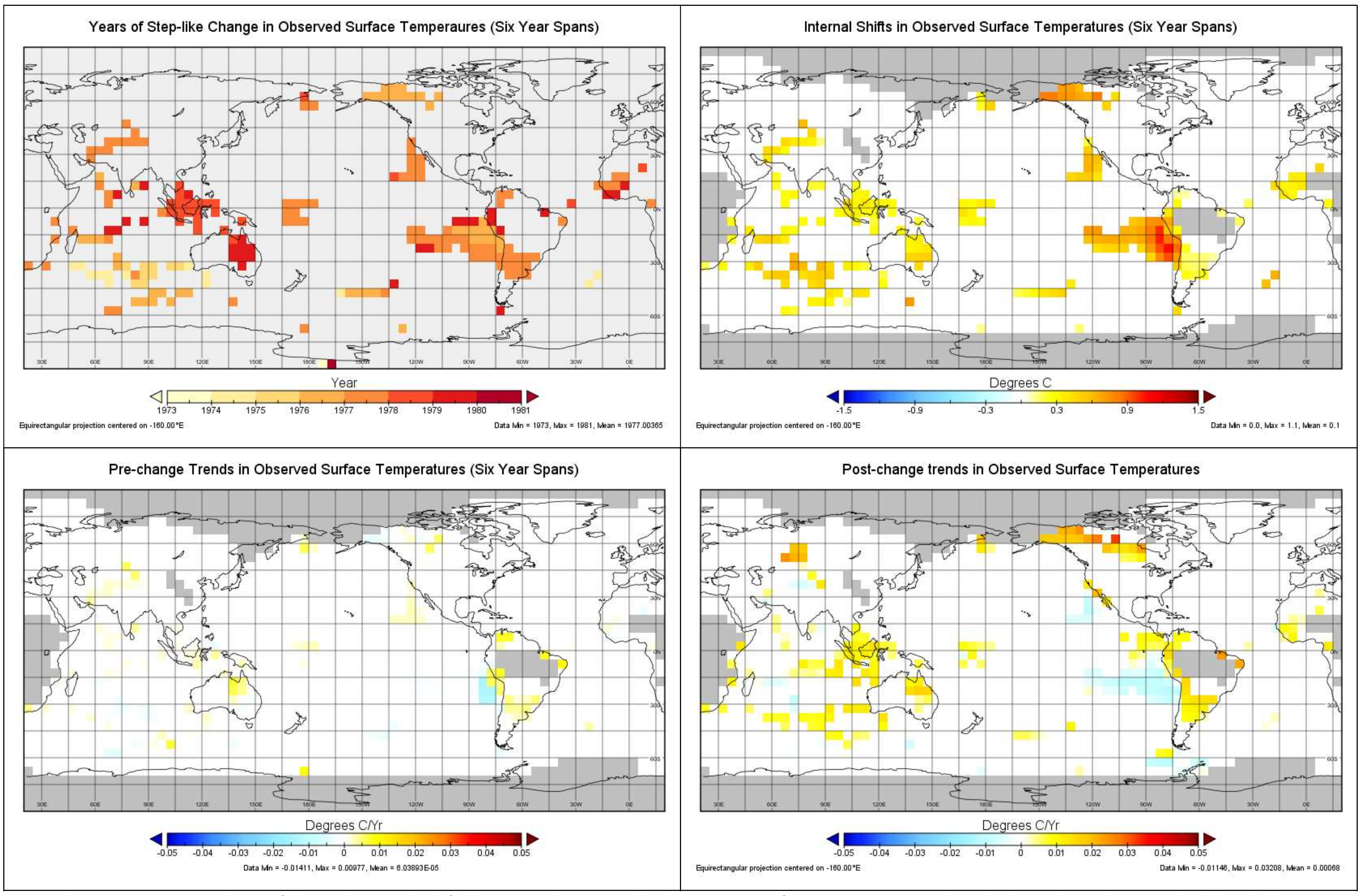


Figure Ch6.43: 1974-1980. Evolution of the predominantly shift-like events that occurred in the Eastern Pacific. Ongoing associated trends changed mostly positively although the Eastern side of the South Pacific shows a cooling trend.

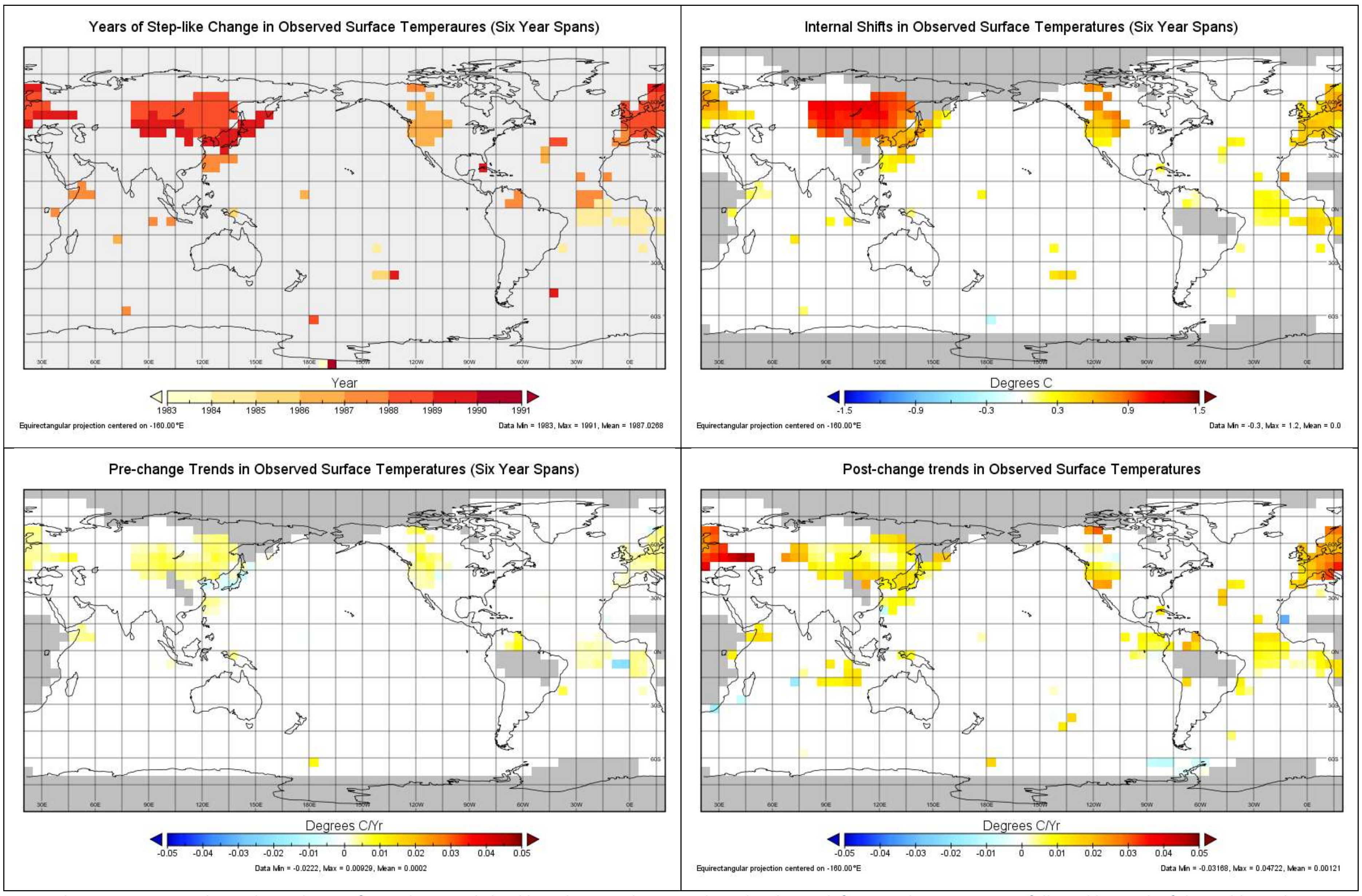


Figure Ch6.44: 1984-1990. Shows the evolution of a predominantly land based event. The tropical South Atlantic shifts up and starts to warm followed by much of the mid-latitudes Northern land surface.

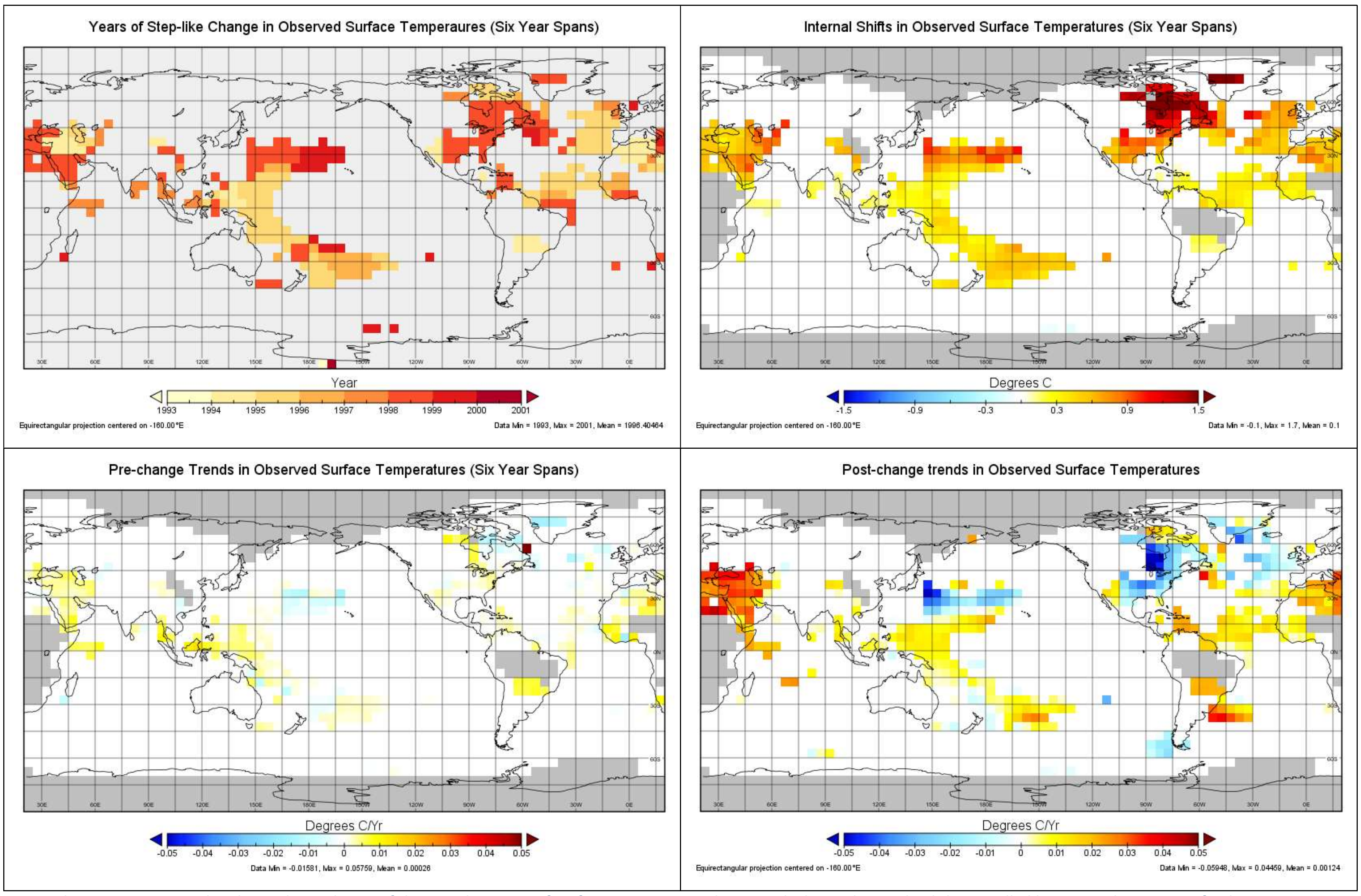


Figure Ch6.45: 1994-2000. This shows the evolution of the complex series of surface temperature changes which may be involved the so-called hiatus. Early West Pacific changes are followed by later land and ocean mid latitude changes.

1994-2000: (See Figure Ch6.45). This corresponds to a wide spread event which has been variably interpreted in a large number of papers. For example, as one of two 20th century slowdowns (Trenberth and Fasullo, 2013, Trenberth, 2015) or as an artefact of statistical methods (Rajaratnam et al., 2015). By this spatial method it shows a substantial and complex evolution. In keeping with the previous two events all of the changes involve positive internal shifts; the bulk also showing strong changes in internal trends, many of which are however negative. The Eastern North Atlantic crescent of 1925 is present again in 1994, possibly more extensive. These years both correspond to a change in sign of the AMO from negative to positive. The Western Pacific changes include the areas of the Western Pacific from 1937-1943. As stated, the earlier date may just precede the change from warm to cool phase of the PDO and this later one correlates well with the same change. The land based changes are at the same latitudes in both cases although, rather than being over India as previously this time they are over the Middle East and Northern Africa. Also two areas in 1994 and 1995 in the central South Pacific align with areas identified by altimetry as part of a spin-up of the South Pacific gyre and suggested as an associated phenomenon (Roemmich et al., 2007).

Section 2: Vertical ocean structure

An examination of globally averaged ocean temperatures at 100m and 700m depth, was presented as a case study in Chapter 3 (MSBV). I look in detail at the step-like changes in vertical ocean structure at the same scale as the previous section.

A considerable body of literature supports the notion that decadal variability modes (e.g. PDO) correspond to variations in ocean heat uptake (e.g. Watanabe et al., 2013), and in the vertical structure of the oceans (Drijfhout, 2018) as well as variations in circulation patterns. Since a change in vertical ocean structure at least regionally, may follow from changes in circulation patterns, it is of interest to see if the patterns in the changes of relationship between shallow and deeper ocean temperatures relate to those of surface temperatures.

Spatial analysis of the Ocean 100m and 700m temperatures

Data was available from 1955 onwards, although temperatures at 2000m are clearly unsuited until much later and hence not used. Four shifts of the second half of the 20th century are analysed for comparison with the temperature analysis in Section 1 (above). The same intervals are shown with the same colouring. In the following figures, the points at which either the 100m temperatures shift as judged by the 700m temperatures, or vise-versa are shown in the top pane, colour coded by year. Then in the middle, the internal-shift at 100m is on the left pane with the post-change internal trend on right. The bottom panes are the internal shifts in temperature, and post-change internal trends for 700m temperatures. All points detected are shown.

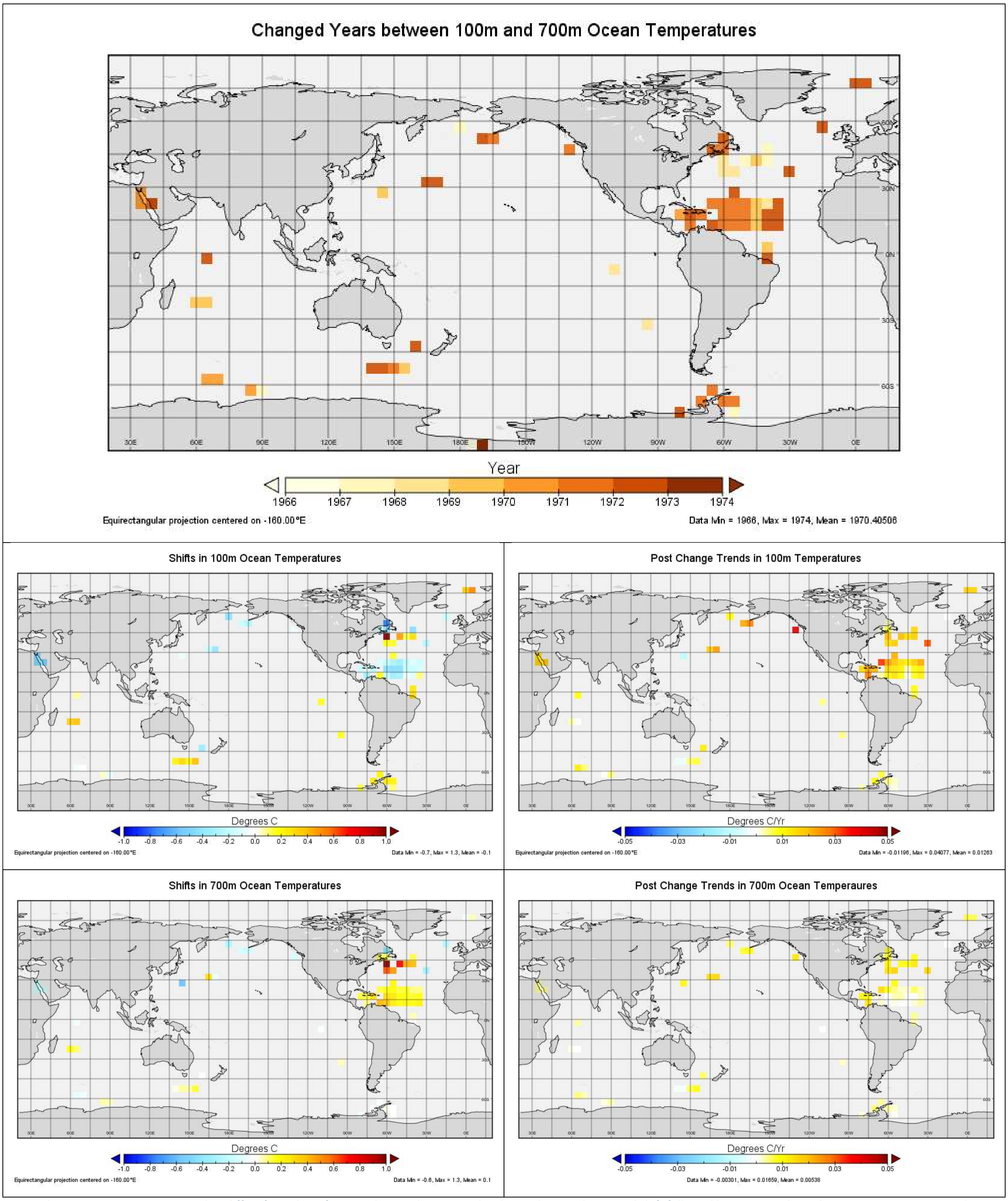


Figure Ch6.46: 1967-1973. A pattern that differs from that of Figure Ch6.42. The SW North Atlantic shows a downward shift followed by an upward trend in 100m temperatures and an upward shift in 700m temperature.

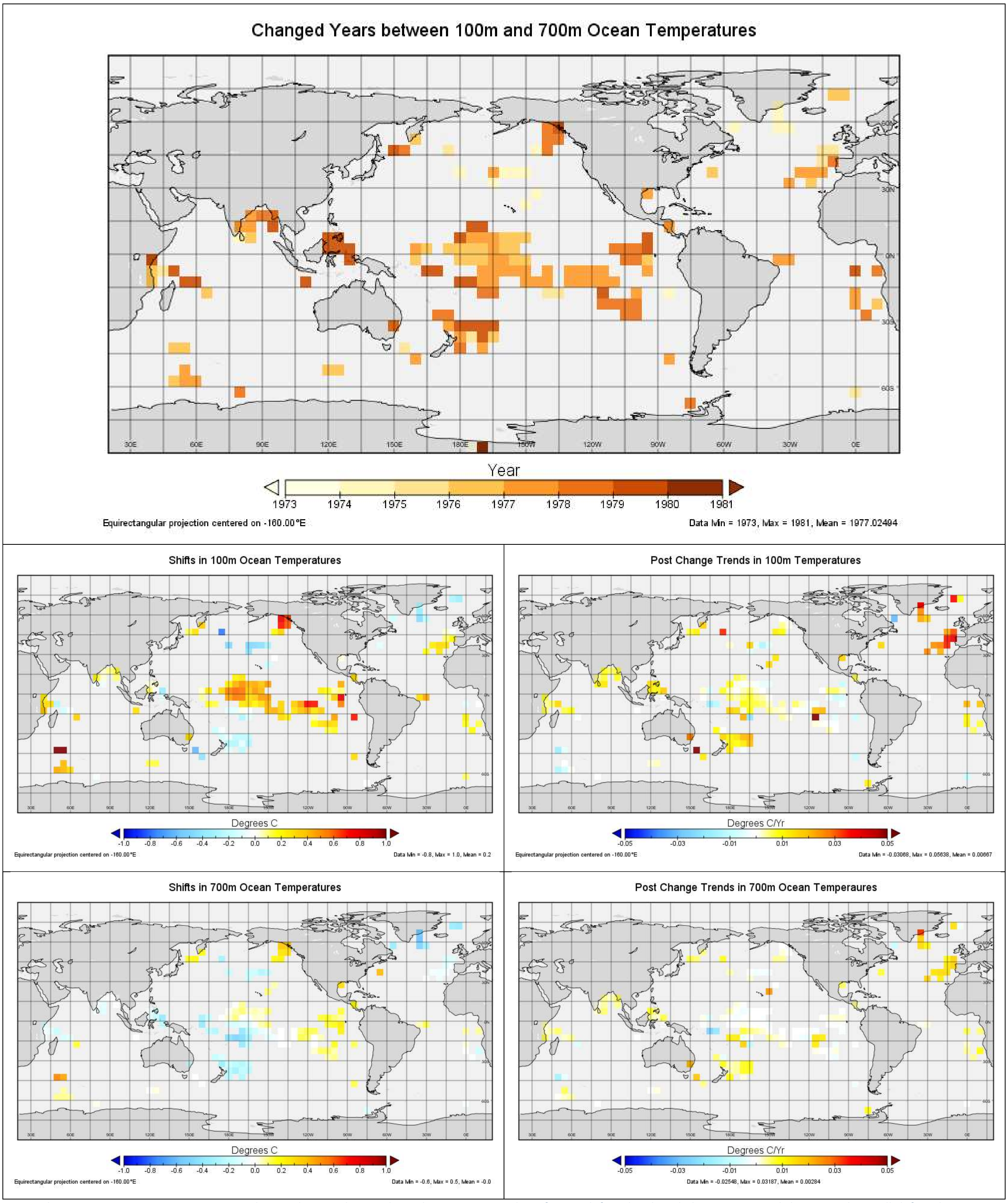


Figure Ch6.47: 1974-1980: Compare to Figure Ch6.43. Changes show in the mid-Western tropical Pacific as a shift with only a slight ongoing trend change, and a small shift with ongoing warming in both layers in the Eastern North Atlantic.

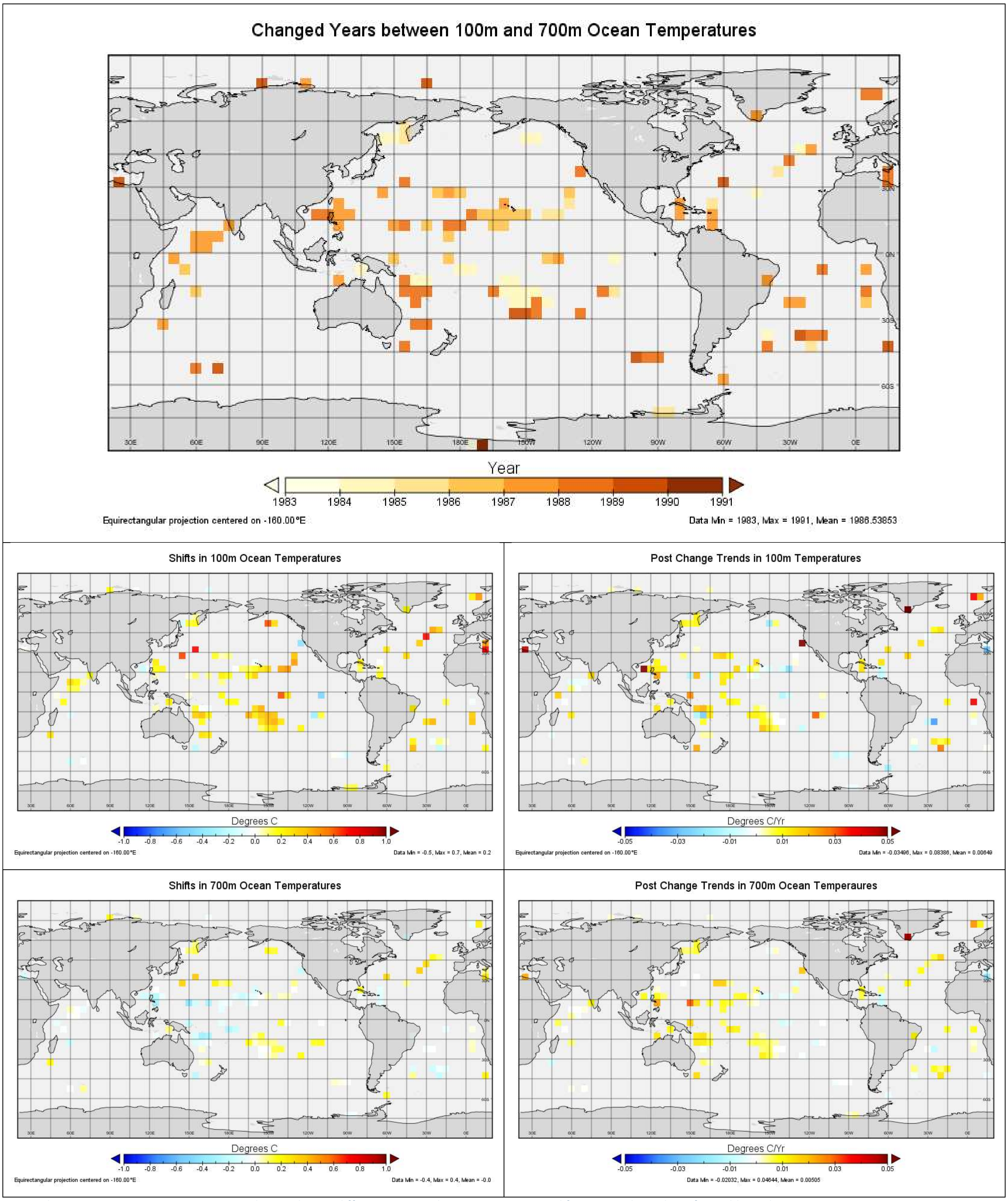


Figure Ch6.48: 1983-1990. Compare to Figure Ch6.44. This is a diffuse pattern with some more organised features in the mid-Pacific, mid-latitudes.

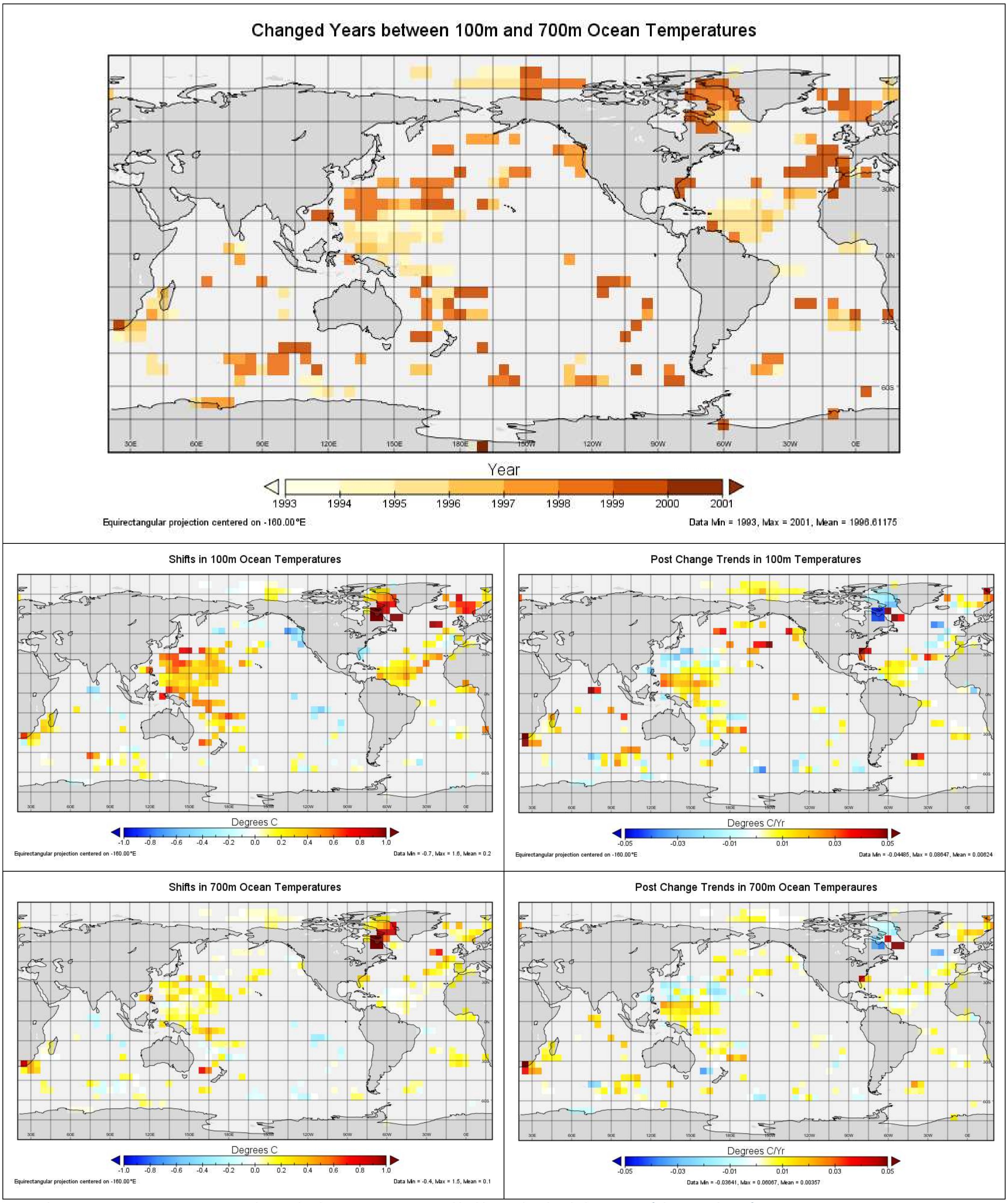


Figure Ch6.49: 1994-2000. Compare to Figure Ch6.45. The changes are extensive and highly correlated with the pattern of change in the surface temperatures.

1967-1973: (See Figure Ch6.46). The patterns contrast strongly with temperature patterns shown in Figure Ch6.42 above. Surface changes are predominantly over Southern Australia and the South Atlantic, but ocean temperature shifts showing a diffuse pattern in the Northern Pacific and a coherent pattern in the West of the North Atlantic. The North Atlantic pattern shows a downward shift at 100 metres with subsequent warming, and an upward shift at 700m, also with slight subsequent warming. The location of the region is within the North Atlantic surface “hotspots” (see Figure Ch6.39 above).

1974-1980: (See Figure Ch6.47). A prominent feature is the area from about 170E to 150W between 15N and 15S. This is mostly not represented in the surface temperatures (see Figure Ch6.43 above). The region to the East of the Pacific only approximately coincides with the more prominent region in Figure Ch6.43. The small region in the East of the North Atlantic is also included in the 1994-2000 event in surface temperatures (Figure Ch6.45).

1983-1990 (See Figure Ch6.48). This mostly shows a very diffuse pattern. Whilst the pattern of surface temperature response shown over the surface in Figure Ch6.44 is firstly in the equatorial South Atlantic and later over the Northern land masses, the ocean response is rather diffuse with the South showing up-shifts in 100m and to a lesser extent 700m temperatures and slight ongoing positive trends.

1994-2000: (See Figure Ch6.49). By contrast, with the prior period, the surface temperatures shown in Figure Ch6.45 match contemporaneous patterns in the vicinity of the Western Warm pool and Eastern North Atlantic with the addition of more diffuse changes in the Southern Hemisphere Oceans. In the Pacific, both variables show an up-step and both then show ongoing downward trends in the Northern mid-latitudes. This is matched in the surface temperatures. The prominent feature on the Eastern North Atlantic also matches the timing and, as in the Pacific, an up-shift in both variables with the more Northern part tending to trend downwards thereafter. The organised changes in temperature profile in the North Atlantic appear along the location of the poleward moving surface flow of the thermohaline circulation in the North Atlantic (Rahmstorf, 2002, see Box 1).

It appears that of the four events shown here, the two major post WW2 surface temperature events that are associated with a change of sign of the PDO (circa 1976 and 1996), are also associated with changes of ocean temperature profile in the Pacific. The other two (circa 1968 and 1986) by contrast, show differing patterns to their surface temperature counterparts, and various atmospheric mechanisms have been proposed. The first of these, already noted above, is associated with rainfall changes over Southern Australia. Frederiksen et al. (2011) suggest the rainfall changes are related to changes in storm tracks and attribute these to changes in the Hadley circulation. Several atmospheric mechanisms have been proposed for the circa 1986 event. Reid et al. (2015) suggest a rapid recovery following a transient cooling due to the El Chichón volcanic eruption, Lo and Hsu (2010) suggest interaction between the Arctic Oscillation (AO) and the PDO, and Xiao et al. (2012) suggest atmospheric structural changes associated with northern SST.

Section 3: Spatial Analysis of AOGCMs

In order to further address Severe Test 2 of JR2017, it is of interest to know if GCMs give similar results spatially to observations up during the period of observations. The previous zonal analysis of climate models suggests that step-like change dates from an ensemble of GCMs aligns with observed dates of documented bio-physical regime changes, especially post WW2. The analysis also shows that for models forced by RCP8.5 the MSBV was pushed beyond its design limits, especially post 2040. Therefore this section combines RCP8.5 runs of a selection of climate data, but considers only changes points within the current observational period. Another related point of interest is whether what appears to be East/West differentials in the number of shifts in the Pacific is also seen in GCMs.

The aim of this section is to address …

1. Whether any of the signature patterns of shifts determined from observations are replicated in any GCMs.
2. If so, to determine if the signature patterns of regional changes, postulated to be related to known internal variability modes, coordinate in the models in the same way and with similar timing. This is of interest because GCMs, by design, are free to develop their own internal states.

   Factors may be combinations of …

   a. inherent properties of gridded climate models in general
   b. specific models or model families
   c. stochastically determined initialisation states
   d. well and poorly modelled behaviour expressed regionally

A detailed examination of these questions and factors is well outside the scope of this thesis, and hence an indicative analysis was performed.

Large numbers of model realisations are available. A reasonable heuristic is required in order to (a) represent regional, timewise variation in models and observations at a scale suited to rapid screening, and (b) provide a ranking scheme that represents similarity, and (c) allow intra-model exploration.

Additional considerations regarding GCMs

Global climate models generate their own internal states, based on their internal physical assumptions, the prescribed atmospheres, and purely stochastic parameters. Due to the manner of their initialisation there is no reason to believe that any particular model will enter the observational period with specific internal states. (For further, see Box A6.2.1 in Appendix 6.2). Rather, the models overall can be shown to have particular spatial patterns of variability analogous to particular observed patterns of variability (Eyring et al., 2016), (their Figure 8 is shown as Figure A6.2.58 in the Appendix). But the timings and phases of those modelled variability modes vary greatly. The ensemble has shown that particular observed change-point dates in zonal and global temperature series align with observations (see Chapter 5). This alignment does not however mean that the internal structures of the model necessarily align with reality. Models may form (a) completely novel structures, (b) build up a set of similar sub-structures (e.g. PDO) but align them differently with different phases, (c) build imperfect versions of the same structures with the same relationships, (d) combinations of (a) and (b).

Producing a similarity index

I wanted to select a representative climate model that had similar regional scale variability to observations. This requires some form of index based on representative regions, and had to allow for differences in timing between model and observed versions of the same events. Ideally such an index would be based on a coarse scale analysis that was informative about any differentials, East/West or North/South in ocean basin or continental responses.

Continent and ocean basin quadrants

The last part of Chapter 5 suggests that a coarse 6x8 grid of the Earth surface is sufficient to capture broad spatial variation in step-like behaviour. It also suggests that ocean and land responses differ, so that a coarse scale scheme should separate land and ocean. Hence for this work I define five ocean basins and five continental land masses. These are North and South Pacific, North and South Atlantic and Indian oceans; plus North and South America, Eurasia, Africa, and Australia. Each is divided into North and South, East and West quadrants. This provides 40 quadrants, 20 land, and 20 ocean. Additionally, for each basin or continent, a whole of basin, and East and West hemi-basins were produced.

Correlation between each model and observations

The observations plus the RCP8.5 runs of the models listed in Table Ch6.20, all regridded to 5°x5° were analysed by MSBV in Rapid Halting mode, and annual patterns of step-like change were extracted. For each of the forty quadrants, the proportion of step-like change-points found in the corresponding 5°x5°grid is computed. In order allow for differences in timing, the time-series were first "blurred" by convolving with eleven element weighting kernel [0.1,0.2,0.3,0.5,0.3,0.2,0.1].

This produces an annual time-series for each quadrant that relates to the proportion of step-change (see Figure Ch6.50). The time-series for each model in each quadrant are regressed against observations, producing a correlation index (the term "index" is used due to the blurring) and the models are then ranked on their correlation index in each quarter.

The result of all of this over each of the forty quarters (ocean basins/continents) is a table of correlation indices and rankings that indicate the degree with which each model predicts the amount of step-like change at about the same time and in about the same region.

At this stage quadrant indices from each continent/ocean basin can be tabulated. There would be ten tables similar to Table Ch6.20, below. In this example IPSL-CM5A-MR would be judged as having the most acceptable overall similarity to observations.

Two final similarity indices were computed. The first is simply the mean correlation index over all 40 quadrants. The second is the grand mean of the ranks.

IPSL-CM5A-LR was second ranked by both similarity indices, and for this model four replicate runs were available. Therefore it was selected for closer examination.

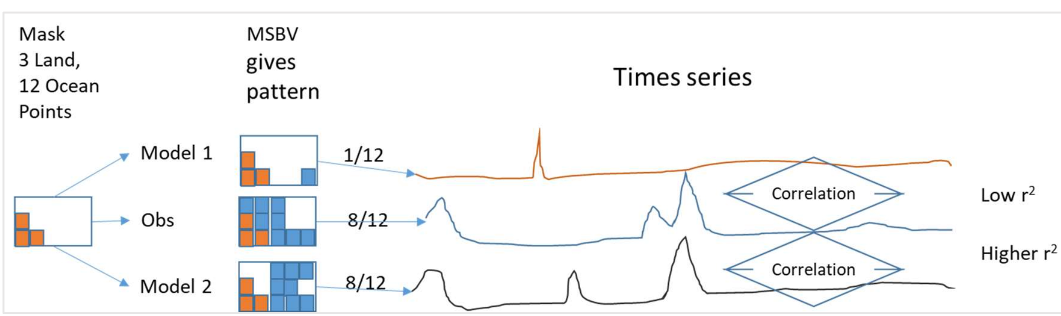


Figure Ch6.50: Schematic of the method for correlating finer generalised patterns between observations and climate models. Here, an area of interest (A quadrant of an ocean basin) is represented as a rectangle, and the smaller squares represent individual grid-points on the 5°x5° grid at which the MSBV was run. In this case orange represents land and is masked out. The particular feature observed is present in 8 of 12 available grids. At the same time it is present in only one from Model 1, but eight of model 2, albeit distributed differently. When this is repeated over time three time-series result. The two models are ranked in this quarter on the degree of correlation between their time-series and observations.

Table Ch6.20: Sample ranking analysis for one continent, South America. Correlation indices are shown for each quadrant as well as their relative ranks. Overall, no model does particularly well, and the selected model in this case is one that does better in the west than any other model.

South America	**Correlation index**				**Rank of Correlation Index**				**Mean correlation-index**	**Rank of mean**
	NE	NW	*SE*	SW	*NE*	NW	*SE*	SW		
ACCESS1-0	0.16	0.07	0.03	0.00	7	7	7	16	7.4	9
ACCESS1-3	0.00	0.02	0.00	0.03	15	15	16	5	10.2	15
bcc-csm1-1	0.12	0.17	0.04	0.00	10	2	6	17	7	8
bcc-csm1-1-m	0.32	0.08	0.00	0.01	2	6	15	11	6.8	5
BNU-ESM	0.20	0.07	0.07	0.04	6	8	4	4	4.4	2
CanESM2	0.00	0.04	0.00	0.01	18	13	17	10	11.6	18
CESM1-BGC	0.13	0.08	0.00	0.00	9	5	18	19	10.2	15
CESM1-CAM5	0.34	0.12	0.02	0.01	1	3	13	8	5	3
IPSL-CM5A-LR	0.25	0.04	0.00	0.05	4	14	19	3	8	10
IPSL-CM5A-MR	***0.10***	***0.22***	***0.05***	***0.20***	***11***	***1***	***5***	***1***	***3.6***	***1***
IPSL-CM5B-LR	0.00	0.00	0.14	0.00	19	18	1	13	10.2	15
MIROC5	0.00	0.00	0.02	0.00	17	19	8	14	11.6	18
MIROC-ESM	0.01	0.04	0.02	0.01	14	12	9	9	8.8	12
MIROC-ESM-CHEM	0.03	0.01	0.10	0.01	13	16	3	12	8.8	12
MPI-ESM-LR	0.16	0.01	0.02	0.03	8	17	11	6	8.4	11
MPI-ESM-MR	0.29	0.05	0.13	0.00	3	11	2	18	6.8	5
MRI-CGCM3	0.05	0.08	0.01	0.05	12	4	14	2	6.4	4
NorESM1-M	0.00	0.06	0.02	0.00	16	9	10	15	10	14
NorESM1-ME	0.22	0.05	0.02	0.02	5	10	12	7	6.8	5

Table Ch6.21: Model ranking for overall correlation of predicted portion of shifts that occurred within each E/W, N/S sectors of ocean basins and land masses. Each model was ranked in two ways. Method 1 ranks each model within each sector, then computes a mean ranking for the land mass/ocean basin and ranks that mean to give a secondary ranking; finally repeating the ranking again at global level for a tertiary ranking. The second computes a mean R2 for all sectors per model and ranks that for a primary ranking.

Name of climate model	Mean Ranking Across All Land masses and Ocean Basins	Ranked by mean ranking	Mean R^2 of all sectors for each model	Ranked by mean R^2
ACCESS1-0	8.333	11	0.13	4
ACCESS1-3	9.400	17	0.05	19
bcc-csm1-1	7.378	6	0.13	5
bcc-csm1-1-m	7.822	9	0.10	8
BNU-ESM	9.533	18	0.06	17
CanESM2	7.956	10	0.09	9
CESM1-BGC	7.222	5	0.09	10
CESM1-CAM5	7.044	3	0.11	7
IPSL-CM5A-LR	***6.511***	***2***	***0.14***	***2***
IPSL-CM5A-MR	7.156	4	0.12	6
IPSL-CM5B-LR	9.378	16	0.05	18
MIROC5	9.111	13	0.09	11
MIROC-ESM	9.133	14	0.07	16
MIROC-ESM-CHEM	10.578	19	0.07	15
MPI-ESM-LR	8.867	12	0.07	14
MPI-ESM-MR	**6.200**	**1**	**0.13**	**3**
MRI-CGCM3	9.200	15	0.07	13
NorESM1-M	7.800	**8**	**0.15**	**1**
NorESM1-ME	7.600	7	0.08	12

General findings

During production of the index it became apparent that patterns in East Pacific are universally poorly modelled, and that for no ocean basin nor continent were all four quadrants well predicted.

Table A6.1.38 in Appendix 6 summarises the overall findings of the observations/models correlations by quadrant. No model correlated well over all four quadrants of any region, in fact nine different models were preferred for at least one region (Appendix 6, Table A6.1.38). In the Pacific, which is of special interest due to its complexity, models performed quite poorly, especially in the Eastern side for which there was essentially no correlation over any model (see example in Figure Ch6.51). This is of some concern since by many measures the Pacific is a major player in climate modulation and change. Visual inspection of the diagnostic plots shows the principal reason is that for most models, a step in the West is accompanied by one to the East, whereas in observations this is not the case at all. Rather steps in the East are rarer. Figure Ch6.52 illustrates other differences between more correlated and non-correlated models at sub-region scale using two sample models. Table A6.1.37 in Appendix 6.1 gives more detail.

The full results of the ranking study do not support the idea of coarse scale quadrant style analysis being synchronous with observations, especially in that the observed differential of the Pacific East and West is replaced by more synchronised behaviour in models. This can be seen by looking for correlated surface responses using Empirical Orthogonal Teleconnections (EOT) (Dool et al., 2000), where the first mode correlations of most models show equatorial East/West communication that is absent in observations (see Appendix, Figure A6.2.57).

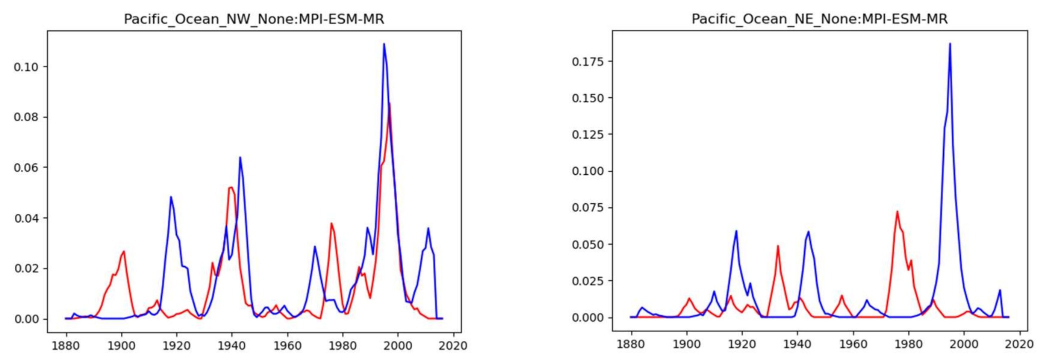


Figure Ch6.51: Typical example of the difference between model predictions of regime changes in the Eastern and Western sides of the Pacific Ocean. MPI-ESM-MR is the GCM which gives the best overall correlation with observations. Here, the Western side of North Pacific (combined NW and SW of North Pacific) is compared to Eastern side for the nominally preferred climate model. Red denotes observed, blue the GCM. The model behaviour in the West (correlation index 0.48) closely captures two of four changes, and possibly predicts the 1976 event too early. However the model's Eastern side (correlation index 0.05) much more resembles its own Western side than it does observations. The quite striking contrast between observed Eastern and Western sides is not well replicated in models.

However the question of whether models show some of the signatures of the AMO and PDO, and at what intervals was explored. Three replicates of the chosen model were then analysed by rapid MSBV at 5 degrees grid using RCP8.5 forcings and examined for correlation with the three signature dates (1976, 1986 and 1996) and timings of the AMO and PDO signatures (Crescent in East North Atlantic, and Tropical Pacific).

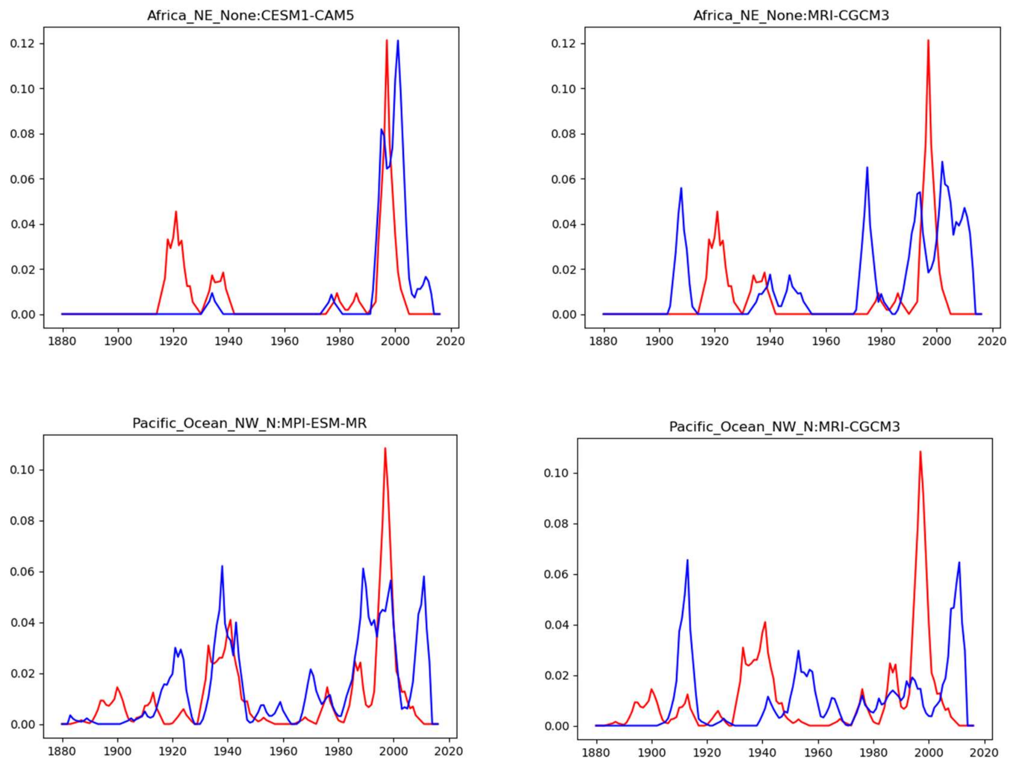


Figure Ch6.52: Sample plots: Good and poor predictions of one land and one ocean region shown by correlation index of quadrants. Left is the highest ranked model for each quadrant, right is an uncorrelated one. Top row NE Africa (CESM1-CAM5 correlation index 0.34, MRI-CGCM3 correlation index 0.01), Bottom Row North-West North Pacific (MPI-ESM-MR correlation index 0.42, MRI-CGCM3 correlation index 0.001)

Features of individual runs IPSL-CM5A-LR. RCP8.5

Note that in common with all models which express an analogue of the Pacific shifts of circa 1976 and 1996 these models are also generally more uniform across the tropical Pacific.

r1i1p1

This replicate evolves very quietly.

- 1898. An AMO like signature in with an up-step and rising trend.
- 1920-21. A tropical Pacific event and propagation 1921 with up-step and rising trend.
- 1975/80 – No analogues of observed reorganisation
- 1983. Southern Ocean/South Atlantic event and Northern North Atlantic, up-step and rising trend.
- 1986/8 – No analogue of Northern mid-latitudes/sub-polar land response
- 1994/8 – No analogues of so-called hiatus.
- 2009/10 Mid-latitudes Southern Atlantic and Eastern tropics, up-step and rising trend.

- 2013-16 Huge tropical break (2014/5 formation of NE Pacific warm blob), up-step and rising trend.
- 2017. South East Australia, up-step and rising trend.

r2i1p1

Generally less quiet and with more similarity to observations.

- 1864. An AMO like signature up-step mixed trends
- 1908, Pacific event up-step small positive trend change
- 1970, A small event near the North Pacific sub-tropical gyre with up-step and increased trend.
- 1974/5, Northern mid-latitudes/sub-polar land response, up-step and rising trend (similar to an event circa 1986 ascribed by Reid et al. (2015) to Hadley cell expansion).
- 1978-1981. Complex event with up-step and rising trends commencing with an up-step in the South Pacific ENSO tongue and N East Atlantic up-step and rise in trend from near zero rate to 0.02 degrees/year.
- 1979 Complex release over WPWP and NE Pacific blob plus SW North Atlantic which propagates in 1980 and also a zonal Southern South Atlantic shift, in 1981 culminating in the AMO signature.
- 1985-89 zonal mainly land, similar to observed and as ascribed by Reid et al. (2015).
- 1993-1998, another complex series of events, all showing up-steps and mostly increased trends, that commences with Atlantic Southern Ocean, and in 1994 the tropical Pacific 160E-130W, the Southern Ocean and SW Indian. 1995-1998 similar to observed with more zonality and indications of teleconnection.
- 2001 Southern coastal Australia, S Indian and S Pacific near the sub-tropical gyre. This shows propagation in 2001
- 2008 Sothern Tropical Atlantic and Mid-latitudes over Western Europe. The latter then propagates East over the Pacific in 2009/10 and towards Alaska in 2011/12
- 2010. An event near the North Pacific sub-tropical gyre with up-step and mostly increased trend.
- Next large event is 2025 with a zonal pattern but with up-steps over the tropical Pacific which show reduction in the rate of warming away from the centre of action. This is the earliest that this mixed response (a feature of the so-called hiatus) is seen.

R3i1p1

- 1918 A small event near the North Pacific sub-tropical gyre with up-step and increased trend.
- 1927, 1929, an AMO signature like event with up-step and essentially no change of trend.
- 1935 Step-like heat release in the mid Tropical Pacific with little change of trend,
- 1939, 1943 Similar response in more Eastern tropical Pacific and in 1943 in the South Pacific ENSO tongue
- 1968 SW Western Australia shows an up-step and increased trends.
- 1976. No analogue of the Great Pacific reorganisation.
- 1981 a small East Pacific ENSO tongue event with a possible sub-tropical North Atlantic heat release that propagates East.
- 1982 a distinctive zonal continuation in the tropics and sub-polar regions with up-steps and strong upward trend changes.
- 1983 tropical Atlantic shows an up-step and strong trend changes.
- 1988 a prominent SE Northern Pacific warm blob appears as an up-step with warming increases.
- 1994/8. No analogue of the so-called hiatus event.
- 1993-2000 Scattered Pacific and progression into continued small scattered events
- 2015. Small possible spin-up over the North Pacific sub-tropical gyre with an up-step and continued warming.
- 2017 a prominent NW North Pacific mid-latitude warm blob.

Characteristic events

Possible North Pacific Sub-tropical Gyre spin-ups.

In observations, spin-up of the South Pacific sub-tropical gyre is seen in 1994 (Roemmich et al., 2007). No such signature is seen in r1i1p1 but they are seen in r2i1p1 (in 1970, and 2010), and r3i1p1 (in 1918, and 2015), which are both well out of phase with observations.

AMO

The AMO index was published for one replicate of this model, r1i1p1, (http://www2.geog.ucl.ac.uk/~ucfaccb/PaleoVar/IPSL-CM5A-LR/amo.timeseries.png). The event which I note in 1898 as a signature corresponds to a prominent negative to positive warm phase shift, but the index shows similar strong changes in the mid-1930s and mid 1980s without any such signature step change. R2i1p1 shows signature events in 1864 and 1981. R3i1p1 in 1927.

PDO

The PDO index computed for r1i1p1 (http://www2.geog.ucl.ac.uk/~ucfaccb/PaleoVar/IPSL-CM5A-LR/pdo.timeseries.png). There was a warm to cool shift in the early 1940s, a cool to warm one in the mid-1970s and a warm to cool one in the mid-1990s, with a cool to warm shift having also occurred circa 2015. The PDO index computed for r1i1p1 shows a mid 1920s cool to warm phase change, with a reverse change in the early 1930s and a cool to warm transition circa 1970 which strengthened further in the mid-1970s. A warm to cool phase shift in the late 1990s is followed by an upward inflection circa 2015. The latter part of this aligns quite well with observations.

Pacific Warm Blobs (r3i1p1)

In the observed data a possible Eastern occurrence is shown in the MSBV in 1976 along with many other features. There are no Western occurrences. A well documented event occurred in 2014 (Bond et al., 2015) but the MSBV will not show this due to the seven year refractory period. The Western location of a blob from R3i1P1 in 2017 is counter to expectation as is a SE Pacific blob – both suggesting that R3i1P1 has an out of phase Pacific circulation.

The results of the replicates examined in detail here, the surface signatures (and where available PDO and AMO indices), suggest that the model ocean circulation configurations are not close analogues of observations.

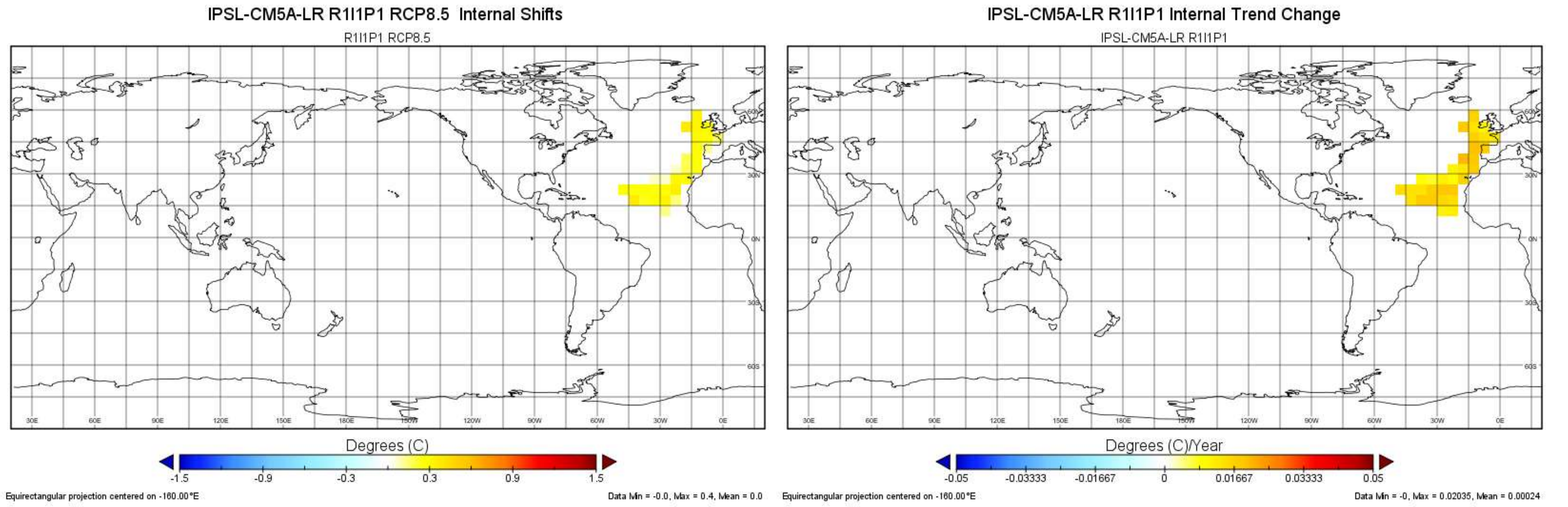
Examples

Some example spatial plots are shown in Figure Ch6. 53. These consist of paired images showing internal shifts and internal trend changes from different realizations of IPSL-CM4A-LR RCP8.5 runs.

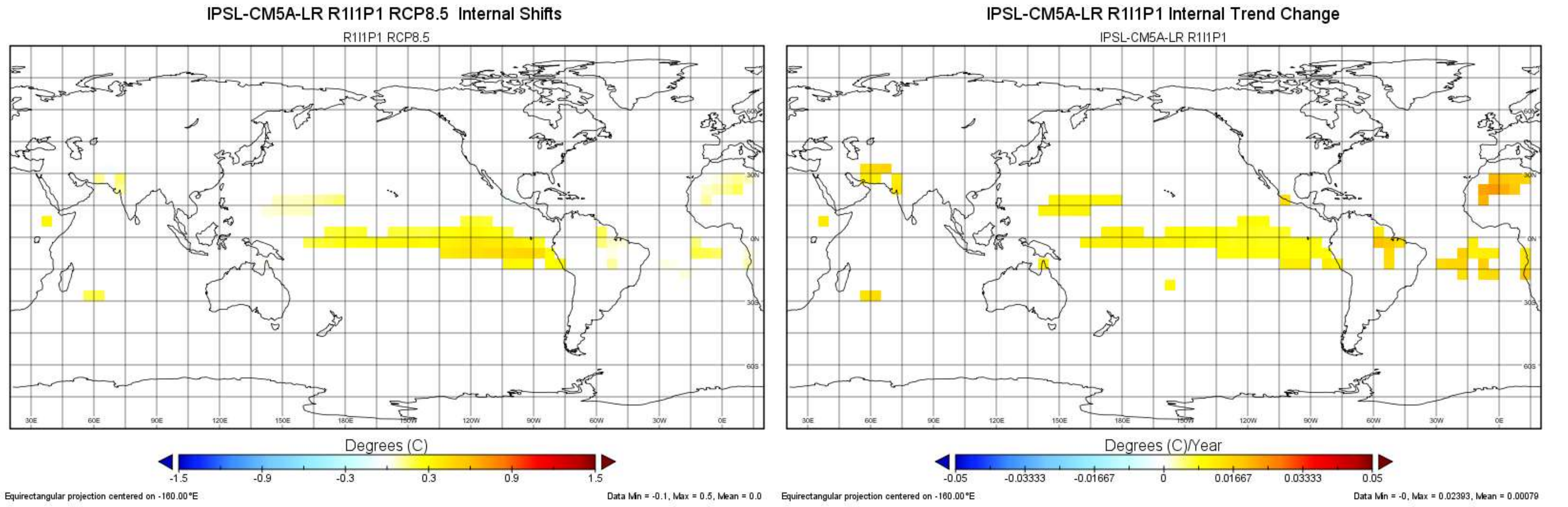
(A) r1i1p1 for 1898/9, aligning with an AMO cool to warm event, and which morphologically resembles the GISTEMP3 1924/5 (Figure Ch6.40). This latter event also aligns with an AMO cool to warm phase shift, as also a 1994 event which also aligns with the same class of change.
(B & C) are the same r1i1p1 realization, 1920 and 1921, and (D and E) 1978 and 1979 showing zonal tropical shifts, with less zonal structure in the second case. These superficially resemble the events associated with PDO changes circa 1976 (Figure Ch6.43) and 1996 (Figure Ch6.45) but are more zonal and less differentiated – GISTEMP3 shows events at the same dates but in 1976 the observed change was from cool to warm and the modelled case it was from warm to cool phase. The modelled 1979 also shows the formation of a NE Pacific warm blob which was present in observations then and in 2014.
(F) Persistent Pacific warm blob formation shown in r3i1p1. Eastern Pacific Warm Blob 1988 and (G) a West Pacific Warm Blob formation 2017. In the observations a somewhat similar Eastern Pacific

response is seen circa 1976, and in 1997 a Western event. Both are part of more complex events and both are further South.

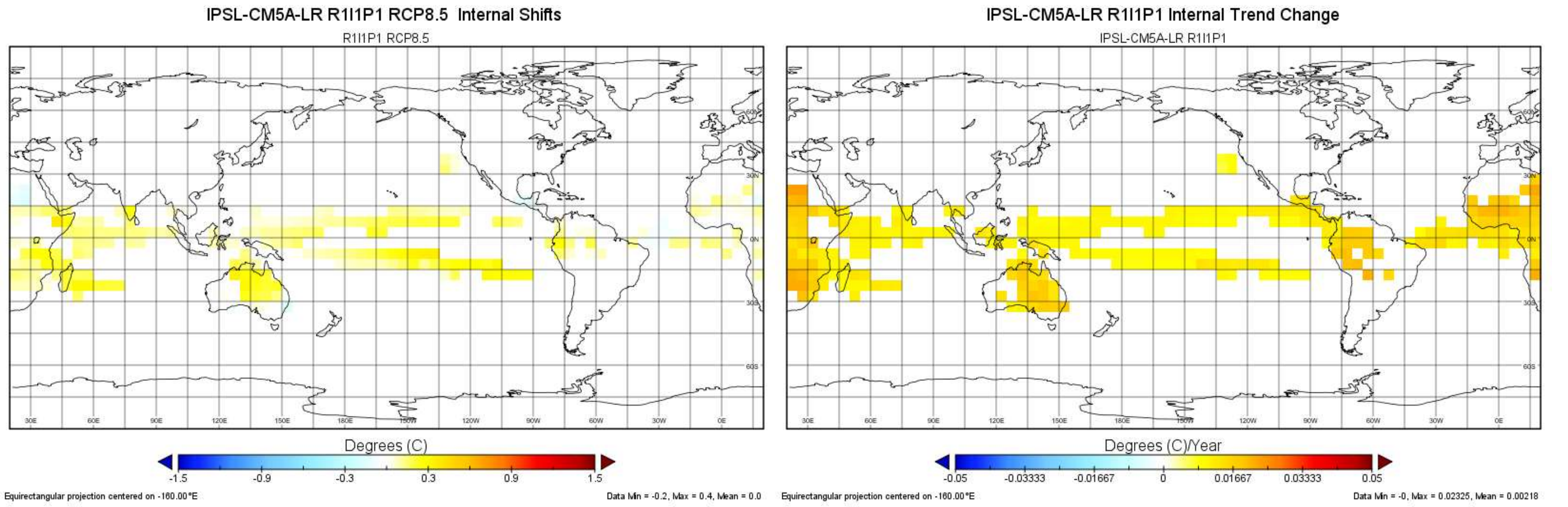
R1I1P1 Sample AMO and PDO Signatures



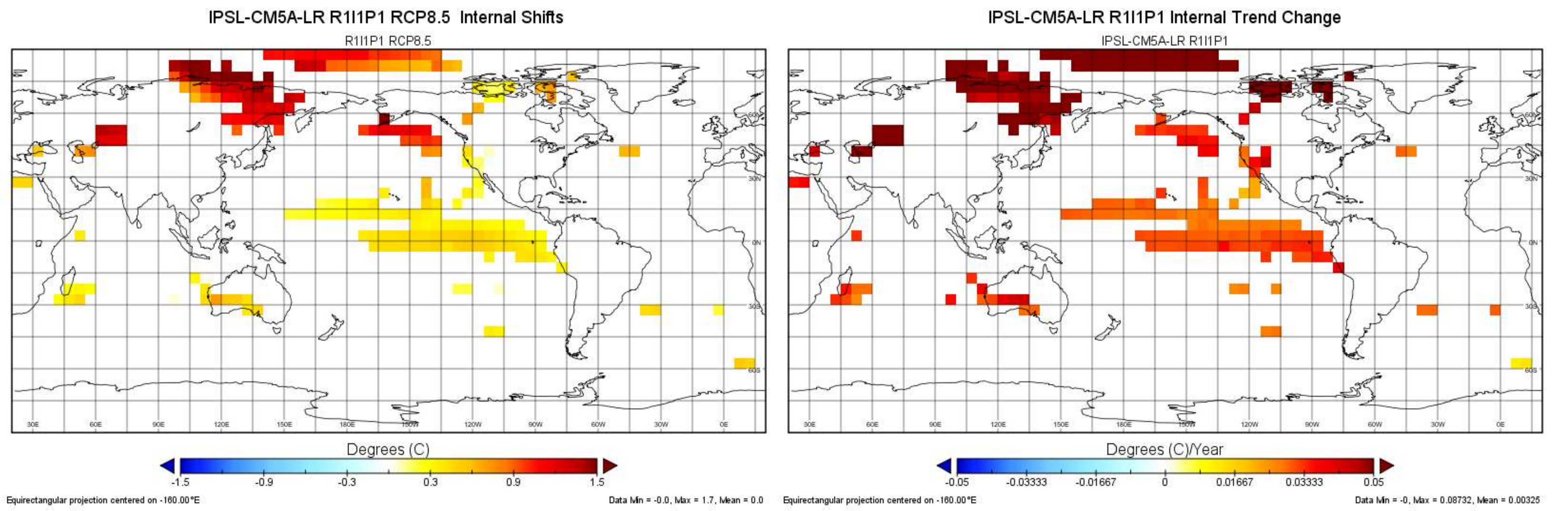
A



B



C



D

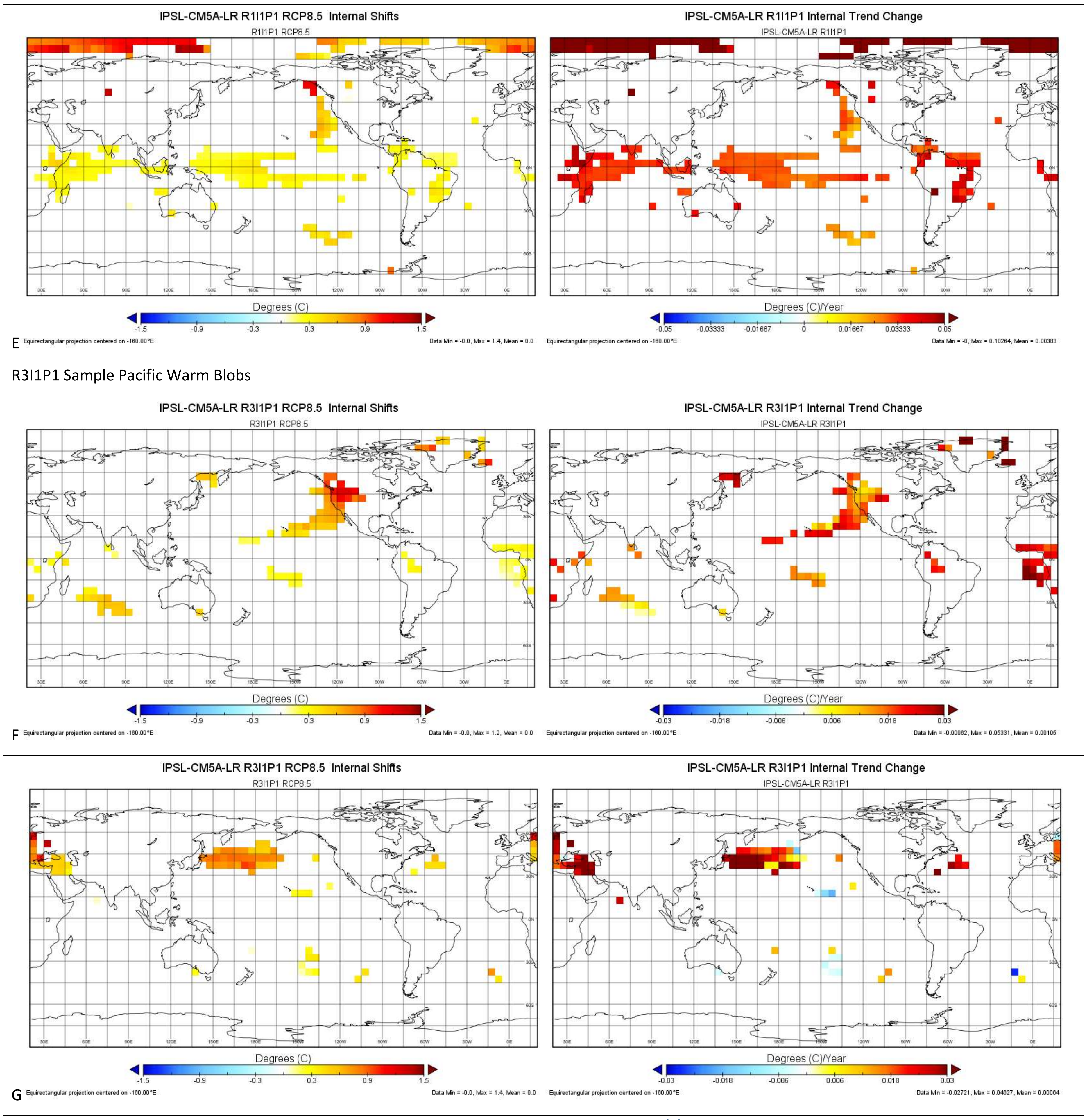


Figure Ch6. 53: Internal shifts and internal trend changes from different realizations of IPSL-CM4A-LR RCP8.5 runs. (A). r1i1p1 1898 morphologically resembles the GISTEMP3 1925 event which aligns with an AMO cool to warm phase shift, and also a 1994 event which also aligns with the same class of change. (B & C) are the same realization, 1920 and 1921, and (D and E) 1978 and 1979 showing zonal tropical shifts, with less zonal structure in the second case. (F) An Eastern Pacific Warm Blob 1988 and (G) a West Pacific Warm Blob formation 2017.

In conclusion, whilst it is probable that the IPSL-CM4A-LR replicate models show some similar surface features to observations, and that these could be diagnostic of the equivalent decadal features operating within the models, they occur out of phase with observations and with other replicates. The preliminary quadrant analysis also suggests that overall models do not closely track surface features. This gives a discrepancy between the model behaviour when examined at global and zonal scale and the detail that emerges at finer scale. It would be consistent with similar global scale features being produced by a variety of configurations of variability modes.

This study is not extensive enough to draw any firmer conclusion. However further work is warranted. A substantial body of work has been performed which attempts forecasts of future weather patterns by pattern scaling, a technique which intrinsically assumes linearity between models and observations (Santer et al., 1990, Mitchell et al., 1999). I have been involved in such work (Ricketts and Page, 2007, Ricketts, 2009, Hennessy et al., 2011, Ricketts et al., 2013). The results in this section so far do not lend support to the underlying assumptions. At the very least, models which show evidence that their decadal variability modes synchronise with those observed should be sought out and assessed.

Section 4: Study of step or shift-like behaviour at differing spatial scales.

This section consolidates the relationship between the shift and trend components at various spatial scales. Chapter 5 lays out the case for compositional misspecification springing from the aggregation of data from regions where changes occur with different timing, or propagate over time within regions, and increased contribution of shifts to total change was shown to be evidence of misspecification.

The contribution of internal shifts to total change in the GISTEMP3 data analysed in section 1 was assessed two ways. Firstly the same quadrant scheme as defined in the previous section was used to produce was used to produce weighted annual average temperature for each quadrant. There were analysed by MSBV as before. The total internal shifts and internal trends were combined by area weighting to produce estimates of global and hemispheric contributions to total warming over land, ocean and combined land-ocean areas. This gives nine estimates of shift to total warming ratios that can be compared to the similar values in Table Ch6.19 in the first section. These are shown side-by-side in Table Ch6.22 below. A students paired *t*-test for gives a two tailed *p*-value of 0.0018 on the hypothesis of no difference, and supports the notion that the data composition makes a systematic difference, in this one data set.

Table Ch6.22: Shift/Total ratios computed for 5°x5° Spatial grids and for quadrants.

GISTEMP3 From 5°x5° Spatial						From Quadrants
Location	**Hemisphere**	**Shift °C**	**Trend °C**	**Total °C**	**Shift Proportions**	**Shift Proportions**
Land	Globe	0.7	0.44	1.13	0.62	0.58
Land	NH	0.79	0.47	1.27	0.63	0.64
Land	SH	0.42	0.32	0.74	0.57	0.38
Ocean	Globe	0.5	0.13	0.63	0.79	0.66
Ocean	NH	0.52	0.15	0.66	0.78	0.65
Ocean	SH	0.49	0.12	0.61	0.8	0.66
Land/Ocean	Globe	0.56	0.22	0.78	0.72	0.63
Land/Ocean	NH	0.64	0.29	0.93	0.69	0.65
Land/Ocean	SH	0.48	0.15	0.63	0.76	0.59
Mean					0.707	0.604
Variance					0.0071	0.0079
t-Stat						4.587
P(T<=t) two-tail						**0.0018**

R.N. Jones (personal communication) recombined the observational data from up to five different data sources published in JR2017, to produce a comparison of the shift/total ratios for global versus sub-global composites (see Table Ch6.23). I have selected the warming over all available times (from 1880 to 2014). Since the underlying data are drawn from the same physical Earth, the differing data sets are treated as independent samples of the same data. Three of these had available data representing some sort of subsampling. The effect of subsampling can be calculated. The main finding is that the warming attributable to internal shifts averages out at 50% across the five global data sets, and these are always less than for 11 subsamples which average 66%. The *t*-test for zero difference with unequal variances returns a two tail *p*-value of 6.96×10^{-8}.

So for GISS data sets looking at combined land-ocean data, the shift/total ratios are, Globally, averages temperatures 0.51, zonally 0.64, for quadrants constructed from spatial grids, 0.63, and from grid-scale 0.72. A clear progression.

Table Ch6.23: Shift/Total ratios computed for single runs of and available breakdowns courtesy of R.N. Jones.

Sub-Global		Global	
Observation Set	**Shift Proportions**	**Observation Set**	**Shift Proportions**
GISS Hemispheres	0.71	BEST	0.53
GISS trop-extrop	0.64	Cowtan&Way	0.50
GISS Zonal	0.64	GISS	0.51
HadCRU Hemispheres	0.66	HadCRU	0.48
HadCRU Land-ocean	0.65	NCDC	0.49
HadCRU L-O N-S	0.68		
NCDC Hemispheres	0.66		
NCDC Land-ocean	0.71		
NCDC L-O N-S	0.64		
NCDC trop-extrop	0.63		
NCDC Zonal	0.62		
Mean	0.66		0.502
Variance	0.00086		0.000387
t-Stat			
P(T<=t) two-tail			**$6.96\text{x}10^{-08}$**

A selection of GCMs from the RCP8.5 runs for which both the global mean annual signal had been analysed and for which gridded scale data was available was also examined, this time as a paired *t*-test between the values return for each model analysed firstly as global average and then analysed spatially. In this case the computation was between 1850 and 2100 (see Table Ch6.24).

Table Ch6.24: Shift/Total ratios for RCP8.5 Global Climate Models

RCP85 Model	5°x5° Spatial	Global mean
IPSL-CM5A-LR	0.53	0.40
MPI-ESM-LR	0.58	0.27
ACCESS1-0	0.64	0.25
ACCESS1-3	0.64	0.28
CanESM2	0.54	0.40
MIROC-ESM-CHEM	0.61	0.27
BNU-ESM	0.61	0.31
IPSL-CM5B-LR	0.64	0.30
NorESM1-ME	0.64	0.45
MIROC5	0.60	0.52
MRI-CGCM3	0.63	0.23
MPI-ESM-MR	0.63	0.35
bcc-csm1-1-m	0.64	0.27
IPSL-CM5A-MR	0.50	0.28
NorESM1-M	0.59	0.32
CESM1-CAM5	0.49	0.38
Mean	0.60	0.33
Variance	0.0027	0.0063
t-Stat		10.32
P(T<=t) two-tail		**$1.76\text{x}10^{-08}$**

Test 6 of JR2017 asks whether temperature progression is more step-like or more trend like. This section has shown that not only is it more step-like but that it becomes even more so when the MSBV test is applied to data at finer scale. This is consistent with step-like shifts being present in temperatures, and being regional. The phenomena are also rather diffuse, so that simple rectangular averaging inevitably induces artefacts or at larger scale, swamps the signal.

Summary and Discussion

The first two sections show serve both to establish a sort of ontology of step change patterns – isolating meaningful spatial signatures, and to establish a link between vertical ocean structures and surface temperature expressions. A crescent like anomaly in the surface temperatures of Eastern North Atlantic in 1925 and 1994 and at no other time aligns with the transition from cool to warm phase in the AMO index. Mosley-Thompson et al. (2005) demonstrated a step-change in the correlation between the NAO and precipitation in the mid-1920s, and also in temperatures, and the AMO and NAO are linked. The 1922 step off Western Greenland has been noted before. Box (2002) suggests a correlation between the North Atlantic Oscillation (NAO) and Western Greenland temperatures (see his Section 4.6). The next time such a shift occurs is in 1997 when there is also a much more wide spread response, although Lloyd et al. (2011) attribute the early to mid-1990s warming of the Greenland waters to the AMO. Nakamura (2013) noted a sudden de-correlation between Greenland sea surface temperatures and NAO phase in the late 1970s suggesting a link to the AMO. The crescent shape is also seen in some of the climate models, especially the ones selected by the quadrant correlation index, and possibly aligned to the AMO there also.

Surface temperature change patterns from 1931 to 1941 appear within the much more wide spread patterns of the 1994 to 1999 event. Although the PDO enters a cooling phase after 1943/5 as given by the PDO index, by this analysis most shifts had happened earlier accompanied by an apparent appearance at the surface of the waters of the West Pacific Warm pool. If the PDO index reflects the circulation of the Pacific then the warm pool release preceded the change of the PDO to cool phase.

1986 This event, by contrast to the other analysed here, is primarily a land based event. Reid et al. (2015) performed a comprehensive analysis of biological and physical variables to conclude that a synchronised change occurred in the mid-1980s suggesting that the El Chichón volcanic eruption of 1982 was contributory to a physical state change. Their method detects changes in SST in 1986 at basin level for which there is little evidence in the spatial analysis, which however finds strong Northern land involvement, suggestive of a predominantly atmospheric state changes, for example Hadley cell expansion.

A series of interlinked events seem to define a signal of the change of phase of the PDO. The pattern of change-points in the spatial surface temperatures shows indicates Pacific changes. The vertical ocean structure in tropical Pacific indicates the sign of the phase change. This is shown by change-points in the relationship between 100m and 700m global averaged ocean temperatures, and in a spatial analysis of the latter two variables.

These two variability modes have often been suggested as governing regional climate, although the NAO (Tsonis et al., 2007) is also a possible factor.

A principal finding from Section 1 is that when internal shifts and internal trends are separated, the picture for oceans is one of heat release in a step-like fashion and with much of the mid-latitude oceans showing nearly compensatory cumulative trends serving to restore thermal equilibrium. The picture for land is different since land trends are mostly quite uniformly positive.

Other findings are that step-like regime shifts …

1. are rarer over land than ocean and overall they contribute much of the observed warming,
2. are commoner in the Southern Hemisphere than the Northern, again mostly oceanic,
3. predominate the Arctic although this analysis under-samples that area due to using only uninterrupted data,
4. are responsible for most of nett ocean surface warming, especially extra-tropically,
5. occur more often on Western basin boundaries, and in the Southern Ocean which has annular flow,
6. are rare to absent near the polar and subtropical gyres which are convergence zones.

At grid scale some of the apparent time-wise precision of global and zonal analyses disappears; events do not happen simultaneously but in a rapid cascade of smaller, linked events.

It is also notable that if one attributed the circa 1939 event to a warm to cool PDO regime shift (more usually attributed to the early 1940s), then in the Pacific, warm to cool regime shifts (1939, 1997) are marked by surface step-like changes further West than the cool to warm one (1976). The effect of the 2015 cool to warm shift remains to be analysed by these methods when more data has accumulated[5].

Overall it would seem that in addition to larger events, there are regional ones that involve shallow ocean circulation unsupported by deeper ocean structures; Indian Ocean, and Western boundaries, and the Southern Ocean (1976 for example). The circa 1976 and 1997 events are accompanied by Pacific changes involving the surface and deeper changes in the

[5] But see the late part of the discussion. There appears to be a corresponding sharp upward step.

Atlantic. The 1986 event remains mostly a land only event more suggestive of an atmospheric reorganisation with minor involvement in the tropical Eastern Pacific and Western North Atlantic not accompanied by ocean structural change.

Some changes are purely regional, for example in the Southern Ocean and south Indian Ocean, some are highly structured and suggest vertical and horizontal changes in the oceans that involved a complex of events, and some are potentially more atmospheric like.

The fact that in 1994/5 the two way analysis of 100m/700m ocean temperature shows indications of warming at the surface, plus the 100m and 700m oceans over the WPWP suggests that the thermal structure warmed right down with warmer waters intruding deeper and to the surface. This would suggest an emplacement of warmer waters a little before the onset of the El Niño and subsequent change of phase of the PDO that is associated with the so-called hiatus. This is consistent with a trigger mechanism related to the Western Pacific Warm Pool mentioned .in the introduction to this chapter.

Section 3 of the chapter was dedicated to bounding the utility of climate models as a tool for making physical inferences about the projection of decadal scale variability into the future. A naïve prediction from models requires that the models somehow represent, not just the separate systems that give rise to variability, but maintain their phases relative to each other and in synch with observations. A conclusion of the previous chapter is that not only do ensembles of global climate runs show sufficient coherence with reality that one can suggest a step change in global temperatures circa 2015, but that one can suggest this be due to differential changes in zonal temperature progression over land and oceans. The ensemble showed alignment with known variability modes. However the reason for this is unclear given the manner by which climate models are initialised.

A method was given for comparing models with observations of reality at finer than global scale, to see if they concur at sub-basin sub-continental scale. There is a systematic variation in these assessments. No model performs equally well in all four quadrants of all five land masses and five ocean basins. A survey by EOT analysis supports the notion that climate models communicate across the Pacific in a manner that reality doesn't (see the example Figure A6.2.57 in Appendix 6.2).

The principal finding in this part of the chapter is that global climate models do behave as if they have internal states reminiscent of decadal variability modes but that variability is not aligned to observations.

This does not invalidate the conclusion of the previous chapter that climate model ensembles tend to predict an upward and persistent shift in temperature in the latter half of the current decade. It does mean that in future work, more careful selection of climate models and features is needed.

The beginning of a type of ontology of sub-regional features of climate regime shifts has been made. For example the crescent-like feature on the Eastern side of the North Atlantic in 1925 and 1994, and aligned with the AMO, is associated with derived AMO indices in some models. The extra-tropical region that surrounds the ENSO-tongue, seen circa 1976 has two features. Firstly, despite the concentration in the literature on the Northern Pacific, the shifts are more extensive in the Southern hemisphere. This also meshes with the zonal analyses. However the quadrants analysis finds them, in fact they are the only shifts detected in the Eastern non-polar Pacific. The second is that by contrast with the models, the region of the ENSO tongue is rarely detected as part of the shift (IPSL-CM5A-MR is somewhat exceptional in this).

The important findings again include that land and ocean based regime shifts differ: land less frequent, more zonal, ocean more variable and more regional. The zonal averages of step counts in Figure Ch6.36 highlight a long suspected difference in the roles of the Northern and Southern hemisphere. The earliest post WW2 response to forced warming is in the Southern mid-latitudes.

A fundamental methodological finding is that the spatial averaging of data is of critical importance, and multiple scales should be used. Since the oceans circulate, the influence of an upwelling event may rapidly smear. Consideration of a rotationally referenced coordinate system may assist.

Section 4 provides an important result in that it is clear that processes are more shift-like at smaller scale. This is considered in the next chapter, the Discussion.

Chapter 7: Discussion

Introduction

This work addresses the nature of previously observed abrupt decadal regime shifts in Earth's climate, and whether there is evidence of these shifts being affected by increases in atmosphere and ocean temperatures due to greenhouse emissions. This is important for climate science, because it goes to the contribution of forced warming on modes of natural variability and any role they have in moderating climate and natural heat flows. It is important for practitioners of climate change risk, because unaccounted for changes to such moderation constitute a risk in their own right.

The work follows on from, and adopts the methods in, previous work of Jones (Jones, 2012, Jones et al., 2013), where abrupt changes in temperature and rainfall records regionally were linked to large-scale and global regime changes in a range of bio-physical systems. It contributed to recent published work that further refines the research by grounding the linkage between the physical climate and statistical induction about it; and then severely testing competing hypotheses of the interaction between forced warming and natural variability (JR2017).

The thesis has focussed on three aspects of the science.

1. How to best measure regime shifts in time-series. What are the limits of the chosen detection method (MSBV)? What are the effects of autocorrelation, red-noise/random-walks, or extreme ongoing trend on the reliability of the MSBV?
   Chapter 3 presented the MSBV, and sensitivity testing of it with brief case studies. Within the limits established by the sensitivity tests the MSBV is found to be fit for purpose.
   Chapter 4 presents a set of post-detection tests which probe the data for features that could deceive the detection method. These can be condensed into two summary metrics, (a) a segment classification, that gives a guide as to whether the data segment within which a change is found shows evidence of underlying non-stationarity, (b) the CP-Index which gives a graded guide to possible false positive change-points. When tested on real data, it further established that the data was generally suited to the analysis.
   Furthermore, when some types of data consistently return many possible false-positives, for example RCP8.5 climate averages whether detected by MSBV or

structural-change, it indicates that the data (and therefore the climate) may no longer be obeying the same statistics. Further investigation is thus indicated in a way that it could not be if the data were not tested-post detection.

2. What evidence is there that regime shifts are real phenomena? To what extent are they influential and to what extent does climate change proceed step-like as postulated by Jones and latterly, others (Bartsev et al., 2017, Belolipetsky et al., 2015, Saltykov et al., 2017, Yan et al., 2016). JR2017 proposed and demonstrated six severe tests that distinguished H1 (forced warming and natural variability change separately) and H2 (they interact) (Chapter 2). This work tests and supports four of those tests, the other two being out of scope.
    a. Comparisons of change-point return periods in zonally averaged, modelled pre-industrial, historical and future climates, matched with observations, showed that during the historical period regime change occurred more frequently as warming progressed, changed to include more zones, and changes occurred simultaneously over more of the Earth. Pre-industrial model runs showed much longer return periods, in general, with only the Southern mid-latitudes showing much activity. 21st Century runs strongly contrasted, with short return periods, and wide-spread shifts. When, under RCP2.6 warming slows, so too the return periods become longer, especially mid-latitude land based ones. This addresses Test 1 of the six severe tests and supports H2.
    b. Examination of the RCP26 and RCP85 ensembles of global climate models showed that clusters of regime shifts in the models corresponded both by date and by zone with those observed. This addresses Test 2 of the six tests in favour of H2.
    c. Analysis of the spatio-temporal patterns of ocean temperature regime shifts, showing that they correspond to surface temperature patterns for those dates associated with regime shifts in the PDO, supports Test 5 of the six tests in favour of H2.
    d. Test 6 of the six, was also resolved in favour of H2, in that step-like behaviour predominates trend-like behaviour. The analysis related to this also clearly indicated that the degree to which step dominates is inversely proportional to the spatial scale over which the climate data are analysed. This is consistent with regimes occurring with somewhat varying timing within areas of roughly half the size of an ocean basin or continent. The analysis of Chapter 6, Section

1 also showed that regime shifts often spread out from an initial event over time, and often that changes over land follow ones involving oceans. And it further showed that trends over ocean, once shifts are accounted for, contribute very little to the GMST. This is very much in keeping with heat being released in bursts accompanied by step-like shifts, and the more stationary periods maintaining equilibrium.

   e. Overall, the reduction of non-stationary behaviour at finer scales, along with the increase in shift to trend ratios, and reduction in AR(1) autocorrelation also suggests that the composition of climate signals is a source of apparent autocorrelation in global and hemispheric data

3. Do regional regime shifts relate to climate variability modes?
   a. Chapter 5 consists of a detailed investigation into the zonal variation of step-like changes in the observed and modelled temperature record. It clearly adds to the knowledge about the non-uniformity of decadal variability attributable to anthropogenic climate change. In the observed and modelled surface temperature records, as demonstrated by step-like shifts, the progress of climate change differs by latitudinal zones with distinct differences over land and over ocean.
   b. Chapter 6 concentrates for the most part on a detailed analysis of spatial patterns formed by abrupt step-like changes in surface temperature observations, ocean temperatures at depth and to a lesser extent global climate models.

Advances

The literature review brings a focus onto statistical climatology. JR2017 built on previous work concerning statistical induction (e.g. Mayo and Spanos, 2006, Roush, 2010) and outlined the Theoretical-Mechanistic/Statistical-Inductive framework (TMSI). This was of use in the latter stages of the thesis, acting as an organising principal, and is treated in more depth due to my contributions to the paper.

The literature review also tabulates different approaches to validation of methods and validation findings. The first is of relevance to those investigating and designing statistical detection and analysis techniques; the second to those using them.

Inspiration is drawn from work on misspecification testing (Mayo and Spanos, 2004). The resulting methods proved to be useful in showing that the MSBV was fit for purpose, and very importantly that at finer scale climate data is deterministic, and thus in sharp contrast of the concerns of Rudnick and Davis (2003) regime shifts can be reliably separated from red-noise in oceans.

MSBV

The work detailed in Chapter 3 advances the objective assessment of multiple change-points in climate time series, when the principle feature being sought is step-like – an abrupt shift and a sustained change. The Multistep Bivariate test (MSBV) is built on the Maronna-Yohai (MYBV) test, and differs from other published methods in several respects.

- Whilst not unique in this, the MYBV test uses a reference series. The use of multiple resampling with a random reference is unusual, and adds sharpness. This relates to stochastic resonance (Moss and Wiesenfeld, 1995, McDonnell and Abbott, 2009) where injection of noise, apparently paradoxically, enhances feature detection, an area of research which has its roots in climatology (Benzi et al., 1981, Benzi et al., 1982). This is the first use that I know of where stochastic resonance is achieved by injection of noise into an auxiliary variable.
- The MSBV differs from the principal multiple change-point methods that have been published. Two of these allow for change-points that include, but are not focussed on, discontinuities. These are "structural change" (SC) from the econometric literature (Zeileis et al., 2001, Zeileis et al., 2010), and "changepoint" (CP), which has its basis in oceanography (Killick and Eckley, 2011, Killick et al., 2010).
- If differs from another method that was published during this work, the "multiple STARS" method (Reid et al., 2015) which was developed from the STARS method of (Rodionov, 2004).
- It also differs from another method, also published during this work, often discussed under the name named "change point" (Cahill et al., 2015), which implements a version of non-disjoint segmented model (one inapplicable to the problem of shift detection). It is based on CP-regression (Carlin et al., 1992). Cahill et al. (2015) also ignore the issue of data dredging (Epstein, 1982, Solow, 1987), whereas the approach used in the MSBV and subsequent analysis does not.

The MSBV builds a disjoint stepwise regression, splitting segments on change-points, and then refines it in a causal order. The statistical model is deemed complete on each iteration when no change-point is found in any segment ($p<0.01$). This is a less general halting criterion and generally terminates sooner than other methods, with mostly less complex models. Given that

AIC based termination used for structural-change methods is known to be prone to overfitting (Spanos, 2010), the relative parsimony of the MSBV is a virtue. The MSBV avoids an assumption of variance stationarity between change-points in the data (by contrast with the AIC).

The work and approach of Jones differs from many others. This thesis has explored that body of work and sought the boundaries of applicability. The step/shift detection method was developed from the description in Vives and Jones (2005) and in consultation with Jones. The method had been limited by a requirement for researcher choice, and some simple rules were developed to capture these choices. Reviewer’s comments have also stimulated an investigation into the rationale for using level shift detection in preference to the more usual trend-change or step-and-trend approaches. It has been shown in the literature from time to time that the problems of shift detection and trend-change detection are different (Jarušková, 1997, Kaiser and Maravall, 1999, Wang and Wang, 1994, Wu, 2005) but methods continue to be proposed that do not take this into account. Several approaches to understanding this have been demonstrated, one using a limit case of a sigmoid curve, another from numerical methods theory (Stoer and Bulirsch, 2013).

Characterisation of change-points

Advances have been made in addressing the issue of mismatches between data and the selection of statistical models of the data which embeds assumptions about the composition of the data. The TMSI requires that the links between physical processes, the measurements of them, and the interpretation of those data be well argued, and that there exist well argued links between the statistical detection (or non-detection) of events and alternative hypotheses about physical processes. This chapter addresses that need. Until now the detection of abrupt shifts in temperature data has generally been conducted without sufficient consideration of unconsidered processes as they affect error-statistics, decreasing the utility of probabilities, and weakening the conclusions.

Because it’s impossible to test for all unconsidered processes, Chapter 4 uses a battery of tests, each with differing assumptions, and which are individually framed as null hypothesis tests. That none of the validation tests share the same null or alternate hypotheses with the detection test is critical[6]. The result is sharper because the data segments are only tested further if they contain a change-point determined by a prior test.

---

[6] The nearest to sharing null hypotheses is between the MYBV and ANCOVA. But the MYBV actually has a null of continuity given an assumption of continuing zero trend and i.i.d. data, whilst ANCOVA has a

Another advance is the combination of stationarity tests to investigation of the effect of averaging of observed temperatures at varying scale. This in turn leads to evidence in later chapters that step-like regime changes occur at regional to sub-regional scale.

Finally the results of later work are improved (e.g. zonal analyses of climate model data) by signalling when data are unsuited to the detection method.

Chapter 4 attempts to address the gap between the error statistics of a detection method, and the assumptions about the data that govern the choice of detection method. It is telling that the MSBV and structural-change methods may return similar sets of change-points, despite one detecting only shifts and the other shifts and/or trend changes, especially in the RCP8.5 futures. When this happens many of the change points may not be statistically significant ($p<0.05$) tested within the data delimited by neighbouring change-points; either by the Chow test or the more general ANCOVA used herein. The fact that both methods give similar results probably indicates that (a) the MSBV detects existing shifts, and (b) the ANCOVA test (equivalent to a Chow test), since it treats shift and trend as interacting parameters is relatively conservative. Regardless it certainly means that care must be taken.

The use of Unit Root tests is not unprecedented, the use made of them is perhaps a bit different. They have been used for detection of change-points, especially where both endogenous changes (those resulting from the unforced structural properties of the system) and exogenous changes (resulting from restructuring of the system after an intervention) are possible. In this domain, a state change under forced warming is exogenous, a random walk which eventually returns is endogenous. The main use made here however is that three tests are used to examine segments within which a supposed exogenous shift has occurred. Each is based on different assumptions and null hypotheses. This form of post-hoc testing yields both extra data and greater nuance.

Climate

The Chapter 5 commences with three questions, and advances have been made in answer to all three.

1. Does the composition of area averaged data bias results? Many papers commence with an assumption that it does not.

---

null of continued linear progress given any trend and i.i.d. data. The tests applied as a conditional pair, given that MSBV has found a change-point, can this be an error due to continued trend? This is the reverse of a multiple testing fallacy.

2. Does the analysis at zonal scale support regime shifts, and if so do they favour regimes being global or sub-global in scale?
3. Is H2 – interaction between warming and natural variability – supported by global climate models?

To answer question 1, and partially address question 2. Area averaging of data does indeed have an effect – and it appears to increase the complexity of the data.

- Combining land and ocean temperatures increases measures of non-stationarity. Data segments containing single change-points from combined-ocean data are more likely to show non-stationarity than those from either land or ocean separately. And in any event land segments, global or zonal are more likely to be stationary than corresponding ocean ones.
- Combining data over the globe increases the same measures relative to zonal averages, whether land, ocean or combined land-ocean averages are involved. When data are averaged to finer scale as sectors of zones, the effect is more pronounced. Measures of stationarity increase and in fact stationarity rates in oceans converge on those for land. False positive rates for the MSBV observable in ocean data fall in going from zones to sectors, and within sectors no confirmed false positives due to continuing trend being identified as shifts occur.
- Lastly the signals become more step-like and less trend-like at finer scale, consistent with shift-like regimes occurring more locally, and not consistent with the absence of shifts locally (hence shifts at global scale do not result from the averaging of trend changes in finer scale data). Evidence that the zonal averages are themselves composites was also given, and the UR tests were used to demonstrate that a plausible scale for ocean regimes was probably sub-basin level.

To further address question 2. Step-like regime shifts show in zonally averaged data. They are also differentiated with some events in the Southern Hemisphere, some in the North. Hence regimes are sub-global phenomena, although some events appear to be more global.

- Post WW2 they tend to be more frequent everywhere.
- An event circa 1968 shows in the Southern mid-latitude oceans but not over land.
- The circa 1976 shows over all of the South and in the Northern tropics.
- The circa 1986 event shows in the Northern extra-tropics.
- The so-called hiatus, circa 1996 is apparent in all zones. It is possibly earlier in the polar North but differs slightly between data sets.

- In the global scale analyses, 1968 does not appear, and the 1986 event shows only in the oceans of the NCDC data set.
- Prior to WW2 there is more variation between data sets and between land and ocean splits. Although a very clear Southern Hemisphere event shows circa 1936.

Question 3, interaction between warming and variability, is addressed by addressing a series of questions, all based on an ensemble of global climate models run with hypothetical pre-industrial, observed, and hypothetical 21st Century atmospheric conditions. As a result the models clearly show warming interacts with the variability modes that give rise to regime shifts.

(a) Are patterns of zonal regimes from models consistent with observations?
(b) If so, do the patterns change with warming, especially either side of the observed time line?
(c) If so does warming affect the periodicity of quasi-oscillatory variability modes? Does it intensify such events? Does it affect the area involved?

(3a) The ensemble of GCMs appears to reproduce approximations of the major events during the 20th Century. There is a general tendency to find events in the early 20th Century which are weakly represented in observed data, but post WW2 there are peaks for events around 1968, 1976, 1986 and 1996, the latter by far the most extensive. (3b&c) The study of return periods within zones shows that without forced warming and under steady state conditions models predict very rare regime shifts except in the Southern mid-latitude oceans. When forced by observed conditions regime shifts converge on observed rates, with more frequent, and more extensive shifts. These do not necessarily become more intensive.

As stated in the introduction to this chapter, analysis of global climate averages alone is inadequate to the task of understanding decadal climate variability.

Evidence that the zonal averages are themselves composites was also given, and the stationarity tests were used to demonstrate that a plausible scale for ocean regimes was probably sub-basin level.

Spatial

The extension of change-point analysis to spatial scale is itself an advance. Although a superficially similar paper was published during this thesis (Yan et al., 2016), the authors simply assumed abrupt changes and attempted to quantify the duration of the transition between states.

- Hot spots of regime changes are apparent, especially along Western ocean boundaries.
- The role of the Southern higher mid-latitudes, and to a lesser extent the Northern sub-polar regions, seems to be earlier regime changes than elsewhere.
- Once step-like regime shifts are accounted for, land regions have a remaining and even accelerating trend, whereas ocean regions are close to having zero residual trend. This would imply that the ocean overall (polar regions excepted) releases heat in well defined events that mark regime changes and largely maintain (or attain) equilibrium in the stable periods between.
- Characteristic patterns associated with known decadal variability were catalogued. How these relate to climate models remains to be explored although replicates of one climate model's representation of the 20th century were examined and found to show strong intra-model variability. In the process they were also shown to contain some similar characteristic patterns to observations, indicating that the models may form similar regional structures to reality, but with variable sequencing with respect to each other and observations. The last might mean that the model runs each find different pathways to energetic equilibrium under the constraint of prescribed atmosphere; or it may simply be an artefact of the models. How this relates to the physical Earth is still an open question.
- Analysis of the ocean temperatures at 100m and 700m also showed that wide-spread events such as 1976 and 1996 are accompanied by changes in the vertical ocean heat structures. These happen extensively and rapidly. There are suggestions in the North Atlantic that these are spatially correlated with the Atlantic Meridional Overturning Circulation. Drijfhout (2018) discusses hiatus periods within GCMs as a function of ocean heat uptake and ocean heat redistribution. While his method applies a smoothing filter and applies different criteria, the results are pertinent to the thesis, JR2017, and JR2019, since I find a link between vertical ocean heat profiles and decadal scale variability. He also makes the point that observed behaviour would be likely to be more severe than suggested by models.
- Sgubin et al. (2017) identify prolonged cooling over the northern Atlantic in their discussion of the potential for an AMOC tipping point. Figure Ch6.35 in the left pane shows the same cooling as mentioned but the story is more nuanced. Much of the Northern Pacific and Northern Atlantic show cumulated negative trends, once the influence of steps is partitioned out (Figure Ch6.34). Figure Ch6.37 (cumulative sum of steps and trends) shows that three NML locations display net cooling (as modelled by a

step and trend model) the North Pacific Sub-polar gyre being another. Further analysis is out of scope of this thesis.

Modelling the hiatus

A model ensemble shows consensus corresponding to the so-called hiatus in the zonal analysis (see Figure Ch5.33), and yet spatially there is evidence that the internal states of replicates of one model differ and do not necessarily align to observations. JR2017 shows that an ensemble of RCP4.5 runs produces a pattern mirroring the observed shift dates (see our Figure 2), although no one ensemble of runs from individual models produced all the observed dates. JR2017 says, “The only event reproduced widely by the models was the 1996–98 step change, peaking in 1997, when 58 of the 107 MME (55 %) underwent a step change …” (ibid, page 190). As I stated in the section “Additional considerations regarding GCMs”, page 187, there is no reason to expect models to closely replicate internal structures and their evolution given prescribed atmospheres. But the consensus for the so-called hiatus is stronger than for any other event. (Guemas et al., 2013) (GU13) used EC-Earth to continually re-initialise model state variables from observations to show that the near-term prediction (1-3 years) improves. Their Figure 1 shows what may represent a consensus marked increase in temperatures circa year 2000, although this is not discussed. Kosaka and Xie (2013) (KC13) shows that with the prescribing of SSTs in the 8.2% of the SSTs represented by the tropical Pacific the predicted GMST closely represents that observed. This was taken further when Kosaka and Xie (2016) (KX16) proposed the tropical Pacific as a key pacemaker of global warming, a proposal not unrelated to that of JR2019. They delineate epochs: 1910 to mid-1940s,1940s-1970s (the “big hiatus”, echoing Trenberth (2015)), 1970s to -1990s, 1990s to now (at 2016). These three papers may cause one to believe that strong nudging (GU13), or pace-making (KX13, KX16) are required for a model consensus of regime changes to meaningfully align to time. However Meehl et al. (2014) also showed that when by chance, GCM simulations have IPOs that align to observations around the hiatus period they tend to also show reduced rates of warming (e.g. their Figure 1). The result illustrated in Figure Ch5.33 may well be explained as a result of an embedded signal in the atmospheric prescriptions, models evolving a selection of internal states which track this signal more or less successfully, and a degree of persistence in the models sufficient to give some decadal predictive capacity. Further work would be required.

Very recent findings

This section discusses some very recent results that I have not yet fully processed, and several recent publications of interest.

Much of the data in this thesis dates back several years. Over time climate data collections are updated in three important ways. (a) New data is added, (b) errors and biases in existing data are addressed, and (c) data processing algorithms change.

At the outset of this work in 2014 the PDO was already thought to be a component of decadal variability, and it was showing indications of a looming change of phase, and in fact Thomson and Emery (2014, p561), as an example of regime change detection, demonstrate a change in the PDO circa 2013 using STARS. The prohibition period in MSBV would mean that changes within the last seven years would not be detected. Hence, to maintain consistency, I elected not to continually update the data included for the body of the thesis.

In preparation for JR2019, Jones analysed more recent datasets, while I continued mainly with data collected earlier. It is now clear that two post WW2 major step-like regime changes have aligned with the phase changes of the PDO, and that detection of another temperature shift aligned with the PDO would serve as a degree of confirmation. Definitive measures of the PDO, suited to assignment of phase, include substantial data and mean that values are not available for some time. However as of now, according to Jet Propulsion Laboratory, the PDO last changed phase in 2014. (see https://sealevel.jpl.nasa.gov/science/elninopdo/pdo/)

Use of the MSBV to investigate recent changes requires that the prohibition period be shortened such that a final change-point observe a three year prohibition from the end of data. Code changes in the MSBV to support this are not fully tested.

Two datasets of interest were examined, GISTEMPv3-2019, a 2°x2° grid and NOAAGlobalTemp-2019, a 5°x5° grid (see Appendix 1.1).

Examination of the global mean surface temperature shows that temperatures spiked (by at least 0.2°C and as much as 0.6°C) in an identifiable step-like change after 2013, and even five years later remain above the previous record temperatures. Over Europe, for example, the internal shifts may exceed 0.6°C after 2013. There is too little data to assess the statistical effect of trend or trend change, nor for the diagnostics to be fully meaningful. None the less it aligns, or slightly precedes the model consensus of a regime change shown in Chapter 5.

The dates obtained may be affected by the central bias issue (change at one end of the data in the presence of continuing trend) that I investigated in Chapter 3. This would give earlier dates for an analysis conducted now than if repeated in another four years.

Further directions

Studies of the recent 21st century regime change

The apparent regime shift of 2013/4 just mentioned would be a critical event in climatology. A simple model of regimes linked to the PDO is that warming rates decline in a cool phase PDO and rise in a warm phase PDO. For example in a study of GCMs, Meehl et al. (2011) suggest increased heat uptake below 300m associated with slower surface warming. Trenberth (2015) links warming rates with the PDO. The PDO switched from cool phase to warm phase in 2014, as it did in 1976/8. Further to that JR217 and JR2019 suggest that heat release events are intimately associated with the regime changes, with the Western Pacific Warm Pool storing and then releasing heat.

The extent of the 2014 event, it's biological and weather related consequences, and underlying mechanisms all remain open questions.

The date itself is increasingly being recognised (Yin et al., 2018). In Chapter 6, spatially distributed patterns in persistent, regime-like changes, signalled as vertical redistributions of 100m and 700m ocean temperatures, were associated with the 1976/8 event as were patterns of surface temperature shifts. Therefore a similar study of current ocean temperature structure is warranted.

For many reasons lead indicators of such a change are needed. These could be physical or ecological. The relationship between physical regime shifts and biological regime changes has been studied (Mantua, 2004, Hare and Mantua, 2000, Reid et al., 2015, Reid and Beaugrand, 2012). Different audiences have different needs, and lead indicators should be generally applicable and easily understood. Work has been done with covariability and coupling between climate indices (Tsonis and Swanson, 2011), and ecologist have looked at changes in variability as lead indicators.

The detail of the recent event is very much an open question. For one studying climate regime changes, a linear study is needed to track the ongoing climate state. The explosion of literature concerning the so-called hiatus is something of an object lesson. The debate has been dogged by imprecise definitions, lack of inductive frameworks, and misused methods.

Statistical induction in climate science

When searching for important statistical markers of variations in climate, it is imperative to have a clear understanding of the link between the physical systems and the appropriate statistical test, and yet it is quite common for papers to be based on ungrounded assumptions.

The particular assumption (usually implicit) that has led to much contention is that mean ocean surface temperatures are suitable proxies for ocean heat content, so that only trend is important, and thus (what JR2017 refers to as H1), variability in surface temperatures given steady ocean heat accumulation must be independent of forced warming) (Risbey et al., 2018, Rahmstorf et al., 2017, Foster and Abraham, 2015, Beaulieu and Killick, 2018). This leads to authors attempting to conflate the statistics of detection and probative testing.

There appears to be a need for a review of statistical induction in climate science that includes such things as the TMSI framework.

Detection of change-points

The optimal detection of steps in the presence of a possibly confounding trend change appears to remain an open problem. The research domain of this thesis dictates that one use a method weighted for the detection of steps and resistant to trend, rather than adapting an *F*-ratio method such as the Chow test. This is because it is postulated that step-like changes are associated with climatic regime changes (Yin et al., 2018, Jones and Ricketts, 2017b, Bartsev et al., 2017, Yan et al., 2016, Reid et al., 2015, Varotsos et al., 2014, Belolipetsky et al., 2015, Belolipetsky, 2014).

However the detection of an abrupt change also has parallels to many other problems from diverse sciences, such as edge detection. When my methods are applied spatially one finds coherent spatial local shifts, and I take the spatial coherence to be meaningful, so perhaps image processing methods also may be adapted. Similarly, the MYBV test itself appears to have much in common with digital filter theory; the formulation of the Ti function would suggest that.

The MYBV is not suited to data in which substantial sustained trend is present. The effects are shown mostly in the climate model analyses, especially of RCP8.5, where high levels of apparent trend are seen in the zonal averages. As a result the numbers of detected change-points which are rejected by ANCOVA rules against further analysis of ensembles after the first quarter of the 21st century. Further work is needed to make the MYBV resistant to trend without losing precision.

The PELT algorithm of Killick et al. (2012) is an exact search method which can work with any change-point detection test and which uses a choice of halting criteria. The difference between the MSBV's approach to the search part of the algorithm (recursion combined with step-wise refinement) and PELT (a tree pruning method) may be of interest. Any test can be combined with PELT.

The role of noise specifically in level change detection has not to my knowledge been fully explored, and should be. The ideas of injecting it into an auxiliary variable rather than the primary independent variable, as per VJ2005, is novel and immensely useful.

Autocorrelation

An issue raised quite recently is that of auto-correlation (Beaulieu and Killick, 2018) (see last part of Chapter 4). Whilst it is unlikely that step-like changes are misidentified as a result of auto-correlation, auto-correlation is mis-quantified, just like any other parameter is, when unconsidered deterministic features such as step-like change are present. The mis-quantification can extend to the identification of a sequence being dominated by unit-root progression. This is important to my selection of unit-root based tests, and the interpretations I give them.

There remains an open question of how to best isolate the auto-correlated component of a signal when it may contain multiple deterministic components, trends, shifts, oscillations etc. as well as random components. This affects both use of autocorrelation as a parameter of interest, and as a parameter to be compensated for. Mizon (1995) outlines the complexities of autocorrelation estimation and warns against OLS approaches and the use of them to correct data for it. It is likely for example that this explains the failure of Belolipetsky et al. (2015) to find a change-point for circa 1976. Rodionov (2006b) acknowledges the issue but finds that, even using more complex estimates of auto-correlation, regime detection sensitivity for the STARS test reduces.

Post-detection testing

I would also strongly recommend that the issues flowing from the composition of the signals and the ability of simple but under analysed averaging to add features or hide features, and at worst, to deceive the detection methods, be considered. It is becoming clear that renewed interest in autocorrelation dictates that care be taken in isolating it.

Despite the various stationarity tests themselves being automatable, the decision rules for their interpretation remain under-determined, however in Chapter 4 the selected suite of tests provides sufficient nuance to inform a strong conclusion about the composition of mean temperature series; specifically that it is the inappropriate composition of temperature data from regional data affected by different processes and timings that leads to the appearance of a degree of non-determinism which in turn leads analysts to attribute deterministic variation to random fluctuations.

Zonal and spatial analyses

I suggest that the spatial analysis be concentrated upon and extended. There are more challenges, but I believe much more information to be gained. There are clearly multiple events within zones, and these events also extend outside zones. As seen with the analysis of ocean temperature at depth, and fact by the presence of shifts in the troposphere (Jones and Ricketts, 2017b Figure 4) regimes involve three dimensional structures.

The atmosphere and the ocean are in constant motion, and hence, although having spatial analyses referenced against Cartesian coordinates is appealing, the motion can confound our understanding. This issue of appropriate coordinates is too, an open question.

Models and internal states

The results of GU13, KX13 and KX16 mentioned above may be taken to argue against my conclusion that the model ensemble consensus location of a step-change corresponding to the so-called hiatus gives credence to a prediction of a corresponding step-like change in the second half of this decade. This creates an open question as to the necessity of close control of internal model states for near term prediction, and the desirability of lead indicators suggests this open question be addressed and the results utilised.

Exploration of other data.

Since the PI-Controls in models are prescribed, paleo-reconstructions could be added to PI-Controls as a source of information on the unperturbed climate, since the imprint of actual climate processes is present. Proxy studies were performed by Jones (1995) and in subsequent work. The main issue is that detection of abrupt shifts can be desensitised by smoothing. Individual tree ring data may well be fine but compositing methods would need to be considered carefully. Other data sources need to have temporal resolution compatible with observations, without intrinsic smoothing and with serial independence.

The attribution studies outlined by Jones (2012) and Jones and Ricketts (2017b) can be performed both in models and in gridded data. It is of interest to find out how general and informative these can be.

Publications

Through the course of this thesis I have published three conference papers, contributed to and co-authored a number of working papers (Jones and Ricketts, 2016a, Jones and Ricketts, 2016b, Jones and Ricketts, 2017a) and co-authored two papers (JR2017 and JR2019).

- The first conference paper built on previous work in machine learning called stepwise symbolic regression (SSD) (Ricketts, 2013) and applied the method to locating abrupt shifts in temperature time-series, comparing it to the MSBV (Ricketts, 2015b).
- The second introduced the MSBV under the name “probabilistic bivariate test” (Ricketts, 2015a).
- The third proposed the tests to be used for validation of change-points (Ricketts and Jones, 2017).
- JR2017 contains both the results obtained using the MSBV applied to zonal and global observations, and global climate models, plus work on regional attribution. In addition it contains the foundational work for the TMSI (Jones and Ricketts, 2017b).
- JR2019 proposes a conceptual model that ties decadal variability to a global heat pump (Jones and Ricketts, 2019).

Conclusion

The nature of abrupt decadal shifts in a changing climate is that they constitute an enhancement of decadal scale natural variability driven by energy imbalances, and are regional, not global in origin. They involve regime state changes in three dimensional structures in the atmosphere and in the oceans that underlie natural decadal variability. They are deterministic in character, rather than random events.

The abruptness detected in climate time-series is a primary signal, certainly more than an artefact of detection methods or data analysis. By and large, trends may be properties of the new regimes, but shifts signal the timing.

Therefore it is important to understand fully the impact of all stages of data preparation. The approach used here is to avoid pre-processing, and to examine the data after change-points have been detected. Overwhelmingly when finer scale data are used non-stationarity is mostly revealed to not factor in the detection.

Appendices

Appendix 1.1: Data sources and preparation

NCDC zonal data version v3.5.4.201504.

Annual and monthly files in ASCII format covering anomalies of land, ocean, and combined land and ocean were downloaded from The National Climate Data Center (NCDC, now known as National Center for Environmental Information) on 29 May 2015 via ftp://ftp.ncdc.noaa.gov/pub/data/mlost/operational/products/ using wget in recursive mode. Each file contains data for one zonal average and for one of land, ocean and combined land and ocean. The zonal averages were over: 90°S– 0°N (Global), 90°– S0°S (Southern hemisphere), 0°N–90°N (Northern hemisphere), 90°S–20°S, 60°S–30°S, 60°S– 60°N, 30°S– 0°N, 0°N – 30°N, 20°S– 20°N, 20°N– 90°N, and 60°N– 90°N. Data in the files labelled as 90°S–60°S for all three subsets was clearly corrupted on receipt and was not used. Annual averages are as provided, rather than simple averages of monthly values. Anomalies are based on a 1971-2000 average. The data format is documented online in the file ftp://ftp.ncdc.noaa.gov/pub/data/mlost/operational/products/readme.timeseries.

GISTEMP3 zonal data

Also reported in Chapter 3 is an analysis of GISTEMP3 zonally averaged data of combined land/ocean temperatures LOTI anomalies (1950-81) downloaded on 15/10/2014 from http://data.giss.nasa.gov/gistemp/tabledata_v3/ZonAnn.Ts+dSST.txt. This data is differently zoned than the above mentioned NCDC zonal data version v3.5.4.201504.

GISTEMP3. Combined Land Ocean Gridded Data

GISTEMP3 monthly gridded combined air surface temperature with a final record of January 2017, was downloaded from Koninklijk Nederlands Meterorologisch Instiuut (KNMI) using their climate data explorer on 6 March 2017 https://climexp.knmi.nl/selectfield_obs2.cgi?id=someone@somewhere =giss_temp_1200

The data are combined Land-Surface Air and Sea-Surface Water Temperature Anomalies (Land-Ocean Temperature Index LOTI) as anomalies from the 1951-1980 mean, smoothed at 1200Km commencing in 1880. (See also documentation at https://data.giss.nasa.gov/gistemp/). Land air temperatures at 2m are combined with sea surface temperatures from the Extended Reconstructed Sea Surface Temperature (ERSSTv4) on a 2^o x 2^o grid, provided by NCDC (see https://www.ncdc.noaa.gov/dataaccess/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v4). This dataset is designated as GISTEMP3 throughout. (GISSTemp Team, 2015, Liu et al., 2015, Huang et al., 2015)

Ocean temperatures

Estimated ocean temperatures anomalies (from mean 1955-2012) at 100m and 700m depths were also downloaded from KNMI at the same time. These are monthly estimates on a 1°x1° grid, between January 1955 and December 2016, and derived from the World Ocean Atlas (WOA12v2) https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/avt_global.html

Ocean basin averaged monthly temperature data for 100m, 700m and 2000m depths, originating the National Oceanographic Data Centre (NODC) from were also downloaded from KNMI on 13 Nov 2015, as NetCDF files.
https://climexp.knmi.nl/select.cgi?id=someone@somewhere&field=nodc_temp100

Monthly data for the standardised depth anomalies (1982-present) of the 20°C isotherm in the tropical West Pacific and the tropical East-Central Pacific were downloaded on 29 January 2018, using the following URLs. Data collection is February 1982 to present (Levitus et al., 2010).

https://iridl.ldeo.columbia.edu/maproom/ENSO/Time_Series/Heat_Storage_ECent_Pac.html

https://iridl.ldeo.columbia.edu/maproom/ENSO/Time_Series/Heat_Storage_West_Pac.html

Global Climate Models.

Modelled surface temperature monthly data used here was originally released to the Program for Climate Model Data & Intercomparison (PCMDI) as part of the Intergovernmental Panel on Climate Change (IPCC) Climate Model Intercomparison Project Phase 5 (CMIP5). A mirror of selected data held on the National Computing Infrastructure (NCI) facility (raijin.nci.org.au), compiled by the Bureau of Meteorology, was used with permission. The data had all been gridded to a one degree common grid using the Climate Model Data Output Rewriter (CMOR). Pre-industrial control runs, Historical and future climates evolved under scenarios defined for four future representative climate pathways (RCPs) (Meinshausen et al., 2011) were obtained. Pre-industrial, historical and RCP runs are derived from a specified initialisations designated by a “RIP” code where all files sharing a RIP code descend from the same original initialisation. Pre-industrial runs are evolve under a stationary atmospheric compositions, historical runs continue on but with observed atmospheres, and RCP runs continue after 2005 but with prescribed future atmospheres. RIP codes look like “r1i1p1” where “r” is followed by a realization number where all realizations are different but start from equally likely initial conditions, “i” by an initialisation number, denoting either a different time of divergence of a run from spin-up, and “p” by a “physics” number denoting alternative implementations of physics or parameterisations.

I acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and thank the climate modelling groups (listed in Table A1.1.25*) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.*

Land Ocean Masks

A 1°x1° gridded ocean basin map was downloaded from NOAA NODC WOA09: World Ocean Atlas 2009 (Levitus et al., 2013) on the 7th March 2017 from http://iridl.ldeo.columbia.edu/SOURCES/NOAA/NODC/WOA09/oceanbasindata_1x1.nc. This was used to delineate ocean basins and extra code was used to delineate continental land masses, with Europe and Asia treated as a single Eurasia. The dataset is referred to as WOA09.

Each climate model is also supplied with its own land/ocean mask.

Climate Modelling Centres

Table A1.1.25: Global Climate Models and institutions. Edited from the table published at https://pcmdi.llnl.gov/mips/cmip5/availability.html

MODELLING CENTER	MODEL	INSTITUTION
BCC	BCC-CSM1.1 BCC-CSM1.1(m)	Beijing Climate Center, China Meteorological Administration
CCCMA	CanESM2	Canadian Centre for Climate Modelling and Analysis
CNRM-CERFACS	CNRM-CM5	Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique
CSIRO-BOM	ACCESS1.0 ACCESS1.3	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)
CSIRO-QCCCE	CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence
EC-EARTH	EC-EARTH	EC-EARTH consortium
GCESS	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University
IPSL	IPSL-CM5A-LR IPSL-CM5A-MR IPSL-CM5B-LR	Institut Pierre-Simon Laplace
LASG-CESS	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University
LASG-IAP	FGOALS-s2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences
MIROC	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
MIROC	MIROC-ESM MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies
MOHC WITH INPE	HadGEM2-ES	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)
MPI-M	MPI-ESM-LR MPI-ESM-MR	Max Planck Institute for Meteorology (MPI-M)
MRI	MRI-CGCM3	Meteorological Research Institute
NASA GISS	GISS-E2-H GISS-E2-R	NASA Goddard Institute for Space Studies

NCAR	CCSM4	National Center for Atmospheric Research
NCC	NorESM1-M NorESM1-ME	Norwegian Climate Centre
NOAA GFDL	GFDL-CM3 GFDL-ESM2G GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory
NSF-DOE-NCAR	CESM1(CAM5)	National Science Foundation, Department of Energy, National Center for Atmospheric Research

Regridding of data

Unless otherwise noted, all gridded data were regridded one of two ways.

Where data was required on a common grid, but necessarily not supplied on the required grid, they were post-processed on the NCI facility using Climate Data Operators (CDO) using flux conservative options. E.g. to produce a common regular 5°x5° grid using flux conservative options the shell command, “cdo remapcon,r72X36 input.nc output.nc” was issued; to produce a 2°x2° grid, “cdo remapcon,r180X90 input.nc output.nc”.

In some cases, where the provided data was on a finer grid that could be overlaid on the required grid without overlap, the required grid was produced by area weighted averaging from the supplied data. Zonally averages were likewise produced when only gridded data was supplied, by area weighted averaging.

Appendix 3.1: Derivation of governing Maronna-Yohai test equations.

Maronna-Yohai's bivariate.

The bivariate test was introduced as an homogeneity test by Maronna and Yohai (1978) who included sample tables of critical levels of their T_0 statistic for various sample sizes. Potter (1981) illustrated the use of the test in the detection of a mean shift in a precipitation series, corrected some typographic errors, and also extended the table of critical values to cover sample size 100.

Bücher and Dessens (1991) derived a formulation of the test that included a normalisation step and applied this to a temperature series from the high Pyrenees. Vives and Jones (2005) used the later formulation and extended the method further using flat random reference series. Jones (2012) uses this work and supplements it with the STARS test (Rodionov, 2004) as an extra validation.

The derivation of Bücher and Dessens (1991) is straightforward enough but I have not seen an assessment of the full effect of the use of a random reference series (Vives and Jones, 2005), although it seems well validated experimentally.

Maronna & Yohai as presented by Potter.

The formulation given by Maronna and Yohai (1978), was published with minor typographic differences in a paper exploring homogenisation of precipitation time series (Potter, 1981).

Assuming that (x_i, y_i) are i.i.d. random vectors of length n,

Let x_i be a stationary reference time series and y_i be a test time-series which is assumed to correlate to x_i except for a single shift at some time i_0.

Step 1 Define running averages, where n is the length of the time series and where "time" really means sample number and no requirement for equi-spacing is implied.

$$X_{i={}^1/_i \sum_{j=1}^{i} x_j},\ Y_{i={}^1/_i \sum_{j=1}^{i} y_j} \qquad \text{(A1)}$$

Step 2, define grand means.

$$\bar{X} = X_n,\ \bar{Y} = Y_n, \qquad \text{(A2)}$$

Step 3, define deviation sums of squares, there are three of these, and are not time varying.

$$S_x = \sum_{j=1}^{n}\left(x_j - \bar{X}\right)^2, S_y = \sum_{j=1}^{n}\left(y_j - \bar{Y}\right)^2, \; S_{xy} = \sum_{j=1}^{n}\left(x_j - \bar{X}\right)\left(y_j - \bar{Y}\right). \tag{A3}$$

Now define two convenience functions F_i and D_i, both of which vary by time

$$F_i = S_x - \frac{(X_i - \bar{X})^2 ni}{(n-i)}, i < n \tag{A4}$$

$$D_i = \frac{\left[S_x(\bar{Y} - Y_i) - S_{xy}(\bar{X} - X_i)\right]n}{(n-i)F_i}, i < n \tag{A5}$$

Then define the statistic T_i

$$T_i = \frac{\left[i(n-i)D_i^2 F_i\right]}{\left(S_x S_y - S_{xy}^2\right)}, i < n \tag{A6}$$

And finally the test itself,

$T_{\mathrm{i0}} = \max\limits_{i<n}\{T_i\}$, and where i_0 is the i value corresponding to T_0 . (A7)

Bücher et al.'s derivation

The formulation published in (Vives and Jones, 2005, Kirono and Jones, 2007, Jones, 2012) is simpler and derived from Appendix B of Bücher and Dessens (1991) in order to apply the published critical values. Here follows a derivation of the latter form.

Below, I have preserved their notation with one exception. In their step 1 (equation A8), because they are about to redefine $\{x_i\}$ and $\{y_i\}$ as normalised versions, they indicate the un-normalised series as $\{\acute{x}_i\}$ and $\{\acute{y}_i\}$. I have extended this to the derived values throughout. This is especially important since S_x and S_y change their meaning from Potter equation (3), although they do not appear in the final equations in either form.

Step 1 : Means and standard deviations.

$$\overline{X'} = \frac{\sum_{j=1}^{n} x\acute{}_J}{n}, \overline{Y'} = \frac{\sum_{j=1}^{n} y\acute{}_J}{n}, \acute{S}_x = \left(\frac{\sum_{j=1}^{n}\left(x\prime_j - \overline{X\prime}\right)^2}{n}\right)^{1/2}, \acute{S}_y = \left(\frac{\sum_{j=1}^{n}\left(y\prime_j - \overline{Y\prime}\right)^2}{n}\right)^{1/2} \tag{A8}$$

Step 2: Standardise.

$$x_j = \frac{(x' - \overline{X\prime})}{\acute{S_x}} \text{ and } y_j = \frac{(y' - \overline{Y\prime})}{\acute{S_y}} \text{ for all } j. \tag{A9}$$

Now Potter's equations apply throughout, but because of the standardisation step (9) we also know that

$$\bar{X} = 0 \text{ and } \bar{Y} = 0 \tag{A10}$$

S_x, S_y and S_{xy} reduce to,

$$S_x = \sum_{j=1}^{n} {x_j}^2, S_y = \sum_{j=1}^{n} {y_j}^2,\ S_{xy} = \sum_{j=1}^{n} x_j y_j. \tag{A11}$$

S_x, S_y can be reduced further, as well. From equations (A9 and A11),

$$S_x = \sum_{j=1}^{n}\left(\frac{(x' - \overline{X'})}{\acute{S_x}}\right)^2 = \sum_{j=1}^{n}\frac{(x' - \overline{X'})^2}{\left(\acute{S_x}\right)^2} = \sum_{j=1}^{n}\frac{(x' - \overline{X'})^2}{\left[\left(\frac{\sum_{j-i}^{n}\left(x'_j - \overline{X'}\right)^2}{n}\right)\right]}$$

Hence,

$$S_x = \left[\frac{n}{\sum_{j=i}^{n}\left(x'_j - \bar{X}'\right)^2}\right]\sum_{j=i}^{n}\left(x'_j - \bar{X}'\right)^2 = n \tag{A12}$$

Similarly $S_y = n$.

So immediately F_i and D_i reduce to the later forms,

$$F_i = n - \frac{{X_i}^2 ni}{(n-i)}, i < n \tag{A13}$$

$$D_i = \frac{[S_{xy}X_i - nY_i]n}{(n-i)F_i}, i < n \tag{A14}$$

And T_i cannot reduce further than,

$$T_i = \frac{[i(n-i)D_i^2 F_i]}{(n^2 - S_{xy}^2)}, i < n \tag{A15}$$

Let $T_{i0} = \max(T_i)$ and let i_0^* be the value of i for which $T_i = T_{i0}$, the time after which a change occurred. Its successor is the first time of the new regime.

D_i^* is defined as the maximum likelihood estimator of a shift at i_0^*.

T_{i0}^* is the test statistic that tested against some constant, discriminates with a specified probability, a null hypothesis H_0 of no shift against H_1 that a shift exists (Maronna and Yohai, 1978).

A mean shift can be computed as $\Delta\bar{y} = D_i^* . \acute{S}_y$.

For the null trend case, analyzed in Maronna and Yohai (1978), critical values of T_i are given for probabilities of (0.25, 0.1, 0.05, and 0.01) for the null hypothesis of no change, given time series lengths *n* of 10, 15, 20, 30, 40 and 70, and a range of correlations (ρ) between *X* and *Y* . Potter (1981) provides these for *n*=100 and ρ=1.

Maronna-Yohai test with random controls

The imposition of a random control essentially simplifies the test further. At all times the variates are normalised such that, absent trend and shift, the cumulative sums converge on zero.

For a flat random control, after normalisation, some limits apply; $\lim_{n\to\infty} S_{xy} \to 0, {X_n}^2 = 0$ and $\lim_{i<n\to\infty} {X_i}^2 \to (n-i)/ni$ due to Brownian progression. Note this latter is maximal at $i = n/2$.

Consider *Fi*

From (A13), given $F_i = n - \frac{X_i{}^2 ni}{(n-i)}$ for all $i < n$, after rearrangement and substitution we have

$F_i = \frac{n(n-i)-\left(\frac{n-i}{ni}\right)ni}{(n-i)}$ for all $i < n$, and can simplify so we can see that for a flat normalised control,

$$\lim_{X_i{}^2 \to 0} F_i = n - 1. \tag{A16}$$

Consider *Di*

Thus given $D_i = \frac{(S_{xy}X_i - nY_i)n}{(n-i)F_i}$ for all $i < n$, without trend, $D_i = \frac{(S_{xy}X_i - nY_i)n}{(n-i)(n-1)}$ for all $i < n$

$$\lim_{i<n, n\to\infty} D_i = -\frac{Y_i n^2}{(n-i)(n-1)} = -\frac{Y_i n}{(n-i)} \tag{A17}$$

Consider *Ti*

We can obtain the exact estimator of this given the limits above.

$$\lim_{i<n, n\to\infty} T_i = \frac{i(n-1)Y_i^2}{(n-i)}, \text{for all } i < n \tag{A18}$$

A18 is the central estimate of the T_i function for a random control, where there is no cross correlation between variates and S_{xy} converges on zero by necessity. The normalisation is crucial.

This version does not appear to be perturbed by red-like drift in the control variate which empirically seems to be an issue with segments of length less than 30 or so.

The relationship between the two main components of T_i is shown in Figure Ch3.6, which also illustrates its "spire" like construct.

Appendix 3.2: Comparisons of MSBV with other methods – table of results

Table A3.2.26: Break dates in artificial data by each of three methods. Bivariate is MSBV with a flat random control. CP is the change point method, SC is the structural change method. Red denotes the target years. In set A 2096 cannot be detected by MSBV due to the minimum seven year segment rule. In B there are two consecutive steps in 1973 and 1974, these cannot be separated by any method. Bolding denotes agreement within one year of the standard. Underscores denote shifts < 1 StdDev.

Set No	Contains	Analysis	Shift Years Found
A1	Random	MSBV	[]
A1	Random	CP	[]
A1	Random	SC	[1958, 1969]
A2	Random + Trend	MSBV	[1979, 2037, 2058]
A2	Random + Trend	CP	[1979, 2037]
A2	Random + Trend	SC	[1979, 2037, 2058]
A3	Random + Shifts	MSBV	[1973, 1998, 2035, 2058]
A3	Random + Shifts	CP	[1973, 1998, 2035, 2058]
A3	Random + Shifts	SC	[1973, 1998, 2035, 2054, 2070]
A4	Random + Shift + Trend	MSBV	[1937, 1973, **1998**, **2035**, **2054**, **2070**]
A4	Random + Shift + Trend	CP	[1969, **1998**, **2035**, 2058, **2070**]
A4	Random + Shift + Trend	SC	[*1937*, *1973*, **1998**, **2028**, **2035**, **2054**, **2070**, 2093]
A	Defined Breaks	Change Years	[1954, <u>1982</u>, 1998, 2029, 2035, 2054, 2070, <u>*2096*</u>]
B1	Random	MSBV	[]
B1	Random	CP	[]
B1	Random	SC	[2039]
B2	Random + Trend	MSBV	[1941, 1971, 2019, 2039, 2057, 2090]
B2	Random + Trend	CP	[1971, 2033, 2057, 2090]
B2	Random + Trend	SC	[1950, 1991, 2024, 2039, 2057, 2090]
B3	Random + Shifts	MSBV	[1973, 2024, 2055, 2074]
B3	Random + Shifts	CP	[1973, 2026, 2055, 2074]
B3	Random + Shifts	SC	[1973, 2024, 2033, 2055, 2074, 2086]
B4	Random + Shift + Trend	MSBV	[1941, **1973**, 2000, 2024, **2033**, **2055**, **2074**, **2086**]
B4	Random + Shift + Trend	CP	[**1973**, 2014, **2030**, **2055**, **2074**, **2086**]
B4	Random + Shift + Trend	SC	[1941, **1973**, 2000, 2024, **2033**, **2055**, **2074**, **2086**]
B	Defined Breaks	Change Years	[*1973*, *1974*, <u>2009</u>, 2026, <u>2031</u>, 2054, 2073, <u>2084</u>]
C1	Random	MSBV	[]
C1	Random	CP	[]
C1	Random	SC	[]
C2	Random + Trend	MSBV	[1945, 1966, 2035, 2065]
C2	Random + Trend	CP	[1966, 2035, 2065]
C2	Random + Trend	SC	[1945, 1969, 2035, 2065]
C3	Random + Shifts	MSBV	[1969, 1994, 2035, 2048, 2070]
C3	Random + Shifts	CP	[1969, 1994, 2035, 2048]

C3	Random + Shifts	SC	[1969, 1994, 2035, 2048, 2072]
C4	Random + Shift + Trend	MSBV	[1945, **1969**, **1994**, **2028**, **2035**, **2048**, 2070]
C4	Random + Shift + Trend	CP	[**1969**, **1994**, **2028**, **2048**, 2072]
C4	Random + Shift + Trend	SC	[1945, **1969**, **1994**, **2028**, **2035**, **2048**, 2070, 2090]
C	Defined Breaks	Change Years	[1969, 1986, 1995, 2027, 2035, 2049, 2075, 2094]
D1	Random	MSBV	[]
D1	Random	CP	[]
D1	Random	SC	[1951, 2062]
D2	Random + Trend	MSBV	[1994, 2062]
D2	Random + Trend	CP	[1994, 2062]
D2	Random + Trend	SC	[1987, 2029, 2062]
D3	Random + Shifts	MSBV	[1979, 2010, 2017, 2060, 2067]
D3	Random + Shifts	CP	[1979, 2012, 2060, 2067]
D3	Random + Shifts	SC	[1979, 2007, 2017, 2060, 2067]
D4	Random + Shift + Trend	MSBV	[**1979**, 2007, **2017**, **2046**, 2060, **2067**, **2084**]
D4	Random + Shift + Trend	CP	[**1979**, 2007, **2017**, **2046**, 2060, **2067**]
D4	Random + Shift + Trend	SC	[1972, **1979**, **2007**, **2017**, **2046**, 2060, **2067**, 2087]
D	Defined Breaks	Change Years	[1950, 1979, 2009, 2017, 2046, 2057, 2067, 2084]
E1	Random	MSBV	[]
E1	Random	CP	[]
E1	Random	SC	[]
E2	Random + Trend	MSBV	[1983, 2022, 2052, 2087]
E2	Random + Trend	CP	[1983, 2022, 2052]
E2	Random + Trend	SC	[1953, 1983, 2022, 2052, 2087]
E3	Random + Shifts	MSBV	[1953, 1983, 1999, 2022, 2052, 2064]
E3	Random + Shifts	CP	[1983, 2022, 2056]
E3	Random + Shifts	SC	[1953, 1983, 1999, 2022, 2052, 2067]
E4	Random + Shift + Trend	MSBV	[**1953**, 1983, **1999**, 2022, **2039**, 2052, 2064, 2087]
E4	Random + Shift + Trend	CP	[**1953**, 1983, **1999**, 2022, 2052, 2067]
E4	Random + Shift + Trend	SC	[**1953**, 1983, **1999**, 2022, **2039**, 2052, 2067, 2087]
E	Defined Breaks	Change Years	[1954, 1980, 2000, 2028, 2038, 2056, 2070, 2079]

Appendix 4.1: Tables of results

Table A4.1.27: Analysis of DS2 using the validation suite. Shift years returned by the bivariate test within two of the prescribed year are bolded.

		Bivariate		Unit Root and Stationarity Tests										ANOVA/ANCOVA			Segment Classification		CP-Index
				KPSS-Level		KPSS-Trend		ADF		Zivot Andrews				ANOVA		ANCOVA	(Stationarity testing)		(segment trends and ANCOVA)
Set	**Defined**	**Break**	**Shift**	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**ZA-year**	**ZA-Diff**	**pShift**	**pTrend**	**pRegime**	**Feature Code**	**Segment Class**	**Index**
A3	1955	1952/3	0.29	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1952	0	0.01	0.01	<0.01	25,26	Single, Stat	6
A	1983	step<1SD																	
A3		1992/3	0.58	0.01	0.10	0.02	0.10	0.01	0.01	0.01	0.01	1998	6	<0.01	0.90	<0.01	25,26	Single, Stat	7
A	1999	In A4 (same dates) 1998/9 is found. 1992/3 may be intermediate																	
A3	**2030**	**2029/30**	1.22	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2029	0	<0.01	<0.01	<0.01	22,26	Single, Stat	7
A	2036	Within prohibition period																	
A	2055	Missed																	
A3	**2071**	**2070/1**	0.65	0.01	0.10	0.05	0.10	0.05	0.01	0.01	0.01	2070	0	<0.01	<0.01	<0.01	26,26	Single, Stat	5
A	2097	Within end prohibition period																	
A4	**1955**	**1954/5**	0.73	0.01	0.10	0.01	0.10	0.1	0.01	0.01	0.01	1954	0	<0.01	0.01	<0.01	22,26	Single, Stat	7
A	1983	step<1SD																	
A4	**1999**	**1998/9**	0.65	0.01	0.10	0.03	0.10	0.01	0.01	0.01	0.01	1998	0	<0.01	0.13	<0.01	25,26	Single, Stat	7
A	2030	Within prohibition period																	
A4	**2036**	**2034/5**	1.21	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2034	0	<0.01	<0.01	<0.01	22,26	Single, Stat	7
A	2055	Missed																	
A4	**2071**	**2070/1**	0.72	0.01	0.10	0.03	0.10	>0.10	0.01	0.01	0.01	2070	0	<0.01	<0.01	<0.01	22,26	Single, Stat	7
A	2097	Within end prohibition period																	
B3	**1974/5**	**1973/4**	1.74	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1973	0	<0.01	<0.01	<0.01	22,26	Single, Stat	6
B	**2010**	Step 1 SD (not guaranteed)																	

B3	**2027**	**2026/7**	0.93	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2026	0	<0.01	0.95	<0.01	22,26	Single, Stat	5
B	2032	Within prohibition period																	
B3	**2055**	**2054/5**	1.23	0.01	0.10	0.06	0.10	>0.10	0.01	0.01	0.01	2054	0	<0.01	0.01	<0.01	20,20	Single, Stat	7
B3	**2074**	**2073/4**	1.06	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2073	0	<0.01	0.88	<0.01	0,2	Single, N/A	5
B3		2081/2	0.38	0.01	0.10	0.10	0.10	0.05	>0.10	0.01	0.01	2088	7	0.05	0.44	0.14	2,2	Single, N/A	4
B	2085	step<1SD, but 2081 may be intermediate																	
B4		1944/4	0.06	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1938	-6	0.63	0.15	0.21	26,26	Single, Stat	3
B4	**1974/5**	**1974/5**	1.72	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1974	0	0.00	0.00	<0.01	22,25	Single, Stat	7
B	**2010**	Step 1 SD (not guaranteed)																	
B4	**2027**	**2026/7**	1.01	0.01	0.10	0.01	0.10	>0.10	>0.10	0.01	0.01	2026	0	<0.01	0.13	<0.01	21,23	Single, Stat	7
B	2032	Within prohibition period																	
B4	**2055**	**2054/5**	0.62	0.01	0.10	0.02	0.10	>0.10	0.01	0.01	0.01	2054	0	<0.01	0.94	<0.01	18,20	Single, Stat	7
B4	**2074**	**2072/3**	0.91	0.01	0.10	0.05	0.10	>0.10	0.1	0.05	0.01	2073	1	<0.01	0.52	<0.01	20,20	Single, Stat	5
B4	**2085**	**2084/5**	0.51	0.01	0.10	0.10	0.10	0.1	0.05	0.01	0.01	2084	0	0.01	0.88	0.04	2,2	Single, N/A	5
C3	1970	1969/70	0.48	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1969	0	<0.01	<0.01	<0.01	22,26	Single, Stat	6
C	1987	Findable-misplaced																	
C3		1990/1	0.59	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2003	13	<0.01	0.07	<0.01	22,26	Single, Stat	7
C	1996	Findable (1990/1 may be intermediate)																	
C	2028	step<1SD																	
C3	**2036**	**2034/5**	0.48	0.01	0.10	0.08	0.10	0.01	0.01	0.01	0.01	2013	-21	0.02	0.56	0.02	26,26	Single, Stat	5
C3	**2050**	**2049/50**	1.01	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2049	0	<0.01	0.42	<0.01	18,20	Single, Stat	4
C3	**2076**	**2075/6**	0.51	0.01	0.10	0.02	0.10	>0.10	0.05	0.01	0.01	2075	0	<0.01	0.01	<0.01	21,26	Single, Stat	7
C	2095	Within end prohibition period																	
C4		1939/40	0.28	0.01	0.10	0.06	0.10	>0.10	0.1	0.01	0.01	1924	-15	0.01	0.63	0.02	23,23	Single, Stat	5
C4	**1970**	**1969/70**	0.66	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1969	0	<0.01	<0.01	<0.01	21,26	Single, Stat	6
C	1987	Missed																	
C4	**1996**	**1995/6**	0.79	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1986	-9	<0.01	<0.01	<0.01	22,26	Single, Stat	5

C4	**2028**	**2029/30**	0.65	0.01	0.10	0.01	0.10	0.1	0.01	0.01	0.01	2028	-1	<0.01	<0.01	<0.01	21,26	Single, Stat	7
C	2036	Was found in C3, but 2028 was not																	
C4	**2050**	**2049/50**	0.70	0.01	0.10	0.01	0.10	>0.10	0.1	0.01	0.01	2049	0	<0.01	0.08	<0.01	18,20	Single, Stat	7
C4	**2076**	**2075/6**	0.22	0.01	0.10	0.03	0.10	0.01	0.01	0.01	0.01	2094	19	0.16	0.03	0.03	24,26	Single, Stat	7
C	2095	Within end prohibition period																	
D	1951	step<1SD																	
D3	**1980**	**1979/80**	0.77	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1979	0	<0.01	0.48	<0.01	22,26	Single, Stat	5
D3	**2010**	**2009/10**	1.23	0.01	0.10	0.01	0.07	>0.10	0.1	0.01	0.01	2009	0	<0.01	0.01	<0.01	22,23	Single, Stat	6
	2018	Missed																	
D3	**2047**	**2046/7**	0.06	0.01	0.10	0.02	0.10	0.05	0.01	0.01	0.01	2037	-9	0.74	0.04	0.04	24,26	Single, Stat	7
	2058	Missed																	
D3	**2068**	**2067/8**	0.84	0.01	0.10	0.10	0.10	>0.10	0.1	0.01	0.01	2067	0	<0.01	0.01	<0.01	20,20	Single, Stat	5
D3		2079/80	0.37	0.01	0.10	0.10	0.10	0.1	0.1	0.01	0.01	2079	0	0.01	0.02	0.02	20,20	Single, Stat	7
D	2085	step<1SD																	
D	1951	step<1SD																	
D4	**1980**	**1979/80**	0.90	0.01	0.10	0.01	0.05	0.05	0.01	0.01	0.01	1979	0	<0.01	0.40	<0.01	25,25	Single, Stat	5
D4	**2010**	**2009/10**	1.15	0.01	0.10	0.01	0.05	>0.10	>0.10	0.01	0.01	2009	0	<0.01	<0.01	<0.01	22,26	Single, Stat	7
	2018	Missed																	
D4	**2047**	**2046/7**	0.35	0.01	0.10	0.01	0.06	>0.10	0.01	0.01	0.01	2028	-18	0.04	0.10	<0.01	21,26	Single, Stat	7
	2058	Missed																	
D4	**2068**	**2067/8**	0.85	0.01	0.10	0.03	0.10	>0.10	0.1	0.01	0.01	2067	0	<0.01	<0.01	<0.01	18,20	Single, Stat	5
D4	**2085**	**2084/5**	0.73	0.01	0.10	0.10	0.10	>0.10	>0.10	>0.10	0.1	2084	0	<0.01	0.61	<0.01	11,11	Single, Non-Stat	4
E	1955	step<1SD																	
E3	**1981**	**1978/9**	0.51	0.01	0.10	0.01	0.08	0.01	0.01	0.01	0.01	1917	-61	<0.01	0.89	<0.01	25,26	Single, Stat	5
E3	**2001**	**2000/1**	0.91	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2000	0	<0.01	0.02	<0.01	21,26	Single, Stat	7
E3	**2029**	**2028/9**	0.78	0.01	0.10	0.01	0.10	>0.10	0.1	0.01	0.01	2028	0	<0.01	<0.01	<0.01	21,23	Single, Stat	7
E	2039	Missed (and found in E4)																	

E3	**2057**	**2056/7**	0.66	0.01	0.10	0.02	0.10	>0.10	0.01	0.01	0.01	2056	0	<0.01	0.86	<0.01	18,20	Single, Stat	5
E3	**2071**	**2069/70**	0.41	0.01	0.10	0.01	0.10	>0.10	0.05	0.01	0.01	2082	13	0.02	0.52	<0.01	18,20	Single, Stat	5
E	2080	step<1SD																	
E	1955	step<1SD																	
E4		1963/4	0.29	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	1954	-9	0.01	<0.01	<0.01	22,26	Single, Stat	7
E	1981	Was found in E3, 1963/4 may be intermediate																	
E4	**2001**	**2000/1**	0.57	0.01	0.10	0.04	0.10	>0.10	0.01	0.01	0.01	2000	0	<0.01	0.49	<0.01	18,20	Single, Stat	7
E4	**2039**	**2038/9**	0.75	0.01	0.10	0.01	0.10	>0.10	0.01	0.01	0.01	2038	0	<0.01	0.32	<0.01	21,26	Single, Stat	7
E4	**2057**	**2056/7**	0.68	0.01	0.10	0.10	0.10	>0.10	0.01	0.01	0.01	2056	0	<0.01	0.10	<0.01	20,20	Single, Stat	7
E	2071	step<1SD, found in E3, but 2080 was not																	
E4	**2080**	**2081/2**	0.56	0.01	0.10	0.08	0.10	0.1	0.01	0.01	0.01	2070	-11	<0.01	0.44	<0.01	20,20	Single, Stat	7

Table A4.1.28: Heteroscedasticity testing of the data reported in Table A4.1.27.

Studentized Breusch-Pagan Test			
Dataset	**Break model**	**Linear model**	**Quad model**
A1	0.4173	0.1083	0.2215
A2	0.5127	0.1723	0.2215
A3	0.1362	0.05181	8.92E-05
A4	0.1828	0.02413	4.41E-05
B1	0.8859	0.7991	0.9164
B2	0.1509	0.1456	0.9164
B3	0.0499	0.000644	0.002369
B4	0.3204	0.001309	0.002369
C1	0.9586	0.9725	0.7825
C2	0.6503	0.9702	0.7825
C3	0.1712	7.62E-07	0.4551
C4	0.2684	5.79E-08	0.4551
D1	0.1267	0.2892	0.07087
D2	0.1115	0.3797	0.07087
D3	0.9354	0.00359	0.002098
D4	0.9567	0.001811	0.002098
E1	0.2575	0.2874	0.539
E2	0.0986	0.3317	0.539
E3	0.3737	0.4607	0.00036
E4	0.3147	0.116	0.00036

Table A4.1.29: classes of data segments/change-points and index values corresponding.

Single, Non-stationary																						
10,11	11,11	12,11	12,14	13,11	13,13	13,14	14,14	18,11	18,9	19,11	19,14	2,11	20,10	20,11	21,11	21,9	22,10	22,11	22,14	25,10	25,14	25,17
Single, Stationary																						
12,19	14,20	14,26	18,18	18,20	19,20	20,20	21,20	21,23	21,26	22,19	22,20	22,23	22,25	22,26	23,20	23,23	23,26	24,26	25,25	25,26	26,26	9,2
Single, N/A																						
0,0	0,0	0,2	11,1	11,2	2,2																	
Multiple, Stationary																						
11,19	11,20	12,18	12,23	13,20	13,26	14,23	14,24	16,26	22,22	9,18	9,20											
Non-stationary																						
12,10	9,0	9,1	9,11	9,9																		

Table A4.1.30: Extended analysis of the MSBV analysis first published by (Jones and Ricketts, 2017b illustrated in Figure 2).

Dataset				MSBV			KPSS-L		KPSS-T		ADF		Zivot Andrews			ANOVA/ANCOVA			Classification		Autocorrelation	
Source	Land/ Ocean	Composition	Boundary	First Changed Year	Internal Shift	Internal Trend Change	A	B	A	B	A	B	A	B	Year Of Change	Trend Change	Internal Shift	Change-point				
				Year	⁰C	⁰C/Year	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Year	Pr	Pr	Pr	Code	Class	Segment	Residual
NCDC	land	Zone	00N-30N	1903	-0.3634	0.0006	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1902	0.000	0.902	0.000	20,20	Single, Stat	0.4677	0.2699
NCDC	land	Zone	00N-30N	1926	0.1579	-0.0041	0.010	0.100	0.012	0.100	0.010	0.010	0.010	0.050	1942	0.044	0.422	0.012	25,26	Single, Stat	0.1628	0.0505
NCDC	land	Zone	00N-30N	1979	0.1574	0.0101	0.010	0.100	0.012	0.100	0.050	0.010	0.010	0.050	1977	0.095	0.170	0.002	25,26	Single, Stat	0.2018	0.0377
NCDC	land	Zone	00N-30N	1998	0.2968	0.0006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1997	0.011	0.956	0.035	20,20	Single, Stat	0.1133	-0.0846
NCDC	land	Hem	00N-90N	1921	0.2461	-0.0019	0.010	0.100	0.013	0.100	0.100	0.010	0.010	0.050	1936	0.001	0.457	0.001	22,26	Single, Stat	0.2443	0.1287
NCDC	land	Hem	00N-90N	1980	0.1979	0.0190	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.076	0.053	0.000	22,26	Single, Stat	0.2487	-0.0188
NCDC	land	Hem	00N-90N	1997	0.3231	-0.0074	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1997	0.012	0.538	0.034	20,20	Single, Stat	0.0417	-0.1054
NCDC	land	Composite	20N-90N	1921	0.2260	-0.0030	0.010	0.100	0.018	0.100	0.050	0.010	0.010	0.050	1963	0.009	0.340	0.004	25,26	Single, Stat	0.1636	0.0780
NCDC	land	Composite	20N-90N	1988	0.4654	0.0008	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.009	0.975	0.000	22,26	Single, Stat	0.2043	-0.0341
NCDC	land	Composite	20N-90N	1998	0.4185	0.0040	0.010	0.100	0.090	0.100	0.100	0.010	0.050	0.050	1997	0.007	0.858	0.019	2,2	Single, N/A	0.1560	-0.0421
NCDC	land	Tropic	20S-20N	1904	-0.1665	0.0021	0.010	0.100	0.100	0.100	1.000	0.010	0.010	0.050	1903	0.001	0.567	0.005	23,23	Single, Stat	0.2617	0.0802
NCDC	land	Tropic	20S-20N	1926	0.2176	0.0046	0.010	0.100	0.028	0.100	0.050	0.010	0.010	0.050	1939	0.001	0.295	0.004	19,20	Single, Stat	0.4317	0.2872
NCDC	land	Tropic	20S-20N	1957	0.1283	-0.0004	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1942	0.123	0.941	0.274	20,20	Single, Stat	0.1764	0.1645
NCDC	land	Tropic	20S-20N	1979	0.1856	0.0106	0.010	0.100	0.100	0.100	0.100	0.010	0.010	0.050	1976	0.069	0.234	0.056	23,20	Single, Stat	0.0677	-0.1185
NCDC	land	Tropic	20S-20N	1997	0.1226	0.0022	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	2001	0.220	0.821	0.447	20,20	Single, Stat	-0.0474	-0.0991
NCDC	land	Zone	30N-60N	1894	0.1648	-0.0134	0.014	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1902	0.198	0.338	0.110	20,20	Single, Stat	0.3264	0.2593
NCDC	land	Zone	30N-60N	1921	0.2804	0.0005	0.010	0.100	0.037	0.100	0.010	0.010	0.010	0.050	1913	0.015	0.936	0.018	25,26	Single, Stat	0.0490	-0.0410
NCDC	land	Zone	30N-60N	1981	0.2729	0.0162	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.050	1963	0.098	0.291	0.002	25,26	Single, Stat	0.0957	-0.0979
NCDC	land	Zone	30N-60N	1997	0.5381	-0.0195	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1997	0.007	0.314	0.014	20,20	Single, Stat	0.1949	0.0535
NCDC	land	Zone	30S-00N	1937	0.3318	-0.0092	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1939	0.000	0.114	0.000	22,26	Single, Stat	0.4862	0.3297
NCDC	land	Zone	30S-00N	1957	0.2612	0.0131	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1957	0.007	0.091	0.008	20,20	Single, Stat	0.0901	-0.0229
NCDC	land	Zone	30S-00N	1979	0.1987	0.0091	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1976	0.035	0.201	0.050	20,20	Single, Stat	0.0156	-0.1711
NCDC	land	Zone	30S-00N	2002	0.1682	-0.0086	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	2010	0.088	0.430	0.227	20,20	Single, Stat	-0.0212	-0.0391
NCDC	land	Zone	60N-90N	1920	0.6122	-0.0056	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.050	1949	0.002	0.439	0.001	25,26	Single, Stat	0.1165	0.0140
NCDC	land	Zone	60N-90N	1988	0.5594	0.0373	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.014	0.003	0.000	22,26	Single, Stat	0.2331	-0.0875
NCDC	land	Zone	60S-30S	1938	0.0919	0.0011	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1902	0.177	0.665	0.257	26,26	Single, Stat	0.2548	0.2349
NCDC	land	Zone	60S-30S	1977	0.2841	0.0005	0.010	0.100	0.030	0.100	0.010	0.010	0.010	0.050	1976	0.001	0.912	0.002	25,26	Single, Stat	0.1590	0.0212
NCDC	land	Zone	60S-30S	2003	0.0661	0.0225	0.010	0.100	0.036	0.100	0.050	0.010	0.010	0.050	1991	0.551	0.100	0.048	18,20	Single, Stat	0.1006	-0.0379
NCDC	land	Composite	60S-60N	1921	0.0952	0.0031	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1902	0.232	0.654	0.165	20,20	Single, Stat	0.4056	0.3660
NCDC	land	Composite	60S-60N	1938	0.1246	-0.0038	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1944	0.117	0.597	0.076	20,20	Single, Stat	0.0307	-0.0703
NCDC	land	Composite	60S-60N	1979	0.2154	0.0150	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1976	0.018	0.043	0.000	25,26	Single, Stat	0.2723	-0.0802
NCDC	land	Composite	60S-60N	1997	0.2926	-0.0065	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1997	0.003	0.472	0.011	20,20	Single, Stat	0.1425	-0.0299
NCDC	land	Hem	90S-00N	1937	0.3162	-0.0095	0.013	0.100	0.011	0.100	0.100	0.010	0.010	0.050	1939	0.000	0.086	0.000	22,26	Single, Stat	0.5056	0.3623
NCDC	land	Hem	90S-00N	1957	0.2448	0.0141	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1957	0.007	0.056	0.006	20,20	Single, Stat	0.0919	-0.0311
NCDC	land	Hem	90S-00N	1979	0.1946	0.0070	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1976	0.025	0.283	0.047	20,20	Single, Stat	0.0049	-0.1932

NCDC	land	Hem	90S-00N	2002	0.1445	-0.0036	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	2010	0.102	0.713	0.229	20,20	Single, Stat	0.0132	-0.0038
NCDC	land	Composite	90S-20S	1926	0.2360	-0.0014	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1943	0.011	0.760	0.033	26,26	Single, Stat	0.3227	0.2871
NCDC	land	Composite	90S-20S	1957	0.2275	0.0014	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1957	0.024	0.848	0.041	20,20	Single, Stat	-0.0060	-0.0548
NCDC	land	Composite	90S-20S	1977	0.2612	0.0071	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1976	0.008	0.353	0.026	20,20	Single, Stat	-0.1539	-0.2246
NCDC	land	Composite	90S-20S	2002	0.1470	0.0083	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	2010	0.168	0.483	0.103	20,20	Single, Stat	0.1099	0.0601
NCDC	land	Glob	90S-90N	1925	0.1829	0.0000	0.010	0.100	0.100	0.100	0.100	0.010	0.010	0.050	1936	0.003	0.994	0.009	23,26	Single, Stat	0.2498	0.1725
NCDC	land	Glob	90S-90N	1980	0.1990	0.0150	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.036	0.070	0.000	22,26	Single, Stat	0.2805	-0.0169
NCDC	land	Glob	90S-90N	1997	0.2500	-0.0042	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1997	0.016	0.672	0.050	20,20	Single, Stat	0.0041	-0.1222
NCDC	land_ocean	Zone	00N-30N	1904	-0.2319	0.0033	0.010	0.100	0.100	0.100	1.000	0.010	0.100	0.050	1906	0.002	0.538	0.008	14,20	Single, Stat	0.4077	0.2901
NCDC	land_ocean	Zone	00N-30N	1926	0.2545	-0.0025	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1942	0.000	0.594	0.000	22,26	Single, Stat	0.5108	0.3306
NCDC	land_ocean	Zone	00N-30N	1979	0.1113	0.0075	0.010	0.100	0.055	0.100	0.100	0.010	0.010	0.050	1970	0.150	0.239	0.007	23,26	Single, Stat	0.4178	0.3231
NCDC	land_ocean	Zone	00N-30N	1997	0.1398	-0.0022	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1986	0.070	0.760	0.186	20,20	Single, Stat	0.1575	0.0461
NCDC	land_ocean	Hem	00N-90N	1925	0.3062	0.0034	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1923	0.000	0.033	0.000	22,26	Single, Stat	0.5789	0.3782
NCDC	land_ocean	Hem	00N-90N	1987	0.2187	0.0050	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.013	0.702	0.000	22,23	Single, Stat	0.4441	0.2655
NCDC	land_ocean	Hem	00N-90N	1997	0.2179	0.0031	0.010	0.100	0.077	0.100	0.100	0.010	0.050	0.050	1996	0.004	0.771	0.010	2,2	Single, N/A	0.2037	-0.0534
NCDC	land_ocean	Composite	20N-90N	1925	0.3284	0.0003	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1920	0.000	0.854	0.000	22,26	Single, Stat	0.4666	0.2348
NCDC	land_ocean	Composite	20N-90N	1988	0.2890	0.0079	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.008	0.625	0.000	22,26	Single, Stat	0.3477	0.0848
NCDC	land_ocean	Composite	20N-90N	1998	0.2831	0.0004	0.010	0.100	0.074	0.100	1.000	0.010	1.000	0.050	1996	0.004	0.975	0.009	11,2	Single, N/A	0.3345	0.0951
NCDC	land_ocean	Tropic	20S-20N	1936	0.2062	0.0027	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1925	0.002	0.249	0.001	25,26	Single, Stat	0.4898	0.4085
NCDC	land_ocean	Tropic	20S-20N	1979	0.1812	0.0067	0.010	0.100	0.018	0.100	1.000	0.010	0.010	0.050	1945	0.041	0.351	0.004	22,26	Single, Stat	0.3748	0.2322
NCDC	land_ocean	Tropic	20S-20N	1997	0.1022	-0.0044	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	2001	0.235	0.588	0.435	20,20	Single, Stat	0.0897	0.0432
NCDC	land_ocean	Zone	30N-60N	1921	0.3372	0.0011	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1920	0.000	0.618	0.000	22,26	Single, Stat	0.3531	0.1448
NCDC	land_ocean	Zone	30N-60N	1988	0.3716	-0.0141	0.042	0.100	0.026	0.100	1.000	0.010	0.010	0.050	1963	0.005	0.513	0.000	22,26	Single, Stat	0.2080	0.0175
NCDC	land_ocean	Zone	30N-60N	1997	0.4279	0.0188	0.010	0.100	0.100	0.100	1.000	0.010	0.100	0.050	1996	0.001	0.336	0.002	11,2	Single, N/A	0.3494	0.0974
NCDC	land_ocean	Zone	30S-00N	1937	0.1599	0.0044	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1902	0.006	0.040	0.000	25,26	Single, Stat	0.5334	0.4487
NCDC	land_ocean	Zone	30S-00N	1979	0.1806	0.0044	0.010	0.100	0.010	0.100	0.100	0.010	0.010	0.050	1945	0.023	0.494	0.004	22,26	Single, Stat	0.4146	0.2756
NCDC	land_ocean	Zone	30S-00N	1997	0.1112	-0.0042	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	2001	0.106	0.516	0.229	20,20	Single, Stat	0.1007	0.0305
NCDC	land_ocean	Zone	60N-90N	1920	0.5775	0.0023	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.050	1919	0.000	0.664	0.000	25,26	Single, Stat	0.1758	0.0589
NCDC	land_ocean	Zone	60N-90N	1988	0.5325	0.0115	0.025	0.100	0.010	0.100	0.010	0.010	0.010	0.050	1963	0.018	0.631	0.000	25,26	Single, Stat	0.1837	-0.0004
NCDC	land_ocean	Zone	60N-90N	2002	0.3442	0.0105	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	2001	0.153	0.727	0.307	2,2	Single, N/A	-0.3315	-0.4835
NCDC	land_ocean	Zone	60S-30S	1887	-0.1860	0.0168	0.010	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1913	0.011	0.356	0.000	13,13	Single, Non-Stat	0.7557	0.6455
NCDC	land_ocean	Zone	60S-30S	1937	0.2404	0.0015	0.010	0.100	0.010	0.100	0.050	0.010	1.000	0.050	1931	0.000	0.478	0.000	16,26	Multiple, Stat	0.7425	0.5928
NCDC	land_ocean	Zone	60S-30S	1968	0.0937	0.0174	0.010	0.100	0.022	0.100	1.000	0.010	1.000	0.050	1962	0.127	0.081	0.000	12,11	Single, Non-Stat	0.6649	0.4876
NCDC	land_ocean	Zone	60S-30S	1977	0.0539	-0.0176	0.010	0.100	0.017	0.100	1.000	0.010	1.000	0.050	1981	0.123	0.006	0.001	9,11	Non-Stat	0.5229	0.2317
NCDC	land_ocean	Zone	60S-30S	1996	0.1049	0.0014	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1995	0.001	0.609	0.004	11,11	Single, Non-Stat	0.4300	0.3308
NCDC	land_ocean	Composite	60S-60N	1903	-0.1383	-0.0041	0.010	0.100	0.081	0.100	0.100	0.010	1.000	0.050	1894	0.018	0.573	0.006	11,11	Single, Non-Stat	0.5104	0.3625
NCDC	land_ocean	Composite	60S-60N	1914	0.1976	0.0023	0.024	0.100	0.100	0.100	0.100	0.010	1.000	0.050	1913	0.002	0.792	0.006	11,2	Single, N/A	0.4064	-0.0483
NCDC	land_ocean	Composite	60S-60N	1925	0.0985	0.0092	0.010	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1915	0.072	0.257	0.109	2,2	Single, N/A	0.1019	-0.0476
NCDC	land_ocean	Composite	60S-60N	1937	0.1147	-0.0048	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1945	0.050	0.534	0.008	20,20	Single, Stat	0.4269	0.3138
NCDC	land_ocean	Composite	60S-60N	1979	0.1392	0.0091	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1945	0.011	0.039	0.000	25,26	Single, Stat	0.5302	0.2636
NCDC	land_ocean	Composite	60S-60N	1997	0.1671	-0.0054	0.010	0.100	0.100	0.100	1.000	0.010	0.050	0.050	1996	0.001	0.251	0.004	20,20	Single, Stat	0.2840	0.0464

NCDC	land_ocean	Hem	90S-00N	1890	-0.1208	0.0143	0.018	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1911	0.027	0.114	0.000	22,22	Multiple, Stat	0.6735	0.5252
NCDC	land_ocean	Hem	90S-00N	1937	0.1979	-0.0010	0.010	0.100	0.026	0.100	1.000	0.010	0.100	0.050	1911	0.000	0.596	0.000	13,26	Multiple, Stat	0.6696	0.5522
NCDC	land_ocean	Hem	90S-00N	1969	0.1819	-0.0005	0.010	0.100	0.015	0.100	0.050	0.010	0.050	0.050	1945	0.019	0.963	0.006	18,20	Single, Stat	0.5191	0.3785
NCDC	land_ocean	Hem	90S-00N	1979	0.1195	0.0058	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1976	0.033	0.485	0.096	2,2	Single, N/A	0.0642	-0.1668
NCDC	land_ocean	Hem	90S-00N	1997	0.1067	-0.0032	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1996	0.008	0.381	0.023	20,20	Single, Stat	0.2041	0.0261
NCDC	land_ocean	Composite	90S-20S	1887	-0.1750	0.0121	0.010	0.100	0.010	0.100	1.000	0.010	0.050	0.050	1911	0.003	0.399	0.000	22,22	Multiple, Stat	0.7203	0.5540
NCDC	land_ocean	Composite	90S-20S	1937	0.2298	-0.0001	0.010	0.100	0.010	0.100	0.050	0.010	1.000	0.050	1931	0.000	0.958	0.000	16,26	Multiple, Stat	0.7108	0.5196
NCDC	land_ocean	Composite	90S-20S	1969	0.1863	0.0046	0.010	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1945	0.008	0.707	0.000	12,14	Single, Non-Stat	0.6678	0.4186
NCDC	land_ocean	Composite	90S-20S	1977	0.0761	-0.0013	0.010	0.100	0.100	0.100	0.100	0.010	0.100	0.050	1976	0.031	0.847	0.030	11,11	Single, Non-Stat	0.1841	0.0736
NCDC	land_ocean	Composite	90S-20S	1997	0.1039	-0.0003	0.010	0.100	0.100	0.100	0.100	0.010	1.000	0.050	1995	0.000	0.887	0.001	11,20	Multiple, Stat	0.4447	0.2407
NCDC	land_ocean	Glob	90S-90N	1930	0.2453	0.0032	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1913	0.000	0.024	0.000	22,26	Single, Stat	0.6645	0.4959
NCDC	land_ocean	Glob	90S-90N	1979	0.1157	0.0089	0.010	0.100	0.010	0.100	0.100	0.010	0.010	0.050	1945	0.033	0.045	0.000	22,26	Single, Stat	0.5345	0.3218
NCDC	land_ocean	Glob	90S-90N	1997	0.1564	-0.0049	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1996	0.002	0.286	0.007	20,20	Single, Stat	0.2471	0.0190
NCDC	ocean	Zone	00N-30N	1907	-0.2312	0.0100	0.015	0.100	0.100	0.100	0.050	0.010	0.100	0.050	1906	0.006	0.128	0.020	11,20	Multiple, Stat	0.3654	0.2672
NCDC	ocean	Zone	00N-30N	1926	0.2385	-0.0061	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1945	0.001	0.281	0.000	22,26	Single, Stat	0.5913	0.4115
NCDC	ocean	Zone	00N-30N	1987	0.1487	0.0051	0.010	0.100	0.010	0.100	0.100	0.010	0.010	0.050	1969	0.016	0.117	0.000	22,26	Single, Stat	0.5029	0.4081
NCDC	ocean	Hem	00N-90N	1903	-0.2528	0.0106	0.010	0.100	0.023	0.100	1.000	0.010	1.000	0.050	1906	0.000	0.015	0.000	12,11	Single, Non-Stat	0.6105	0.3778
NCDC	ocean	Hem	00N-90N	1926	0.2671	-0.0079	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1945	0.000	0.012	0.000	22,26	Single, Stat	0.7637	0.4696
NCDC	ocean	Hem	00N-90N	1987	0.1271	0.0071	0.037	0.100	0.023	0.100	0.100	0.010	0.050	0.050	1969	0.073	0.501	0.001	22,26	Single, Stat	0.5975	0.5100
NCDC	ocean	Hem	00N-90N	1997	0.1307	0.0001	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1996	0.035	0.989	0.074	11,11	Single, Non-Stat	0.3100	0.1302
NCDC	ocean	Composite	20N-90N	1902	-0.2142	-0.0035	0.010	0.100	0.043	0.100	1.000	0.010	1.000	0.050	1901	0.000	0.505	0.000	9,11	Non-Stat	0.5963	0.3657
NCDC	ocean	Composite	20N-90N	1915	0.1361	0.0245	0.010	0.100	0.020	0.100	1.000	0.010	1.000	0.050	1907	0.006	0.000	0.000	9,11	Non-Stat	0.4710	0.1796
NCDC	ocean	Composite	20N-90N	1930	0.1341	-0.0154	0.010	0.100	0.010	0.100	1.000	0.010	0.100	0.050	1945	0.005	0.002	0.000	13,20	Multiple, Stat	0.6873	0.3986
NCDC	ocean	Composite	20N-90N	1964	-0.1560	-0.0036	0.010	0.100	0.025	0.100	0.100	0.010	0.050	0.050	1963	0.000	0.181	0.000	22,20	Single, Stat	0.5088	0.3497
NCDC	ocean	Composite	20N-90N	1988	0.1968	0.0056	0.014	0.100	0.019	0.100	1.000	0.010	1.000	0.050	1987	0.007	0.605	0.001	9,11	Non-Stat	0.4869	0.2180
NCDC	ocean	Composite	20N-90N	1997	0.1951	0.0067	0.010	0.100	0.090	0.100	1.000	0.010	1.000	0.050	1993	0.003	0.514	0.008	11,11	Single, Non-Stat	0.5214	0.2013
NCDC	ocean	Tropic	20S-20N	1926	0.2562	0.0060	0.010	0.100	0.032	0.100	0.010	0.010	0.010	0.050	1924	0.000	0.008	0.000	25,26	Single, Stat	0.4736	0.3423
NCDC	ocean	Tropic	20S-20N	1979	0.1610	0.0049	0.010	0.100	0.034	0.100	0.100	0.010	0.010	0.050	1945	0.058	0.477	0.006	22,26	Single, Stat	0.3812	0.2794
NCDC	ocean	Tropic	20S-20N	1997	0.0959	-0.0064	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	2001	0.261	0.431	0.405	20,20	Single, Stat	0.1333	0.0840
NCDC	ocean	Zone	30N-60N	1902	-0.3433	0.0031	0.010	0.100	0.038	0.100	1.000	0.010	0.100	0.050	1901	0.000	0.619	0.000	9,11	Non-Stat	0.5852	0.1488
NCDC	ocean	Zone	30N-60N	1915	0.2074	0.0135	0.010	0.100	0.100	0.100	0.050	0.010	0.100	0.050	1914	0.001	0.057	0.001	11,2	Single, N/A	0.3203	0.0827
NCDC	ocean	Zone	30N-60N	1930	0.1320	-0.0113	0.010	0.100	0.017	0.100	0.010	0.010	0.100	0.050	1938	0.050	0.103	0.002	10,11	Single, Non-Stat	0.4667	0.3173
NCDC	ocean	Zone	30N-60N	1964	-0.1841	-0.0063	0.010	0.100	0.020	0.100	0.100	0.010	0.010	0.050	1963	0.003	0.086	0.000	22,26	Single, Stat	0.4164	0.2517
NCDC	ocean	Zone	30N-60N	1989	0.2803	0.0068	0.022	0.100	0.010	0.100	0.100	0.010	0.050	0.050	1988	0.002	0.621	0.000	18,20	Single, Stat	0.4497	0.0546
NCDC	ocean	Zone	30N-60N	1998	0.2108	0.0074	0.010	0.100	0.100	0.100	1.000	0.010	0.100	0.050	1993	0.012	0.586	0.034	11,2	Single, N/A	0.4639	0.2389
NCDC	ocean	Zone	30S-00N	1937	0.1489	0.0029	0.010	0.100	0.016	0.100	0.050	0.010	0.010	0.050	1911	0.012	0.184	0.006	25,26	Single, Stat	0.5315	0.4737
NCDC	ocean	Zone	30S-00N	1979	0.1830	0.0036	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1945	0.024	0.576	0.006	22,26	Single, Stat	0.4570	0.3387
NCDC	ocean	Zone	30S-00N	1997	0.1307	-0.0056	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	2001	0.045	0.356	0.095	20,20	Single, Stat	0.1490	0.0353
NCDC	ocean	Zone	60N-90N	1926	0.4000	0.0020	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1925	0.000	0.243	0.000	22,25	Single, Stat	0.6016	0.3729
NCDC	ocean	Zone	60N-90N	2000	0.4789	0.0249	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1981	0.000	0.005	0.000	22,25	Single, Stat	0.8191	0.3889
NCDC	ocean	Zone	60S-30S	1887	-0.1958	0.0171	0.010	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1913	0.011	0.371	0.000	13,13	Single, Non-Stat	0.7703	0.6657

NCDC	ocean	Zone	60S-30S	1937	0.2358	0.0029	0.010	0.100	0.010	0.100	0.050	0.010	1.000	0.050	1931	0.000	0.149	0.000	16,26	Multiple, Stat	0.7551	0.6146
NCDC	ocean	Zone	60S-30S	1970	0.1396	0.0111	0.010	0.100	0.025	0.100	1.000	0.010	1.000	0.050	1962	0.060	0.463	0.001	12,11	Single, Non-Stat	0.6659	0.4450
NCDC	ocean	Zone	60S-30S	1977	0.0496	-0.0122	0.010	0.100	0.084	0.100	1.000	0.010	1.000	0.050	1981	0.169	0.146	0.018	11,11	Single, Non-Stat	0.4941	0.2948
NCDC	ocean	Zone	60S-30S	1996	0.1064	0.0001	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1995	0.001	0.971	0.004	11,11	Single, Non-Stat	0.4756	0.3842
NCDC	ocean	Composite	60S-60N	1890	-0.1357	0.0080	0.016	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1913	0.039	0.452	0.003	13,13	Single, Non-Stat	0.7049	0.6223
NCDC	ocean	Composite	60S-60N	1930	0.2165	0.0011	0.010	0.100	0.022	0.100	1.000	0.010	1.000	0.050	1913	0.000	0.540	0.000	13,26	Multiple, Stat	0.7148	0.6017
NCDC	ocean	Composite	60S-60N	1977	0.1088	-0.0017	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1945	0.116	0.872	0.071	20,20	Single, Stat	0.5643	0.5356
NCDC	ocean	Composite	60S-60N	1987	0.1004	0.0013	0.013	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1986	0.027	0.853	0.076	11,11	Single, Non-Stat	0.0923	0.0138
NCDC	ocean	Composite	60S-60N	1997	0.1364	0.0013	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1996	0.005	0.849	0.011	11,11	Single, Non-Stat	0.3655	0.0690
NCDC	ocean	Hem	90S-00N	1890	-0.1235	0.0161	0.018	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1911	0.022	0.072	0.000	22,22	Multiple, Stat	0.7140	0.5597
NCDC	ocean	Hem	90S-00N	1937	0.2039	-0.0023	0.010	0.100	0.039	0.100	1.000	0.010	1.000	0.050	1911	0.000	0.254	0.000	13,26	Multiple, Stat	0.7079	0.5961
NCDC	ocean	Hem	90S-00N	1969	0.1806	0.0025	0.017	0.100	0.010	0.100	1.000	0.010	0.050	0.050	1945	0.017	0.815	0.002	21,11	Single, Non-Stat	0.6186	0.4779
NCDC	ocean	Hem	90S-00N	1979	0.0993	0.0027	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1976	0.035	0.702	0.092	2,2	Single, N/A	0.1025	-0.1034
NCDC	ocean	Hem	90S-00N	1997	0.1158	-0.0044	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1996	0.001	0.164	0.003	20,20	Single, Stat	0.3183	0.0568
NCDC	ocean	Composite	90S-20S	1887	-0.1840	0.0130	0.010	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1911	0.003	0.386	0.000	13,13	Single, Non-Stat	0.7778	0.6367
NCDC	ocean	Composite	90S-20S	1937	0.2456	-0.0001	0.010	0.100	0.010	0.100	0.100	0.010	1.000	0.050	1933	0.000	0.975	0.000	13,26	Multiple, Stat	0.7678	0.5928
NCDC	ocean	Composite	90S-20S	1969	0.1653	0.0121	0.013	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1944	0.017	0.322	0.000	12,14	Single, Non-Stat	0.7423	0.5159
NCDC	ocean	Composite	90S-20S	1977	0.0404	-0.0096	0.010	0.100	0.069	0.100	1.000	0.010	1.000	0.050	1987	0.185	0.118	0.021	11,11	Single, Non-Stat	0.4108	0.2906
NCDC	ocean	Composite	90S-20S	1996	0.1084	-0.0002	0.010	0.100	0.100	0.100	1.000	0.010	1.000	0.050	1995	0.000	0.924	0.001	11,11	Single, Non-Stat	0.5678	0.4655
NCDC	ocean	Glob	90S-90N	1890	-0.1330	0.0083	0.014	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1913	0.040	0.433	0.002	13,13	Single, Non-Stat	0.7077	0.6258
NCDC	ocean	Glob	90S-90N	1930	0.2200	0.0010	0.010	0.100	0.020	0.100	1.000	0.010	1.000	0.050	1913	0.000	0.560	0.000	13,26	Multiple, Stat	0.7247	0.6065
NCDC	ocean	Glob	90S-90N	1977	0.1081	-0.0017	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1945	0.111	0.865	0.066	20,20	Single, Stat	0.5733	0.5439
NCDC	ocean	Glob	90S-90N	1987	0.0973	0.0021	0.012	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1986	0.028	0.771	0.075	11,11	Single, Non-Stat	0.0952	0.0162
NCDC	ocean	Glob	90S-90N	1997	0.1350	0.0011	0.010	0.100	0.100	0.100	0.100	0.010	1.000	0.050	1996	0.004	0.875	0.010	11,11	Single, Non-Stat	0.3782	0.0784
GISSTEMP	land_ocean	Zone	24N-44N	1930	0.2245	0.0042	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1902	0.000	0.081	0.000	25,26	Single, Stat	0.6076	0.4507
GISSTEMP	land_ocean	Zone	24N-44N	1964	-0.2122	-0.0016	0.010	0.100	0.013	0.100	0.050	0.010	0.010	0.050	1963	0.000	0.640	0.000	19,20	Single, Stat	0.2224	-0.0122
GISSTEMP	land_ocean	Zone	24N-44N	1987	0.2263	0.0014	0.010	0.100	0.090	0.100	0.050	0.010	0.050	0.050	1986	0.025	0.916	0.022	20,20	Single, Stat	0.1992	-0.0138
GISSTEMP	land_ocean	Zone	24N-44N	1998	0.4256	-0.0032	0.010	0.100	0.096	0.100	1.000	0.010	0.010	0.050	1997	0.000	0.813	0.000	2,2	Single, N/A	0.3765	0.0128
GISSTEMP	land_ocean	Composite	24N-90N	1921	0.3951	-0.0008	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1929	0.000	0.712	0.000	22,26	Single, Stat	0.4537	0.2004
GISSTEMP	land_ocean	Composite	24N-90N	1988	0.3278	0.0077	0.050	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.014	0.697	0.000	22,26	Single, Stat	0.3438	0.1436
GISSTEMP	land_ocean	Composite	24N-90N	1998	0.3281	0.0041	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1993	0.009	0.822	0.024	11,2	Single, N/A	0.2835	0.1003
GISSTEMP	land_ocean	Tropic	24S-24N	1903	-0.2330	0.0044	0.018	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1906	0.004	0.456	0.014	20,20	Single, Stat	0.3603	0.2485
GISSTEMP	land_ocean	Tropic	24S-24N	1926	0.1913	-0.0028	0.010	0.100	0.024	0.100	0.050	0.010	0.010	0.050	1945	0.008	0.552	0.002	25,26	Single, Stat	0.4005	0.2838
GISSTEMP	land_ocean	Tropic	24S-24N	1979	0.1669	0.0063	0.010	0.100	0.014	0.100	0.100	0.010	0.010	0.050	1945	0.040	0.342	0.001	22,26	Single, Stat	0.3825	0.2473
GISSTEMP	land_ocean	Tropic	24S-24N	1997	0.1121	-0.0041	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	2001	0.189	0.615	0.379	20,20	Single, Stat	0.0856	0.0293
GISSTEMP	land_ocean	Zone	24S-00N	1937	0.1606	0.0030	0.010	0.100	0.018	0.100	0.010	0.010	0.010	0.050	1911	0.015	0.208	0.008	25,26	Single, Stat	0.4539	0.3946
GISSTEMP	land_ocean	Zone	24S-00N	1979	0.1907	0.0065	0.010	0.100	0.010	0.100	0.100	0.010	0.010	0.050	1945	0.032	0.366	0.004	22,26	Single, Stat	0.3844	0.2408
GISSTEMP	land_ocean	Zone	24S-00N	1997	0.1156	-0.0085	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	2001	0.172	0.295	0.244	20,20	Single, Stat	0.0888	0.0173
GISSTEMP	land_ocean	Zone	44N-64N	1920	0.3010	-0.0028	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1963	0.003	0.444	0.001	25,26	Single, Stat	0.0866	-0.0244
GISSTEMP	land_ocean	Zone	44N-64N	1988	0.5660	-0.0242	0.024	0.100	0.011	0.100	1.000	0.010	0.010	0.050	1963	0.008	0.495	0.001	22,26	Single, Stat	0.1050	-0.0807
GISSTEMP	land_ocean	Zone	44N-64N	1997	0.4503	0.0283	0.010	0.100	0.100	0.100	1.000	0.010	0.050	0.050	2008	0.008	0.302	0.026	2,2	Single, N/A	0.1519	0.0027

GISSTEMP	land_ocean	Zone	44S-24S	1887	-0.1067	0.0222	0.012	0.100	0.010	0.100	0.100	0.010	0.010	0.050	1912	0.092	0.166	0.000	22,22	Multiple, Stat	0.6199	0.4882
GISSTEMP	land_ocean	Zone	44S-24S	1933	0.2179	0.0007	0.010	0.100	0.025	0.100	0.050	0.010	0.050	0.050	1931	0.000	0.671	0.000	25,26	Single, Stat	0.6707	0.5614
GISSTEMP	land_ocean	Zone	44S-24S	1970	0.2322	0.0028	0.010	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1944	0.000	0.594	0.000	13,11	Single, Non-Stat	0.7084	0.4786
GISSTEMP	land_ocean	Zone	44S-24S	1985	0.0799	-0.0029	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1986	0.088	0.629	0.226	11,11	Single, Non-Stat	0.4384	0.4186
GISSTEMP	land_ocean	Zone	44S-24S	1997	0.1312	0.0100	0.010	0.100	0.100	0.100	1.000	0.010	1.000	0.050	1996	0.001	0.028	0.001	11,11	Single, Non-Stat	0.5676	0.2629
GISSTEMP	land_ocean	Zone	64N-90N	1903	0.7491	-0.0336	0.010	0.100	0.061	0.100	0.050	0.010	0.010	0.050	1902	0.000	0.057	0.001	20,20	Single, Stat	0.1026	-0.1170
GISSTEMP	land_ocean	Zone	64N-90N	1920	1.0669	0.0312	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1947	0.000	0.135	0.000	22,25	Single, Stat	0.3371	0.1444
GISSTEMP	land_ocean	Zone	64N-90N	1995	0.7989	0.0025	0.035	0.100	0.010	0.100	0.010	0.010	0.010	0.050	1963	0.013	0.958	0.000	25,25	Single, Stat	0.2709	0.0625
GISSTEMP	land_ocean	Zone	64N-90N	2005	0.7699	-0.0179	0.016	0.100	0.100	0.100	0.100	0.010	0.010	0.050	2004	0.026	0.745	0.078	2,2	Single, N/A	-0.1172	-0.1973
GISSTEMP	land_ocean	Zone	64S-44S	1887	-0.1932	0.0052	0.010	0.100	0.013	0.100	1.000	0.010	1.000	0.050	1886	0.001	0.640	0.000	9,2	Single, Stat	0.5013	0.0192
GISSTEMP	land_ocean	Zone	64S-44S	1904	-0.0518	0.0028	0.010	0.100	0.100	0.100	1.000	0.010	0.050	0.050	1911	0.170	0.401	0.108	23,23	Single, Stat	0.3676	0.2985
GISSTEMP	land_ocean	Zone	64S-44S	1937	0.2183	0.0068	0.010	0.100	0.010	0.100	1.000	0.010	1.000	0.050	1935	0.000	0.068	0.000	12,23	Multiple, Stat	0.7739	0.3562
GISSTEMP	land_ocean	Zone	64S-44S	1952	0.0822	-0.0073	0.010	0.100	0.100	0.100	0.100	0.010	0.010	0.050	1951	0.032	0.085	0.027	20,20	Single, Stat	0.1124	-0.0693
GISSTEMP	land_ocean	Zone	64S-44S	1968	0.1529	0.0037	0.019	0.100	0.021	0.100	1.000	0.010	0.100	0.050	1967	0.012	0.708	0.006	9,2	Single, Stat	0.2782	-0.1078
GISSTEMP	land_ocean	Zone	64S-44S	1976	0.1762	-0.0016	0.010	0.100	0.010	0.100	1.000	0.010	0.050	0.050	1974	0.000	0.849	0.000	21,20	Single, Stat	0.5279	0.0659
GISSTEMP	land_ocean	Zone	64S-44S	2007	-0.1806	0.0212	0.023	0.100	0.100	0.100	0.050	0.010	0.010	0.050	2006	0.000	0.017	0.002	20,20	Single, Stat	0.2526	0.0239
GISSTEMP	land_ocean	Composite	90S-24S	1887	-0.1124	0.0165	0.010	0.100	0.048	0.100	1.000	0.010	0.100	0.050	1886	0.021	0.116	0.006	9,11	Non-Stat	0.2798	-0.1914
GISSTEMP	land_ocean	Composite	90S-24S	1900	-0.1050	0.0039	0.082	0.100	0.011	0.100	1.000	0.010	1.000	0.050	1912	0.008	0.402	0.001	13,14	Single, Non-Stat	0.6658	0.5115
GISSTEMP	land_ocean	Composite	90S-24S	1937	0.1888	-0.0009	0.010	0.100	0.063	0.100	1.000	0.010	0.100	0.050	1935	0.000	0.580	0.000	14,26	Single, Stat	0.6277	0.4447
GISSTEMP	land_ocean	Composite	90S-24S	1970	0.2312	0.0038	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1968	0.000	0.141	0.000	22,26	Single, Stat	0.5638	0.2263
GISSTEMP	land_ocean	Composite	90S-24S	1996	0.0681	0.0009	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1981	0.132	0.803	0.245	20,20	Single, Stat	0.0507	0.0319
GISSTEMP	land_ocean	Zone	90S-64S	1912	-0.4806	0.0204	0.010	0.100	0.059	0.100	0.050	0.010	0.010	0.050	1935	0.099	0.152	0.049	26,26	Single, Stat	0.1476	0.0910
GISSTEMP	land_ocean	Zone	90S-64S	1955	0.4264	0.0062	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1935	0.069	0.462	0.183	26,26	Single, Stat	0.1021	0.0787
GISSTEMP	land_ocean	Zone	00N-24N	1904	-0.2392	0.0044	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1906	0.006	0.479	0.019	20,20	Single, Stat	0.3111	0.1900
GISSTEMP	land_ocean	Zone	00N-24N	1926	0.2670	-0.0035	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1925	0.001	0.512	0.000	22,26	Single, Stat	0.4483	0.2669
GISSTEMP	land_ocean	Zone	00N-24N	1979	0.1689	0.0043	0.010	0.100	0.073	0.100	0.100	0.010	0.010	0.050	1942	0.063	0.605	0.008	23,26	Single, Stat	0.3738	0.2710
GISSTEMP	land_ocean	Zone	00N-24N	1995	0.1201	0.0038	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1986	0.184	0.673	0.397	20,20	Single, Stat	0.1266	0.0829
GISSTEMP	land_ocean	Glob	90S-90N	1902	-0.1442	0.0050	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1906	0.008	0.282	0.027	11,11	Single, Non-Stat	0.4493	0.3964
GISSTEMP	land_ocean	Glob	90S-90N	1920	0.0655	0.0058	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1906	0.223	0.271	0.228	20,20	Single, Stat	0.2552	0.1911
GISSTEMP	land_ocean	Glob	90S-90N	1937	0.1013	-0.0090	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1944	0.037	0.041	0.000	22,23	Single, Stat	0.4055	0.1996
GISSTEMP	land_ocean	Glob	90S-90N	1980	0.1581	0.0102	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.005	0.034	0.000	22,26	Single, Stat	0.5160	0.1997
GISSTEMP	land_ocean	Glob	90S-90N	1997	0.1463	-0.0020	0.010	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1996	0.014	0.728	0.045	20,20	Single, Stat	0.1760	0.0052
GISSTEMP	land_ocean	Hem	00N-90N	1924	0.3340	0.0016	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1920	0.000	0.356	0.000	22,26	Single, Stat	0.5573	0.3358
GISSTEMP	land_ocean	Hem	00N-90N	1987	0.2295	0.0079	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.022	0.598	0.000	22,23	Single, Stat	0.4379	0.2598
GISSTEMP	land_ocean	Hem	00N-90N	1997	0.2283	0.0036	0.010	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1994	0.013	0.786	0.034	2,2	Single, N/A	0.1535	-0.0507
GISSTEMP	land_ocean	Hem	90S-00N	1890	-0.1074	0.0143	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1911	0.039	0.099	0.000	22,19	Single, Stat	0.6578	0.5132
GISSTEMP	land_ocean	Hem	90S-00N	1937	0.1989	-0.0004	0.010	0.100	0.014	0.100	0.100	0.010	0.050	0.050	1911	0.000	0.809	0.000	22,26	Single, Stat	0.6573	0.4907
GISSTEMP	land_ocean	Glob	90S-00N	1969	0.2135	-0.0040	0.010	0.100	0.010	0.100	1.000	0.010	0.050	0.050	1945	0.001	0.660	0.000	21,20	Single, Stat	0.4561	0.1736
GISSTEMP	land_ocean	Glob	90S-00N	1979	0.1399	0.0079	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.050	1976	0.030	0.408	0.091	11,2	Single, N/A	0.1511	-0.0668
GISSTEMP	land_ocean	Glob	90S-00N	1996	0.0890	0.0007	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1995	0.058	0.877	0.161	20,20	Single, Stat	0.0547	0.0037
CW	land_ocean	Glob	90S-90N	1920	0.1204	0.0085	0.065	0.100	0.013	0.100	0.010	0.010	0.010	0.050	1901	0.051	0.121	0.000	25,26	Single, Stat	0.5864	0.4551

CW	land_ocean	Glob	90S-90N	1937	0.1495	-0.0099	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1936	0.003	0.030	0.000	22,26	Single, Stat	0.4293	0.1207
CW	land_ocean	Glob	90S-90N	1980	0.1424	0.0131	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1976	0.011	0.007	0.000	22,26	Single, Stat	0.4929	0.1192
CW	land_ocean	Glob	90S-90N	1997	0.1530	-0.0002	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1996	0.016	0.974	0.053	20,20	Single, Stat	0.1315	-0.0268
HadCRU	land_ocean	Tropic	30S-30N	1926	0.2502	0.0002	0.010	0.100	0.020	0.100	0.010	0.010	0.010	0.050	1925	0.000	0.872	0.000	25,26	Single, Stat	0.4686	0.3577
HadCRU	land_ocean	Tropic	30S-30N	1979	0.1153	0.0089	0.010	0.100	0.024	0.100	0.100	0.010	0.010	0.050	1973	0.115	0.140	0.001	22,26	Single, Stat	0.3580	0.2231
HadCRU	land_ocean	Tropic	30S-30N	1997	0.1527	-0.0046	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	2001	0.054	0.539	0.134	20,20	Single, Stat	0.1289	0.0145
HadCRU	land_ocean	Hem	00N-90N	1925	0.3107	0.0016	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1920	0.000	0.169	0.000	25,26	Single, Stat	0.5199	0.3156
HadCRU	land_ocean	Hem	00N-90N	1987	0.2119	0.0078	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1970	0.021	0.570	0.000	22,23	Single, Stat	0.4267	0.2463
HadCRU	land_ocean	Hem	00N-90N	1997	0.2148	0.0038	0.010	0.100	0.080	0.100	0.100	0.010	0.100	0.050	1996	0.012	0.756	0.031	11,2	Single, N/A	0.1781	0.0129
HadCRU	land_ocean	Hem	90S-00N	1937	0.2477	-0.0003	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1883	0.000	0.846	0.000	22,26	Single, Stat	0.5959	0.4345
HadCRU	land_ocean	Hem	90S-00N	1979	0.1559	0.0056	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1976	0.006	0.215	0.000	22,26	Single, Stat	0.3858	0.1257
HadCRU	land_ocean	Hem	90S-00N	1997	0.1308	-0.0043	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.050	1996	0.005	0.316	0.015	20,20	Single, Stat	0.1354	-0.0733
HadCRU	land_ocean	Glob	90S-90N	1930	0.2683	0.0004	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1929	0.000	0.739	0.000	22,26	Single, Stat	0.6145	0.4295
HadCRU	land_ocean	Glob	90S-90N	1979	0.0949	0.0111	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1963	0.085	0.016	0.000	22,26	Single, Stat	0.4569	0.2199
HadCRU	land_ocean	Glob	90S-90N	1997	0.1665	-0.0046	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1996	0.003	0.368	0.010	20,20	Single, Stat	0.1961	-0.0104
BERKLEY	land_ocean	Glob	90S-90N	1921	0.1255	0.0064	0.100	0.100	0.022	0.100	1.000	0.010	0.010	0.050	1901	0.067	0.325	0.000	22,26	Single, Stat	0.5610	0.4558
BERKLEY	land_ocean	Glob	90S-90N	1937	0.1798	-0.0071	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.050	1936	0.001	0.151	0.000	22,26	Single, Stat	0.4223	0.1021
BERKLEY	land_ocean	Glob	90S-90N	1980	0.1411	0.0114	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1976	0.011	0.017	0.000	25,26	Single, Stat	0.4458	0.0851
BERKLEY	land_ocean	Glob	90S-90N	1997	0.1566	-0.0016	0.010	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1996	0.015	0.797	0.048	20,20	Single, Stat	0.1314	-0.0394

Appendix 5.1: Detailed tables of results and analyses

Table A5.1.31: Results of the studentized Breusch-Pagan tests for each zone and for combined land/ocean, land and ocean sub-divisions. Three different models are applied. Probabilities are listed for each for the null case of homoscedasticity. The quadratic ("Quad") model tests for a simple quadratic change as would be the case if temperature were determined solely by the CO_2 concentrations. The Linear model simply provides a base case of at most a constant linear change. The Break model applies the deduced change-points and tests that the residuals of the segmented model are homoscedastic. Throughout red or green indicate probabilities: red is Pr <=0.01, green is 0.01<Pr<=0.05. Grey shading indicates differences between the two data sets.

Dataset			NCDC-z			GISS-g		
Split	**Composition**	**Zone**	**Break**	**Linear**	**Quad**	**Break**	**Linear**	**Quad**
Land/ocean	**Zone**	**60N.90N**	**0.6783**	**0.0197**	**0.0959**	**0.1138**	***<0.0001***	**0.0016**
Land/ocean	**Zone**	**30N.60N**	**0.1264**	**<0.0001**	**0.0069**	**0.1953**	***<0.0001***	**0.0241**
Land/ocean	**Zone**	**00N.30N**	**0.5625**	**0.2948**	**0.102**	**0.7182**	**0.8594**	**0.1212**
Land/ocean	**Zone**	**30S.00N**	**0.3024**	**0.0609**	**0.3988**	**0.0151**	**0.1153**	**0.0796**
Land/ocean	**Zone**	**60S.30S**	**0.0348**	**<0.0001**	**0.0061**	**0.2603**	**0.0001**	**0.1301**
Land/ocean	Comp	60S.60N	0.1873	0.2028	0.0381	0.4146	0.3587	0.0822
Land/ocean	Comp	20N.90N	0.1367	0	0.0015	0.1670	0.0000	0.0003
Land/ocean	Tropic	20S.20N	0.5204	0.1527	0.4462	0.7780	0.7189	0.3341
Land/ocean	Comp	90S.20S	0.0898	<0.0001	0.0739	0.6994	0.0113	0.0611
Land/ocean	Hem	*00N.90N*	*0.2204*	*0.0085*	*0.0059*	0.2837	*<0.0001*	0.0019
Land/ocean	Hem	*90S.00N*	*0.0911*	*0.0001*	*0.3718*	0.0258	0.0047	0.0795
Land/ocean	Glob	*90S.90N*	*0.1386*	*0.3796*	*0.0118*	0.7458	0.2164	0.0120
Land	**Zone**	**60N.90N**	**0.5756**	**0.0718**	**0.1943**	**0.0564**	**0.0033**	**0.0038**
Land	**Zone**	**30N.60N**	**0.0891**	**0.0002**	**0.0463**	**0.0963**	**0.0001**	**0.0515**
Land	**Zone**	**00N.30N**	**0.4082**	**0.0131**	**0.1123**	**0.7605**	**0.0002**	**0.1290**
Land	**Zone**	**30S.00N**	**0.4905**	**0.4223**	**0.6494**	**0.4038**	**0.9471**	**0.3275**
Land	**Zone**	**60S.30S**	**0.9761**	**0.6877**	**0.4879**	**0.7878**	**0.6080**	**0.3418**
Land	Comp	60S.60N	0.4914	0.0005	0.1592	0.1819	0.0124	0.1347
Land	Comp	20N.90N	0.0735	<0.0001	0.0076	0.1177	0.0000	0.0038
Land	Tropic	20S.20N	0.1013	0.9545	0.0551	0.0828	0.0340	0.0679
Land	Comp	90S.20S	0.0293	0.5468	0.3089	0.0007	0.0003	0.0016
Land	Hem	*00N.90N*	*0.167*	*<0.0001*	*0.0077*	0.0908	<0.0001	0.0040
Land	Hem	*90S.00N*	*0.3226*	*0.4167*	*0.4958*	0.0163	0.0392	0.0090
Land	Glob	*90S.90N*	*0.358*	*0.0001*	*0.018*	0.0849	<0.0001	0.0065
Ocean	**Zone**	**60N.90N**	**0.0459**	**<0.0000**	**0.0186**	**0.1316**	**0.0000**	**0.0006**
Ocean	**Zone**	**30N.60N**	**0.1553**	**0.6011**	**0.0624**	**0.0064**	**0.5199**	**0.2017**
Ocean	**Zone**	**00N.30N**	**0.5979**	**0.0071**	**0.0344**	**0.2652**	**0.0126**	**0.0459**
Ocean	**Zone**	**30S.00N**	**0.1859**	**0.0429**	**0.2434**	**0.0145**	**0.0461**	**0.0430**
Ocean	**Zone**	**60S.30S**	**0.0356**	**<0.0001**	**0.0027**	**0.2718**	***<0.0001***	**0.1463**
Ocean	Comp	60S.60N	0.0735	0.0001	0.0098	0.1063	0.0160	0.0398
Ocean	Comp	20N.90N	0.2484	0.9081	0.0197	0.0095	0.0085	0.0361
Ocean	Tropic	20S.20N	0.3875	0.0673	0.3192	0.5194	0.2038	0.3779
Ocean	Comp	90S.20S	0.0402	<0.0001	0.0375	0.3650	0.0002	0.1105
Ocean	Hem	*00N.90N*	*0.37*	*0.0663*	*0.0099*	0.2497	0.9076	0.0287
Ocean	Hem	*90S.00N*	*0.0324*	*<0.0001*	*0.2237*	0.0037	0.0002	0.0919
Ocean	Glob	*90S.90N*	*0.0693*	*0.0001*	*0.0083*	0.0027	0.0037	0.0230

Table A5.1.32: Results of the main findings, break years, shifts and diagnostics for GISSS-g observed zonal annual average temperatures from the MSBV. The bivariate test shows the year of change (the start of a new regime is the next year), the internal shift and change of trend at that time. The unit root tests show the statistical significance of each test under two conditions. The A column shows the test applied to the segment of data containing a change-point and, the B column shows the same test on the residuals after the implicit shift and trend-changes are removed. The year of exogenous for the ZA test is also shown. ANOVA tests are provided for the significance of a change of trend, and independently a change of intercept where the time is relative to the year of change (both should be considered in the context of the ANCOVA). The ANCOVA tests the two segment regression at the change-point against the single regression and is equivalent to a Chow test. The segment classification is as per Table A4.1.29 . For the unit-root tests, pink fill indicates a finding of non-stationarity, green indicates stationarity and no colour indicates insufficient the p-value is based on insufficient data. For the ANOVA and ANCOVA tests, green fill indicate tests with p-values that support acceptance of an H1 of a change of level, trend or persistent regime in each case. In the year of change column, red text indicates that the consideration of the ANCOVA together with non-zero trends casts means that the data is mis-specified for the MSBV and continued trends cannot be discounted. Green shading signals that only one of the pre and post trends is non-zero, while ANCOVA does not.

Zone	Bivariate Test			Unit root and stationarity tests									ANOVA		ANCOVA	Classifications		
	Year of Change	Internal Shift	Trend Change	KPSS-L		KPSS-T		ADF		Zivot-Andrews			Level	Trend	Regime	Change Type		Change-Index
	Year	^{0}C	^{0}C/Year	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Year	Pr	Pr	Pr	Code	Class	Ix
				A	B	A	B	A	B	A	B							
LandOcean.00N.30N	1935	0.26	-0.0008	0.010	0.100	0.025	0.100	0.010	0.010	0.010	0.010	1923	0.0001	0.7150	0.0003	25,26	Single, Stationary	4
LandOcean.00N.30N	1978	0.21	0.0087	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0157	0.2103	0.0003	25,26	Single, Stationary	4
LandOcean.00N.30N	1996	0.13	0.0028	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1986	0.1171	0.7195	0.2628	20,20	Single, Stationary	0
LandOcean.30S.00N	1902	-0.26	0.0025	0.085	0.100	0.050	0.100	0.010	0.010	0.010	0.010	1902	0.0002	0.5627	0.0002	26,26	Single, Stationary	6
LandOcean.30S.00N	1939	0.18	-0.0055	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1945	0.0129	0.0935	0.0111	26,26	Single, Stationary	5
LandOcean.30S.00N	1978	0.28	0.0083	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1945	0.0045	0.2784	0.0002	19,20	Single, Stationary	4
LandOcean.30S.00N	1996	0.09	-0.0021	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2001	0.2070	0.7630	0.4337	20,20	Single, Stationary	0
LandOcean.30N.60N	1920	0.28	0.0010	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0000	0.6840	0.0001	25,26	Single, Stationary	4
LandOcean.30N.60N	1987	0.37	-0.0066	0.015	0.100	0.015	0.100	0.010	0.010	0.010	0.010	1963	0.0054	0.7390	0.0001	25,26	Single, Stationary	4
LandOcean.30N.60N	1997	0.41	0.0094	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.010	1997	0.0029	0.6303	0.0090	11,2	Single, N/A	4
LandOcean.60S.30S	1896	-0.07	0.0095	0.010	0.100	0.035	0.100	1.000	1.000	1.000	1.000	1911	0.2201	0.0776	0.0094	13,14	Single, Non-stationary	5
LandOcean.60S.30S	1936	0.27	-0.0003	0.010	0.100	0.038	0.100	1.000	0.050	1.000	0.010	1936	0.0000	0.9018	0.0000	13,26	Multiple, Stationary	4
LandOcean.60S.30S	1968	0.15	0.0106	0.036	0.100	0.017	0.100	1.000	0.050	1.000	0.050	1944	0.0379	0.4025	0.0004	12,20	Single, Stationary	6
LandOcean.60S.30S	1976	0.12	-0.0040	0.010	0.100	0.041	0.100	1.000	0.100	0.100	0.050	1981	0.0057	0.6525	0.0002	12,23	Multiple, Stationary	7
LandOcean.60N.90N	1919	0.65	-0.0069	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1954	0.0000	0.1783	0.0000	25,26	Single, Stationary	5

LandOcean.60N.90N	1994	0.69	0.0212	0.010	0.100	0.010	0.031	0.010	0.010	0.010	0.010	1963	0.0154	0.6218	0.0000	25,25	Single, Stationary	4
LandOcean.60N.90N	2004	0.61	-0.0367	0.013	0.100	0.100	0.100	1.000	0.100	0.010	0.050	2004	0.0527	0.4751	0.1282	2,2	Single, N/A	0
LandOcean.90S.60S	1969	0.31	0.0072	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1965	0.0001	0.0048	0.0000	25,26	Single, Stationary	4
LandOcean.20S.20N	1939	0.23	-0.0002	0.010	0.100	0.048	0.100	0.010	0.010	0.010	0.010	1902	0.0046	0.9503	0.0106	25,26	Single, Stationary	4
LandOcean.20S.20N	1978	0.29	0.0114	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1950	0.0018	0.0078	0.0002	25,26	Single, Stationary	6
LandOcean.60S.60N	1939	0.27	-0.0011	0.010	0.100	0.019	0.100	0.010	0.010	0.010	0.010	1902	0.0000	0.6450	0.0001	25,26	Single, Stationary	4
LandOcean.60S.60N	1978	0.26	0.0095	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1945	0.0034	0.1688	0.0001	19,20	Single, Stationary	4
LandOcean.60S.60N	1996	0.11	0.0003	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.010	2001	0.1390	0.9650	0.3272	20,20	Single, Stationary	0
LandOcean.90S.20S	1899	-0.04	0.0061	0.010	0.100	0.049	0.100	1.000	1.000	1.000	1.000	1911	0.3063	0.0579	0.0239	13,14	Single, Non-stationary	5
LandOcean.90S.20S	1936	0.26	-0.0017	0.010	0.100	0.029	0.100	1.000	0.050	0.100	0.010	1936	0.0000	0.4249	0.0000	13,26	Multiple, Stationary	4
LandOcean.90S.20S	1968	0.23	0.0022	0.074	0.100	0.010	0.100	1.000	0.100	1.000	0.010	1944	0.0032	0.8744	0.0001	12,23	Multiple, Stationary	4
LandOcean.90S.20S	1976	0.15	0.0031	0.010	0.100	0.100	0.100	1.000	0.050	1.000	0.100	1987	0.0522	0.8270	0.0990	11,11	Single, Non-stationary	0
LandOcean.90S.20S	1995	0.05	0.0064	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1981	0.3718	0.1825	0.2618	20,20	Single, Stationary	2
LandOcean.20N.90N	1920	0.34	-0.0001	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1919	0.0000	0.9710	0.0000	25,26	Single, Stationary	4
LandOcean.20N.90N	1987	0.31	0.0046	0.018	0.100	0.010	0.071	0.050	0.010	0.010	0.010	1963	0.0111	0.8043	0.0000	25,26	Single, Stationary	4
LandOcean.20N.90N	1997	0.31	0.0094	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.010	1993	0.0078	0.5768	0.0248	11,2	Single, N/A	6
LandOcean.90S.90N	1936	0.24	-0.0012	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1936	0.0000	0.4400	0.0000	25,26	Single, Stationary	4
LandOcean.90S.90N	1978	0.21	0.0112	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.010	1976	0.0007	0.0199	0.0000	22,26	Single, Stationary	6
LandOcean.90S.90N	1996	0.11	0.0013	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1986	0.0535	0.8182	0.1436	20,20	Single, Stationary	3
LandOcean.00N.90N	1924	0.27	0.0008	0.010	0.100	0.011	0.100	0.010	0.010	0.010	0.010	1920	0.0000	0.6300	0.0000	25,26	Single, Stationary	4
LandOcean.00N.90N	1986	0.26	0.0035	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1963	0.0115	0.8184	0.0001	25,26	Single, Stationary	4
LandOcean.00N.90N	1996	0.22	0.0104	0.010	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1994	0.0158	0.4359	0.0501	2,2	Single, N/A	2
LandOcean.90S.00N	1901	-0.12	0.0060	0.010	0.100	0.014	0.100	0.050	0.010	0.050	0.010	1911	0.0067	0.0399	0.0002	19,20	Single, Stationary	4
LandOcean.90S.00N	1938	0.18	-0.0003	0.010	0.100	0.100	0.100	0.050	0.010	0.100	0.010	1936	0.0003	0.8823	0.0014	17,26	Single, Stationary	4
LandOcean.90S.00N	1976	0.24	0.0051	0.010	0.100	0.010	0.075	1.000	0.050	0.010	0.010	1945	0.0002	0.2861	0.0000	22,20	Single, Stationary	4
LandOcean.90S.00N	1995	0.04	0.0013	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1991	0.4177	0.7668	0.6779	20,20	Single, Stationary	2
Land.00N.30N	1923	0.22	0.0011	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1923	0.0009	0.6280	0.0038	26,26	Single, Stationary	4
Land.00N.30N	1978	0.21	0.0118	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0245	0.0951	0.0000	25,26	Single, Stationary	4

Land.00N.30N	1997	0.29	0.0014	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1997	0.0138	0.8994	0.0415	20,20	Single, Stationary	4
Land.30S.00N	1939	0.25	0.0041	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1902	0.0002	0.1227	0.0000	25,26	Single, Stationary	4
Land.30S.00N	1976	0.19	0.0149	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0409	0.0507	0.0004	19,20	Single, Stationary	6
Land.30S.00N	1994	0.04	-0.0039	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2010	0.6510	0.6680	0.8141	20,20	Single, Stationary	1
Land.30N.60N	1893	0.25	-0.0097	0.010	0.100	0.056	0.100	0.050	0.010	0.050	0.050	1902	0.0378	0.4580	0.0245	20,20	Single, Stationary	4
Land.30N.60N	1920	0.26	0.0033	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1963	0.0197	0.5990	0.0414	26,26	Single, Stationary	4
Land.30N.60N	1985	0.47	-0.0058	0.010	0.100	0.011	0.100	0.010	0.010	0.010	0.010	1963	0.0152	0.8242	0.0010	25,26	Single, Stationary	4
Land.30N.60N	1996	0.57	0.0031	0.010	0.100	0.100	0.100	0.050	0.010	0.050	0.050	1997	0.0039	0.9051	0.0103	20,20	Single, Stationary	4
Land.60S.30S	1931	0.18	0.0011	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1931	0.0069	0.6373	0.0180	26,26	Single, Stationary	4
Land.60S.30S	1976	0.28	0.0014	0.010	0.100	0.012	0.100	0.010	0.010	0.010	0.010	1976	0.0006	0.7587	0.0004	25,26	Single, Stationary	4
Land.60S.30S	2002	0.07	0.0247	0.010	0.100	0.030	0.100	0.100	0.010	0.010	0.010	1991	0.5513	0.0674	0.0281	18,20	Single, Stationary	6
Land.60N.90N	1899	0.49	-0.0127	0.010	0.100	0.087	0.100	0.050	0.010	0.010	0.010	1902	0.0083	0.4083	0.0236	20,20	Single, Stationary	4
Land.60N.90N	1919	0.79	0.0134	0.010	0.100	0.010	0.039	0.010	0.010	0.010	0.010	1948	0.0003	0.4337	0.0001	25,25	Single, Stationary	4
Land.60N.90N	1994	0.66	0.0380	0.010	0.100	0.010	0.044	0.050	0.010	0.010	0.010	1963	0.0053	0.0317	0.0000	25,25	Single, Stationary	6
Land.90S.60S	2000	0.37	0.0048	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1957	0.0376	0.8074	0.0004	25,26	Single, Stationary	4
Land.20S.20N	1900	-0.21	-0.0060	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1900	0.0033	0.2474	0.0096	20,20	Single, Stationary	4
Land.20S.20N	1925	0.23	0.0020	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1923	0.0024	0.6599	0.0045	26,26	Single, Stationary	4
Land.20S.20N	1978	0.27	0.0125	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0042	0.1023	0.0000	25,26	Single, Stationary	4
Land.20S.20N	1996	0.13	-0.0027	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1986	0.2660	0.8110	0.5241	20,20	Single, Stationary	0
Land.60S.60N	1925	0.19	0.0053	0.010	0.100	0.041	0.100	0.010	0.010	0.010	0.010	1902	0.0011	0.0099	0.0005	25,26	Single, Stationary	5
Land.60S.60N	1978	0.19	0.0116	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0208	0.0646	0.0000	25,26	Single, Stationary	6
Land.60S.60N	1997	0.18	-0.0014	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1997	0.0796	0.8817	0.2089	20,20	Single, Stationary	1
Land.90S.20S	1935	0.09	0.0015	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1971	0.0997	0.4394	0.1320	26,26	Single, Stationary	0
Land.90S.20S	1979	0.28	-0.0047	0.010	0.100	0.043	0.100	0.010	0.010	0.010	0.010	1979	0.0087	0.4973	0.0223	25,26	Single, Stationary	4
Land.90S.20S	2001	0.27	0.0135	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1992	0.0979	0.4615	0.0691	20,20	Single, Stationary	0
Land.20N.90N	1893	0.24	-0.0017	0.010	0.100	0.047	0.100	0.050	0.010	0.100	0.050	1906	0.0166	0.8769	0.0241	9,20	Multiple, Stationary	4
Land.20N.90N	1920	0.35	0.0010	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0005	0.8557	0.0003	25,26	Single, Stationary	4
Land.20N.90N	1987	0.47	-0.0007	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1963	0.0079	0.9795	0.0001	25,26	Single, Stationary	4

Land.20N.90N	1997	0.46	0.0081	0.010	0.100	0.100	0.100	0.100	0.010	0.100	0.010	1997	0.0051	0.7267	0.0148	11,2	Single, N/A	4
Land.90S.90N	1920	0.20	-0.0017	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1929	0.0005	0.4160	0.0004	25,26	Single, Stationary	4
Land.90S.90N	1979	0.20	0.0157	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1963	0.0195	0.0266	0.0000	25,26	Single, Stationary	4
Land.90S.90N	1997	0.24	-0.0039	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1997	0.0306	0.7087	0.0920	20,20	Single, Stationary	2
Land.00N.90N	1895	0.23	-0.0082	0.016	0.100	0.045	0.100	0.050	0.010	1.000	0.050	1896	0.0096	0.3129	0.0070	9,20	Multiple, Stationary	4
Land.00N.90N	1920	0.33	0.0058	0.010	0.100	0.012	0.100	0.010	0.010	0.010	0.010	1920	0.0002	0.2662	0.0003	25,26	Single, Stationary	4
Land.00N.90N	1986	0.36	0.0096	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1963	0.0109	0.6139	0.0000	25,26	Single, Stationary	4
Land.00N.90N	1997	0.37	-0.0010	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.010	1997	0.0064	0.9552	0.0163	2,2	Single, N/A	4
Land.90S.00N	1939	0.14	0.0040	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1902	0.0049	0.0347	0.0001	25,26	Single, Stationary	4
Land.90S.00N	1979	0.25	0.0024	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1979	0.0028	0.6645	0.0015	25,26	Single, Stationary	4
Land.90S.00N	2001	0.20	0.0015	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2001	0.0770	0.9070	0.1231	20,20	Single, Stationary	0
Ocean.00N.30N	1935	0.33	0.0004	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1924	0.0000	0.8850	0.0000	25,26	Single, Stationary	4
Ocean.00N.30N	1978	0.20	0.0076	0.010	0.100	0.016	0.100	0.010	0.010	0.050	0.050	1945	0.0150	0.1660	0.0008	25,26	Single, Stationary	4
Ocean.00N.30N	2000	0.12	-0.0040	0.010	0.100	0.100	0.100	0.010	0.010	0.100	0.100	1985	0.1450	0.6350	0.3369	11,11	Single, Non-stationary	0
Ocean.30S.00N	1939	0.29	0.0001	0.013	0.100	0.044	0.100	0.010	0.010	0.010	0.010	1938	0.0001	0.9660	0.0001	25,26	Single, Stationary	4
Ocean.30S.00N	1978	0.30	0.0069	0.010	0.100	0.010	0.072	0.010	0.010	0.010	0.010	1945	0.0036	0.4011	0.0003	19,20	Single, Stationary	4
Ocean.30S.00N	1996	0.11	-0.0015	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2001	0.1220	0.8230	0.2934	20,20	Single, Stationary	0
Ocean.30N.60N	1901	-0.48	0.0043	0.010	0.100	0.028	0.100	1.000	0.010	1.000	0.050	1901	0.0000	0.6830	0.0000	9,20	Multiple, Stationary	4
Ocean.30N.60N	1914	0.30	-0.0001	0.010	0.100	0.099	0.100	0.100	0.010	0.050	0.010	1914	0.0001	0.9909	0.0004	20,20	Single, Stationary	4
Ocean.30N.60N	1931	0.16	-0.0048	0.090	0.100	0.010	0.048	0.010	0.010	0.010	0.010	1963	0.0390	0.5190	0.0026	25,25	Single, Stationary	4
Ocean.30N.60N	1997	0.34	0.0166	0.010	0.100	0.010	0.054	1.000	0.010	0.010	0.010	1967	0.0001	0.0273	0.0000	22,26	Single, Stationary	6
Ocean.60S.30S	1896	-0.07	0.0102	0.010	0.100	0.039	0.100	1.000	1.000	1.000	1.000	1911	0.2582	0.0696	0.0104	13,14	Single, Non-stationary	5
Ocean.60S.30S	1936	0.28	0.0000	0.010	0.100	0.035	0.100	1.000	0.050	1.000	0.010	1936	0.0000	0.9957	0.0000	13,26	Multiple, Stationary	4
Ocean.60S.30S	1968	0.15	0.0129	0.035	0.100	0.017	0.100	1.000	0.050	1.000	0.050	1944	0.0396	0.3165	0.0002	12,20	Single, Stationary	6
Ocean.60S.30S	1976	0.11	-0.0066	0.010	0.100	0.035	0.100	0.100	0.050	0.100	0.050	1981	0.0094	0.4486	0.0002	12,20	Single, Stationary	7
Ocean.60N.90N	1919	0.68	-0.0013	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1919	0.0000	0.7694	0.0000	25,26	Single, Stationary	4
Ocean.60N.90N	1994	0.63	0.0347	0.027	0.100	0.010	0.047	0.050	0.010	0.010	0.010	1962	0.0108	0.3540	0.0000	25,25	Single, Stationary	4
Ocean.60N.90N	2004	0.65	-0.0447	0.010	0.100	0.100	0.100	1.000	0.050	0.050	0.100	2004	0.0301	0.3569	0.0722	2,11	Single, Non-stationary	0

Ocean.90S.60S	1969	0.41	0.0055	0.010	0.100	0.010	0.042	0.010	0.010	0.010	0.010	1935	0.0000	0.0141	0.0000	25,25	Single, Stationary	6
Ocean.20S.20N	1939	0.27	-0.0007	0.010	0.100	0.083	0.100	0.010	0.010	0.010	0.010	1902	0.0038	0.8418	0.0100	26,26	Single, Stationary	4
Ocean.20S.20N	1978	0.31	0.0096	0.010	0.100	0.022	0.100	0.010	0.010	0.010	0.010	1956	0.0025	0.0382	0.0008	25,26	Single, Stationary	6
Ocean.60S.60N	1939	0.32	-0.0015	0.010	0.100	0.019	0.100	0.010	0.010	0.010	0.010	1938	0.0000	0.5650	0.0000	25,26	Single, Stationary	4
Ocean.60S.60N	1978	0.27	0.0084	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1945	0.0031	0.2406	0.0001	19,20	Single, Stationary	4
Ocean.60S.60N	1996	0.09	-0.0003	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	2001	0.2100	0.9640	0.4503	20,20	Single, Stationary	0
Ocean.90S.20S	1899	-0.05	0.0069	0.010	0.100	0.060	0.100	1.000	1.000	1.000	1.000	1911	0.3159	0.0584	0.0251	14,14	Single, Non-stationary	5
Ocean.90S.20S	1936	0.29	-0.0010	0.010	0.100	0.030	0.100	1.000	0.050	1.000	0.100	1936	0.0000	0.6520	0.0000	13,17	Non-stationary	4
Ocean.90S.20S	1968	0.21	0.0106	0.070	0.100	0.010	0.100	1.000	0.050	1.000	0.050	1944	0.0081	0.4357	0.0000	12,20	Single, Stationary	4
Ocean.90S.20S	1976	0.12	-0.0054	0.010	0.100	0.039	0.100	1.000	0.100	1.000	0.100	1987	0.0257	0.5934	0.0162	9,11	Non-stationary	4
Ocean.90S.20S	1995	0.05	0.0035	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1981	0.1536	0.3119	0.2045	20,20	Single, Stationary	2
Ocean.20N.90N	1901	-0.27	0.0014	0.010	0.100	0.070	0.100	0.050	0.010	0.100	0.050	1901	0.0001	0.8310	0.0001	11,20	Multiple, Stationary	4
Ocean.20N.90N	1914	0.15	0.0147	0.010	0.100	0.100	0.100	0.050	0.010	1.000	1.000	1906	0.0053	0.0256	0.0037	11,11	Single, Non-stationary	6
Ocean.20N.90N	1929	0.15	-0.0151	0.025	0.100	0.010	0.061	0.050	0.010	0.010	0.010	1962	0.0350	0.0575	0.0000	25,26	Single, Stationary	5
Ocean.20N.90N	1994	0.26	0.0278	0.019	0.100	0.010	0.019	1.000	0.010	0.050	0.010	1962	0.0196	0.1885	0.0000	22,25	Single, Stationary	6
Ocean.20N.90N	2002	0.06	-0.0118	0.010	0.100	0.100	0.100	1.000	0.100	0.010	0.050	2007	0.3210	0.3490	0.2984	2,2	Single, N/A	3
Ocean.90S.90N	1902	-0.14	-0.0011	0.010	0.100	0.100	0.100	0.050	0.010	1.000	1.000	1894	0.0279	0.8865	0.0236	11,11	Single, Non-stationary	4
Ocean.90S.90N	1913	0.15	0.0067	0.010	0.100	0.100	0.100	0.050	0.010	0.100	1.000	1911	0.0044	0.3324	0.0147	11,11	Single, Non-stationary	4
Ocean.90S.90N	1936	0.23	-0.0026	0.010	0.100	0.010	0.100	0.050	0.010	1.000	0.010	1945	0.0000	0.3980	0.0000	16,26	Multiple, Stationary	4
Ocean.90S.90N	1976	0.19	0.0097	0.010	0.100	0.010	0.100	0.100	0.010	0.010	0.010	1945	0.0005	0.0118	0.0000	22,26	Single, Stationary	6
Ocean.90S.90N	1996	0.06	0.0011	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.050	1981	0.1420	0.7629	0.2961	20,20	Single, Stationary	3
Ocean.00N.90N	1901	-0.28	0.0090	0.010	0.100	0.019	0.100	0.100	0.010	0.100	0.050	1906	0.0000	0.0638	0.0000	12,20	Single, Stationary	6
Ocean.00N.90N	1925	0.22	-0.0077	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.050	1945	0.0002	0.0387	0.0000	25,26	Single, Stationary	5
Ocean.00N.90N	1986	0.13	0.0126	0.010	0.100	0.010	0.097	0.050	0.010	0.050	0.010	1969	0.0806	0.1235	0.0000	25,26	Single, Stationary	4
Ocean.00N.90N	2000	0.11	-0.0018	0.010	0.100	0.100	0.100	0.050	0.010	1.000	1.000	1991	0.1293	0.8349	0.3072	11,11	Single, Non-stationary	2
Ocean.90S.00N	1901	-0.10	0.0068	0.010	0.100	0.018	0.100	0.050	0.010	0.050	0.050	1911	0.0219	0.0291	0.0006	19,20	Single, Stationary	5
Ocean.90S.00N	1938	0.20	-0.0001	0.010	0.100	0.100	0.100	0.050	0.010	1.000	0.010	1936	0.0001	0.9716	0.0005	17,26	Single, Stationary	4
Ocean.90S.00N	1976	0.25	0.0041	0.010	0.100	0.010	0.044	1.000	0.050	0.010	0.010	1945	0.0001	0.3870	0.0000	22,19	Single, Stationary	4

Ocean.90S.00N	1995	0.05	-0.0001	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1980	0.2225	0.9823	0.4697	20,20	Single, Stationary	2

Table A5.1.33: MSBV and diagnostics based on sector analysis (See Figure Ch5.28). Columns denoted as A are of segments of data containing a change-point. Those denoted B are of the residual of the segments after internal trend and shifts are removed. Throughout red or green highlights indicate probabilities: red is Pr <=0.01, green is 0.01<Pr<=0.05.

Zone	Sector	Bivariate Test			Unit root and stationarity tests									ANOVA		ANCOVA	Classifications		
		Year of Change	Internal Shift	Trend Change	KPSS-L		KPSS-T		ADF		Zivot-Andrews			Level	Trend	Regime	Change Type		
		Year	⁰C	⁰C/Year	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Year	Pr	Pr	Pr	Code	Change-Class	Change-Index
					A	B	A	B	A	B	A	B							
LandOcean.00N.30N	105E.150E	1940	0.35	-0.003	0.045	0.100	0.022	0.100	0.010	0.010	0.010	0.010	1940	0.0000	0.3961	0.0000	25,26	Single,Stat	7
LandOcean.00N.30N	105E.150E	1977	0.21	0.013	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1946	0.0097	0.0334	0.0001	19,20	Single,Stat	5
LandOcean.00N.30N	105E.150E	1997	0.35	-0.011	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.050	1997	0.0002	0.1695	0.0007	20,20	Single,Stat	4
LandOcean.30S.00N	105E.150E	1956	0.21	-0.001	0.010	0.100	0.039	0.100	0.010	0.010	0.010	0.010	1901	0.0625	0.9501	0.0331	25,26	Single,Stat	4
LandOcean.30S.00N	105E.150E	1977	0.24	0.005	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.010	1968	0.0340	0.5925	0.0812	20,20	Single,Stat	0
LandOcean.30S.00N	105E.150E	1997	0.08	0.008	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2010	0.5470	0.5100	0.6113	20,20	Single,Stat	0
LandOcean.30N.60N	105E.150E	1937	0.16	-0.001	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1913	0.2190	0.7990	0.4659	26,26	Single,Stat	0
LandOcean.30N.60N	105E.150E	1987	0.70	0.009	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1987	0.0001	0.3240	0.0000	25,26	Single,Stat	4
LandOcean.60S.30S	105E.150E	1899	-0.26	-0.006	0.025	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1923	0.0126	0.4899	0.0244	26,26	Single,Stat	4
LandOcean.60S.30S	105E.150E	1956	0.21	0.006	0.014	0.100	0.038	0.100	0.050	0.010	0.010	0.010	1923	0.0672	0.4552	0.0062	25,26	Single,Stat	4
LandOcean.60S.30S	105E.150E	1975	0.24	-0.005	0.010	0.100	0.024	0.100	0.050	0.010	0.010	0.010	1976	0.0058	0.4828	0.0054	19,20	Single,Stat	4
LandOcean.60S.30S	105E.150E	2006	0.05	0.051	0.010	0.100	0.010	0.100	1.000	0.100	0.010	0.010	2005	0.6233	0.0103	0.0002	21,23	Single,Stat	6
LandOcean.60N.90N	105E.150E	1987	0.36	0.039	0.010	0.100	0.010	0.092	0.010	0.010	0.010	0.010	1949	0.2449	0.0255	0.0000	25,26	Single,Stat	7
LandOcean.90S.60S	105E.150E	1970	0.43	0.000	0.010	0.100	0.015	0.100	0.010	0.010	0.010	0.010	1970	0.0003	0.9525	0.0001	25,26	Single,Stat	5
LandOcean.00N.30N	150E.165W	1939	0.36	0.004	0.010	0.100	0.019	0.100	0.010	0.010	0.010	0.010	1920	0.0000	0.0540	0.0000	25,26	Single,Stat	4
LandOcean.00N.30N	150E.165W	1999	0.30	-0.001	0.010	0.100	0.010	0.049	0.050	0.010	0.010	0.010	1970	0.0025	0.9206	0.0001	25,25	Single,Stat	4
LandOcean.30S.00N	150E.165W	1941	0.47	-0.008	0.036	0.100	0.010	0.091	0.010	0.010	0.010	0.010	1939	0.0000	0.0794	0.0000	25,26	Single,Stat	7
LandOcean.30S.00N	150E.165W	1968	0.25	0.014	0.021	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1946	0.0004	0.0035	0.0000	19,20	Single,Stat	5
LandOcean.30S.00N	150E.165W	1994	0.21	0.008	0.010	0.100	0.025	0.100	0.010	0.010	0.010	0.010	1994	0.0044	0.1282	0.0017	18,20	Single,Stat	6
LandOcean.30N.60N	150E.165W	1896	-0.65	-0.020	0.014	0.100	0.026	0.100	0.050	0.010	0.050	0.010	1894	0.0001	0.1778	0.0003	25,26	Single,Stat	4
LandOcean.30N.60N	150E.165W	1941	0.35	-0.005	0.010	0.100	0.045	0.100	0.010	0.010	0.010	0.010	1941	0.0009	0.1827	0.0006	25,26	Single,Stat	4

LandOcean.30N.60N	150E.165W	1998	0.32	0.010	0.100	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1988	0.1070	0.7670	0.0073	25,26	Single,Stat	4
LandOcean.30N.60N	150E.165W	2007	0.23	-0.001	0.010	0.100	0.100	0.100	0.050	0.050	0.010	0.050	2010	0.0816	0.9723	0.1796	0,0	Single, N/A	0
LandOcean.60S.30S	150E.165W	1953	0.31	0.004	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1896	0.0015	0.2189	0.0002	25,26	Single,Stat	4
LandOcean.60S.30S	150E.165W	1997	0.23	-0.003	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.010	1967	0.0611	0.7860	0.0853	26,26	Single,Stat	0
LandOcean.60N.90N	150E.165W	1920	0.70	0.017	0.011	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1920	0.0009	0.0265	0.0028	26,26	Single,Stat	5
LandOcean.60N.90N	150E.165W	1994	0.60	0.046	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0504	0.0488	0.0000	25,26	Single,Stat	4
LandOcean.90S.60S	150E.165W	1969	0.35	0.008	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0011	0.0303	0.0000	25,26	Single,Stat	5
LandOcean.00N.30N	165W.120W	1935	0.23	0.003	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1976	0.0533	0.3851	0.1443	26,26	Single,Stat	2
LandOcean.30S.00N	165W.120W	1981	0.30	0.003	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1906	0.0065	0.4970	0.0001	25,26	Single,Stat	4
LandOcean.30N.60N	165W.120W	1933	0.38	0.008	0.010	0.100	0.049	0.100	0.010	0.010	0.010	0.010	1916	0.0010	0.0157	0.0017	25,26	Single,Stat	6
LandOcean.60S.30S	165W.120W	1989	0.28	0.006	0.010	0.100	0.010	0.081	0.010	0.010	0.010	0.010	1906	0.0066	0.3241	0.0000	25,26	Single,Stat	4
LandOcean.60N.90N	165W.120W	1910	1.12	0.016	0.010	0.100	0.028	0.100	0.010	0.010	0.010	0.010	1901	0.0016	0.3363	0.0040	25,26	Single,Stat	4
LandOcean.60N.90N	165W.120W	1986	0.65	0.031	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0653	0.0999	0.0003	25,26	Single,Stat	6
LandOcean.90S.60S	165W.120W	1986	0.49	-0.001	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0000	0.9170	0.0000	25,26	Single,Stat	4
LandOcean.00N.30N	120W.75W	1981	0.21	0.006	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1905	0.0384	0.2233	0.0003	25,26	Single,Stat	4
LandOcean.30S.00N	120W.75W	1976	0.62	-0.002	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1975	0.0001	0.7225	0.0000	25,26	Single,Stat	4
LandOcean.30N.60N	120W.75W	1920	0.27	-0.012	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1946	0.1305	0.0714	0.0074	25,26	Single,Stat	5
LandOcean.30N.60N	120W.75W	1997	1.07	-0.038	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1997	0.0002	0.1200	0.0001	25,26	Single,Stat	4
LandOcean.60S.30S	120W.75W	1890	-0.34	0.028	0.036	0.100	0.010	0.010	0.050	0.050	0.010	0.010	1937	0.0175	0.2141	0.0000	25,25	Single,Stat	4
LandOcean.60S.30S	120W.75W	1978	0.43	-0.005	0.010	0.100	0.010	0.016	0.010	0.010	0.010	0.010	1937	0.0000	0.1810	0.0000	25,25	Single,Stat	4
LandOcean.60N.90N	120W.75W	1922	0.58	-0.010	0.010	0.100	0.013	0.100	0.010	0.010	0.010	0.010	1954	0.0276	0.2665	0.0101	25,26	Single,Stat	4
LandOcean.60N.90N	120W.75W	1994	0.98	0.035	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1971	0.0166	0.2564	0.0000	25,26	Single,Stat	4
LandOcean.90S.60S	120W.75W	1905	-0.04	0.008	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1957	0.5625	0.0829	0.0379	26,26	Single,Stat	5
LandOcean.90S.60S	120W.75W	1969	0.53	-0.003	0.037	0.100	0.012	0.100	0.010	0.010	0.010	0.010	1964	0.0003	0.7751	0.0000	25,26	Single,Stat	4
LandOcean.90S.60S	120W.75W	1987	0.78	-0.002	0.010	0.100	0.053	0.100	0.050	0.010	0.010	0.010	1987	0.0007	0.9322	0.0017	20,20	Single,Stat	4
LandOcean.00N.30N	75W.30W	1903	-0.27	-0.002	0.011	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1903	0.0173	0.8320	0.0512	20,20	Single,Stat	0
LandOcean.00N.30N	75W.30W	1925	0.42	-0.001	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1925	0.0000	0.9360	0.0000	25,26	Single,Stat	4
LandOcean.00N.30N	75W.30W	1977	0.09	0.008	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1970	0.3540	0.1540	0.0509	26,26	Single,Stat	0

LandOcean.00N.30N	75W.30W	2002	0.29	-0.007	0.010	0.100	0.082	0.100	0.050	0.010	0.010	0.010	1986	0.0257	0.6231	0.0524	20,20	Single,Stat	0
LandOcean.30S.00N	75W.30W	1902	-0.26	0.004	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1915	0.0030	0.5054	0.0032	19,20	Single,Stat	4
LandOcean.30S.00N	75W.30W	1938	0.29	0.003	0.010	0.100	0.075	0.100	0.010	0.010	0.010	0.010	1925	0.0016	0.4757	0.0056	26,26	Single,Stat	4
LandOcean.30S.00N	75W.30W	1976	0.19	0.012	0.010	0.100	0.022	0.100	0.010	0.010	0.010	0.010	1976	0.0776	0.2099	0.0059	19,20	Single,Stat	6
LandOcean.30S.00N	75W.30W	1993	0.10	-0.006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2006	0.2865	0.5195	0.4125	20,20	Single,Stat	1
LandOcean.30N.60N	75W.30W	1926	0.52	-0.001	0.010	0.100	0.030	0.100	0.010	0.010	0.010	0.010	1926	0.0000	0.8329	0.0000	25,26	Single,Stat	4
LandOcean.30N.60N	75W.30W	1961	-0.42	0.008	0.010	0.100	0.028	0.100	0.050	0.010	0.010	0.010	1961	0.0001	0.1361	0.0003	25,26	Single,Stat	4
LandOcean.30N.60N	75W.30W	1997	0.61	-0.001	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1997	0.0000	0.9370	0.0000	19,20	Single,Stat	4
LandOcean.60S.30S	75W.30W	1939	0.32	0.004	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1939	0.0000	0.1927	0.0000	25,26	Single,Stat	4
LandOcean.60S.30S	75W.30W	1976	0.26	0.002	0.010	0.100	0.045	0.100	0.010	0.010	0.010	0.010	1945	0.0007	0.6361	0.0026	25,26	Single,Stat	6
LandOcean.60N.90N	75W.30W	1922	1.55	-0.026	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1922	0.0000	0.0058	0.0000	25,26	Single,Stat	6
LandOcean.60N.90N	75W.30W	1997	1.24	0.046	0.027	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1994	0.0004	0.1364	0.0000	25,26	Single,Stat	5
LandOcean.90S.60S	75W.30W	1980	0.68	0.003	0.010	0.100	0.010	0.023	0.010	0.010	0.010	0.010	1926	0.0009	0.7271	0.0000	25,25	Single,Stat	4
LandOcean.00N.30N	30W.15E	1978	0.34	-0.003	0.010	0.100	0.027	0.100	0.010	0.010	0.010	0.010	1986	0.0093	0.8138	0.0004	25,26	Single,Stat	4
LandOcean.00N.30N	30W.15E	1994	0.33	0.014	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1994	0.0212	0.3240	0.0560	20,20	Single,Stat	0
LandOcean.30S.00N	30W.15E	1919	0.33	0.006	0.010	0.100	0.062	0.100	0.010	0.010	0.010	0.010	1911	0.0017	0.1474	0.0045	26,26	Single,Stat	4
LandOcean.30S.00N	30W.15E	1967	0.27	-0.011	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1945	0.0628	0.4300	0.1346	26,26	Single,Stat	0
LandOcean.30S.00N	30W.15E	1982	0.31	0.011	0.010	0.100	0.100	0.100	0.010	0.010	0.100	0.010	1982	0.0188	0.4149	0.0561	11,20	Multiple,Stat	0
LandOcean.30N.60N	30W.15E	1925	0.30	0.003	0.010	0.100	0.048	0.100	0.010	0.010	0.010	0.010	1900	0.0026	0.4037	0.0034	25,26	Single,Stat	4
LandOcean.30N.60N	30W.15E	1961	-0.46	0.007	0.017	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1961	0.0000	0.1096	0.0000	25,26	Single,Stat	6
LandOcean.30N.60N	30W.15E	1994	0.49	-0.010	0.010	0.100	0.029	0.100	1.000	0.050	0.010	0.010	1994	0.0000	0.1970	0.0001	22,20	Single,Stat	5
LandOcean.60S.30S	30W.15E	1968	0.53	0.002	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1967	0.0000	0.4359	0.0000	25,26	Single,Stat	4
LandOcean.60N.90N	30W.15E	1921	0.62	-0.010	0.010	0.100	0.010	0.072	0.010	0.010	0.010	0.010	1960	0.0000	0.0438	0.0000	25,26	Single,Stat	4
LandOcean.60N.90N	30W.15E	2001	1.08	0.015	0.010	0.100	0.010	0.014	1.000	0.010	0.010	0.010	1976	0.0000	0.5600	0.0000	22,25	Single,Stat	4
LandOcean.90S.60S	30W.15E	1973	0.65	-0.014	0.017	0.100	0.046	0.100	0.010	0.010	0.010	0.010	1973	0.0000	0.0024	0.0000	25,26	Single,Stat	6
LandOcean.00N.30N	15E.60E	1902	-0.26	0.001	0.024	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1913	0.0106	0.8472	0.0363	20,20	Single,Stat	4
LandOcean.00N.30N	15E.60E	1923	0.30	-0.005	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1942	0.0035	0.5239	0.0003	25,26	Single,Stat	4
LandOcean.00N.30N	15E.60E	1978	0.20	0.009	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1942	0.0649	0.2790	0.0022	25,26	Single,Stat	4

LandOcean.00N.30N	15E.60E	1997	0.33	0.010	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.010	1994	0.0168	0.4237	0.0309	20,20	Single,Stat	4
LandOcean.30S.00N	15E.60E	1976	0.25	0.006	0.010	0.100	0.010	0.096	0.010	0.010	0.010	0.010	1903	0.0088	0.3657	0.0000	25,26	Single,Stat	4
LandOcean.30S.00N	15E.60E	1997	0.10	0.005	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2002	0.3130	0.5860	0.4388	20,20	Single,Stat	0
LandOcean.30N.60N	15E.60E	1933	0.10	0.001	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1988	0.4847	0.7868	0.7699	26,26	Single,Stat	0
LandOcean.30N.60N	15E.60E	1997	0.73	0.015	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1997	0.0019	0.4524	0.0000	25,26	Single,Stat	4
LandOcean.60S.30S	15E.60E	1929	0.31	-0.010	0.046	0.100	0.030	0.045	1.000	1.000	0.100	0.050	1905	0.1530	0.8310	0.0773	13,22	Multiple,Stat	0
LandOcean.60S.30S	15E.60E	1936	0.30	0.008	0.100	0.100	0.066	0.100	0.050	0.010	0.010	0.010	1944	0.0509	0.8399	0.0342	20,20	Single,Stat	4
LandOcean.60S.30S	15E.60E	1976	0.23	0.009	0.052	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1963	0.1370	0.7540	0.0196	18,20	Single,Stat	4
LandOcean.60S.30S	15E.60E	1984	0.17	-0.007	0.024	0.100	0.068	0.100	0.050	0.010	1.000	1.000	2001	0.0804	0.7210	0.0344	11,11	Single, Non-stationary	4
LandOcean.60N.90N	15E.60E	1919	0.68	-0.003	0.010	0.100	0.053	0.100	0.010	0.010	0.010	0.010	1962	0.0104	0.7318	0.0076	26,26	Single,Stat	4
LandOcean.60N.90N	15E.60E	1999	0.88	0.038	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1962	0.0307	0.3613	0.0000	25,26	Single,Stat	4
LandOcean.90S.60S	15E.60E	1935	0.14	0.004	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2001	0.2290	0.2360	0.3258	26,26	Single,Stat	0
LandOcean.00N.30N	60E.105E	1935	0.18	-0.001	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1902	0.0141	0.6701	0.0468	26,26	Single,Stat	4
LandOcean.00N.30N	60E.105E	1976	0.22	0.005	0.010	0.100	0.022	0.100	0.010	0.010	0.010	0.010	1946	0.0208	0.4544	0.0066	25,26	Single,Stat	4
LandOcean.00N.30N	60E.105E	1997	0.22	-0.001	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1997	0.0366	0.9344	0.1010	20,20	Single,Stat	0
LandOcean.30S.00N	60E.105E	1976	0.28	0.006	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1945	0.0036	0.3583	0.0000	25,26	Single,Stat	4
LandOcean.30S.00N	60E.105E	2000	0.10	0.012	0.010	0.100	0.086	0.100	0.010	0.010	0.010	0.010	1991	0.2466	0.1995	0.0710	20,20	Single,Stat	2
LandOcean.30N.60N	60E.105E	1912	0.28	-0.006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1926	0.0753	0.3877	0.0348	26,26	Single,Stat	4
LandOcean.30N.60N	60E.105E	1976	0.40	0.016	0.010	0.100	0.019	0.100	0.010	0.010	0.010	0.010	1976	0.0633	0.3043	0.0015	25,26	Single,Stat	4
LandOcean.30N.60N	60E.105E	1996	0.71	-0.031	0.010	0.100	0.100	0.100	0.100	0.010	0.010	0.010	1996	0.0067	0.1849	0.0165	20,20	Single,Stat	4
LandOcean.60S.30S	60E.105E	1897	-0.24	-0.001	0.010	0.100	0.100	0.100	0.100	0.050	0.010	0.010	1913	0.0064	0.9090	0.0175	23,20	Single,Stat	4
LandOcean.60S.30S	60E.105E	1925	-0.39	0.018	0.022	0.100	0.015	0.100	1.000	0.010	0.010	0.010	1913	0.0033	0.4736	0.0010	18,20	Single,Stat	4
LandOcean.60S.30S	60E.105E	1932	0.49	-0.021	0.033	0.100	0.010	0.100	1.000	0.100	0.050	0.010	1941	0.0005	0.5351	0.0000	22,23	Single,Stat	4
LandOcean.60S.30S	60E.105E	1975	0.42	0.008	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.010	1974	0.0000	0.0059	0.0000	22,26	Single,Stat	6
LandOcean.60N.90N	60E.105E	1919	0.80	-0.013	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1919	0.0097	0.2628	0.0027	25,26	Single,Stat	4
LandOcean.60N.90N	60E.105E	1987	0.65	0.041	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0874	0.0490	0.0003	25,26	Single,Stat	4
LandOcean.90S.60S	60E.105E	1999	0.37	-0.006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1991	0.1190	0.8039	0.0587	26,26	Single,Stat	1

Land.00N.30N	105E.150E	1936	0.25	-0.006	0.010	0.100	0.079	0.100	0.010	0.010	0.010	0.010	1936	0.0020	0.0224	0.0015	26,26	Single,Stat	5
Land.00N.30N	105E.150E	1986	0.37	-0.003	0.033	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0099	0.8776	0.0008	25,26	Single,Stat	4
Land.00N.30N	105E.150E	1997	0.41	0.001	0.013	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1997	0.0104	0.9493	0.0282	2,2	Single, N/A	4
Land.30S.00N	105E.150E	1909	0.29	0.008	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1909	0.0137	0.1882	0.0416	26,26	Single,Stat	4
Land.30S.00N	105E.150E	1956	0.21	0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1916	0.0726	0.7682	0.0617	26,26	Single,Stat	0
Land.30S.00N	105E.150E	1978	0.32	-0.001	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.010	1997	0.0247	0.9415	0.0777	20,20	Single,Stat	0
Land.30S.00N	105E.150E	2001	0.15	0.016	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2010	0.3840	0.4130	0.2588	20,20	Single,Stat	0
Land.30N.60N	105E.150E	1918	0.24	-0.000	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1951	0.1250	0.9710	0.2328	26,26	Single,Stat	0
Land.30N.60N	105E.150E	1987	0.80	0.007	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1987	0.0000	0.4830	0.0000	25,26	Single,Stat	4
Land.60S.30S	105E.150E	1971	0.35	0.007	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1942	0.0027	0.2100	0.0000	25,26	Single,Stat	4
Land.60S.30S	105E.150E	2004	0.15	0.041	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.010	1983	0.4190	0.1404	0.0128	21,20	Single,Stat	4
Land.60N.90N	105E.150E	1987	0.52	0.027	0.010	0.100	0.010	0.052	0.010	0.010	0.010	0.010	1949	0.1223	0.1617	0.0004	25,26	Single,Stat	5
Land.90S.60S	105E.150E	2006	0.78	-0.062	0.033	0.100	0.079	0.100	0.010	0.010	0.010	0.010	1992	0.0068	0.2616	0.0014	26,26	Single,Stat	4
Land.30S.00N	150E.165W	1972	0.36	-0.001	0.010	0.100	0.010	0.053	0.010	0.010	0.010	0.010	1901	0.0006	0.9149	0.0000	25,26	Single,Stat	4
Land.30S.00N	150E.165W	1994	0.33	0.010	0.010	0.100	0.100	0.100	0.050	0.010	0.010	0.010	1994	0.0020	0.2481	0.0030	20,20	Single,Stat	4
Land.30N.60N	150E.165W	1988	0.40	0.022	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1916	0.0802	0.0938	0.0000	25,26	Single,Stat	4
Land.60S.30S	150E.165W	1969	0.44	-0.008	0.010	0.100	0.039	0.100	0.010	0.010	0.010	0.010	1969	0.0007	0.2673	0.0007	25,26	Single,Stat	4
Land.60S.30S	150E.165W	1997	0.43	0.001	0.036	0.100	0.100	0.100	0.050	0.010	0.050	0.010	1997	0.0069	0.9129	0.0104	20,20	Single,Stat	4
Land.60N.90N	150E.165W	1920	0.82	0.021	0.014	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1920	0.0011	0.0251	0.0036	26,26	Single,Stat	5
Land.60N.90N	150E.165W	1999	0.86	0.036	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0220	0.3429	0.0000	25,26	Single,Stat	4
Land.90S.60S	150E.165W	2006	0.96	0.002	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	2006	0.0020	0.9677	0.0000	25,26	Single,Stat	4
Land.00N.30N	165W.120W	1933	0.50	-0.016	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1939	0.0014	0.0744	0.0057	26,26	Single,Stat	4
Land.00N.30N	165W.120W	1958	0.37	0.018	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1956	0.0048	0.0262	0.0132	26,26	Single,Stat	4
Land.30N.60N	165W.120W	1976	0.31	0.005	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1947	0.1550	0.5280	0.0606	26,26	Single,Stat	1
Land.60N.90N	165W.120W	1910	1.24	-0.005	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1944	0.0018	0.7890	0.0009	25,26	Single,Stat	4
Land.60N.90N	165W.120W	1975	0.88	0.027	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1944	0.0110	0.0408	0.0002	25,26	Single,Stat	4
Land.90S.60S	165W.120W	1987	0.68	0.012	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1987	0.0022	0.3371	0.0000	25,26	Single,Stat	4
Land.00N.30N	120W.75W	1989	0.31	0.010	0.010	0.100	0.010	0.075	0.010	0.010	0.010	0.010	1965	0.0007	0.0723	0.0000	25,26	Single,Stat	6

Land.30S.00N	120W.75W	1978	0.64	0.006	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1905	0.0011	0.4555	0.0000	25,26	Single,Stat	4
Land.30N.60N	120W.75W	1920	0.28	-0.012	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1946	0.1228	0.0683	0.0064	25,26	Single,Stat	5
Land.30N.60N	120W.75W	1997	1.06	-0.039	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1997	0.0002	0.1150	0.0002	25,26	Single,Stat	4
Land.60N.90N	120W.75W	1922	0.65	-0.015	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1954	0.0260	0.1530	0.0048	25,26	Single,Stat	4
Land.60N.90N	120W.75W	1993	0.93	0.034	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1971	0.0286	0.2547	0.0001	25,26	Single,Stat	4
Land.90S.60S	120W.75W	1987	1.03	-0.004	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1987	0.0000	0.7330	0.0000	25,26	Single,Stat	4
Land.00N.30N	75W.30W	1925	0.19	-0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1884	0.0996	0.6480	0.1812	26,26	Single,Stat	0
Land.00N.30N	75W.30W	1986	0.23	0.015	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1976	0.0327	0.0112	0.0000	25,26	Single,Stat	6
Land.30S.00N	75W.30W	1908	-0.28	0.011	0.010	0.100	0.012	0.100	0.050	0.010	0.010	0.010	1915	0.0055	0.0655	0.0046	19,20	Single,Stat	4
Land.30S.00N	75W.30W	1938	0.29	-0.000	0.010	0.100	0.099	0.100	0.010	0.010	0.010	0.010	1925	0.0098	0.9889	0.0311	26,26	Single,Stat	4
Land.30S.00N	75W.30W	1976	0.19	0.015	0.010	0.100	0.030	0.100	0.010	0.010	0.010	0.010	1976	0.1130	0.1410	0.0055	19,20	Single,Stat	6
Land.30S.00N	75W.30W	1993	0.14	-0.009	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2006	0.2140	0.3741	0.2640	20,20	Single,Stat	1
Land.30N.60N	75W.30W	1929	0.48	-0.013	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1961	0.0196	0.0389	0.0020	25,26	Single,Stat	4
Land.30N.60N	75W.30W	1997	1.07	0.003	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1997	0.0011	0.9093	0.0000	25,26	Single,Stat	4
Land.60S.30S	75W.30W	1931	0.36	-0.005	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1942	0.0007	0.1891	0.0023	26,26	Single,Stat	4
Land.60S.30S	75W.30W	1976	0.27	0.005	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1945	0.0130	0.2391	0.0134	26,26	Single,Stat	4
Land.60N.90N	75W.30W	1922	1.62	-0.030	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1925	0.0000	0.0025	0.0000	25,26	Single,Stat	6
Land.60N.90N	75W.30W	1997	1.29	0.042	0.025	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1994	0.0003	0.1783	0.0000	25,26	Single,Stat	5
Land.90S.60S	75W.30W	1967	0.57	-0.014	0.100	0.100	0.031	0.100	0.010	0.010	0.010	0.010	1967	0.0003	0.2481	0.0001	25,26	Single,Stat	4
Land.90S.60S	75W.30W	1987	0.53	0.022	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1980	0.0356	0.2600	0.0863	20,20	Single,Stat	0
Land.00N.30N	30W.15E	1978	0.42	0.010	0.100	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1915	0.0163	0.5298	0.0000	25,26	Single,Stat	5
Land.00N.30N	30W.15E	1995	0.37	0.005	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1995	0.0273	0.7433	0.0826	20,20	Single,Stat	0
Land.30N.60N	30W.15E	1892	0.70	0.048	0.010	0.100	0.023	0.054	0.010	0.010	0.010	0.010	1891	0.0002	0.0477	0.0004	25,26	Single,Stat	5
Land.30N.60N	30W.15E	1986	0.41	0.016	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1961	0.0056	0.0429	0.0000	25,26	Single,Stat	6
Land.60N.90N	30W.15E	1921	0.95	-0.021	0.010	0.100	0.010	0.087	0.010	0.010	0.010	0.010	1921	0.0000	0.0030	0.0000	25,26	Single,Stat	5
Land.60N.90N	30W.15E	2000	1.23	0.020	0.010	0.100	0.010	0.013	1.000	0.010	0.010	0.010	1965	0.0000	0.5150	0.0000	22,25	Single,Stat	4
Land.00N.30N	15E.60E	1902	-0.29	-0.002	0.021	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1902	0.0113	0.8464	0.0348	20,20	Single,Stat	4
Land.00N.30N	15E.60E	1923	0.26	-0.000	0.010	0.100	0.042	0.029	0.010	0.010	0.010	0.010	1942	0.0243	0.9630	0.0101	25,25	Single,Stat	4

Land.00N.30N	15E.60E	1994	0.41	0.026	0.010	0.100	0.010	0.023	0.050	0.010	0.010	0.010	1973	0.0008	0.0035	0.0000	25,25	Single,Stat	6
Land.30S.00N	15E.60E	1982	0.25	0.011	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1982	0.0024	0.0047	0.0000	25,26	Single,Stat	6
Land.30N.60N	15E.60E	1933	0.08	0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1988	0.5990	0.6444	0.8045	26,26	Single,Stat	0
Land.30N.60N	15E.60E	1997	0.74	0.014	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1997	0.0031	0.5017	0.0000	25,26	Single,Stat	4
Land.60S.30S	15E.60E	1901	-0.98	0.038	0.010	0.100	0.075	0.100	1.000	1.000	1.000	0.050	1901	0.0000	0.0107	0.0000	11,20	Multiple,Stat	4
Land.60S.30S	15E.60E	1917	0.14	-0.030	0.010	0.100	0.023	0.100	0.010	0.010	0.010	0.010	1914	0.3524	0.0541	0.0015	25,26	Single,Stat	6
Land.60S.30S	15E.60E	1982	0.43	-0.014	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1982	0.0068	0.3622	0.0056	26,26	Single,Stat	5
Land.60S.30S	15E.60E	1997	0.44	0.017	0.010	0.100	0.100	0.100	0.050	0.050	0.100	0.050	1997	0.0068	0.3241	0.0182	11,20	Multiple,Stat	4
Land.60N.90N	15E.60E	1987	0.48	0.024	0.010	0.100	0.058	0.100	0.010	0.010	0.010	0.010	1962	0.1759	0.2367	0.0029	26,26	Single,Stat	5
Land.00N.30N	60E.105E	1937	0.20	-0.006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1902	0.0210	0.0635	0.0340	26,26	Single,Stat	5
Land.00N.30N	60E.105E	1978	0.25	0.005	0.010	0.100	0.028	0.100	0.010	0.010	0.010	0.010	1976	0.0252	0.5253	0.0072	25,26	Single,Stat	4
Land.00N.30N	60E.105E	1997	0.32	0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2008	0.0159	0.8841	0.0468	20,20	Single,Stat	4
Land.30S.00N	60E.105E	1967	0.18	0.007	0.017	0.100	0.048	0.100	0.010	0.010	0.010	0.010	1889	0.1660	0.5240	0.0140	25,26	Single,Stat	4
Land.30S.00N	60E.105E	1986	0.20	0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1977	0.1346	0.8590	0.3124	20,20	Single,Stat	0
Land.30N.60N	60E.105E	1912	0.28	-0.006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1926	0.0753	0.3877	0.0348	26,26	Single,Stat	4
Land.30N.60N	60E.105E	1976	0.40	0.016	0.010	0.100	0.019	0.100	0.010	0.010	0.010	0.010	1976	0.0633	0.3043	0.0015	25,26	Single,Stat	4
Land.30N.60N	60E.105E	1996	0.71	-0.031	0.010	0.100	0.100	0.100	0.100	0.010	0.010	0.010	1996	0.0067	0.1849	0.0165	20,20	Single,Stat	4
Land.60N.90N	60E.105E	1919	0.82	-0.025	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1945	0.0141	0.0469	0.0009	25,26	Single,Stat	4
Land.60N.90N	60E.105E	1980	0.84	0.035	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1980	0.0276	0.0392	0.0004	25,26	Single,Stat	4
Land.90S.60S	60E.105E	2001	0.62	-0.019	0.100	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2001	0.0166	0.5435	0.0026	26,26	Single,Stat	4
Ocean.00N.30N	105E.150E	1900	-0.40	-0.011	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1900	0.0002	0.1209	0.0008	26,26	Single,Stat	4
Ocean.00N.30N	105E.150E	1940	0.30	-0.006	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1940	0.0000	0.0681	0.0001	26,26	Single,Stat	4
Ocean.00N.30N	105E.150E	1977	0.20	0.011	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1946	0.0165	0.0794	0.0004	19,20	Single,Stat	4
Ocean.00N.30N	105E.150E	1997	0.35	-0.010	0.010	0.100	0.100	0.100	0.100	0.050	0.010	0.050	1997	0.0002	0.2172	0.0009	20,20	Single,Stat	4
Ocean.30S.00N	105E.150E	1968	0.21	0.009	0.010	0.100	0.010	0.083	0.010	0.010	0.010	0.010	1900	0.0511	0.1404	0.0001	25,26	Single,Stat	4
Ocean.30S.00N	105E.150E	1994	0.10	0.000	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1977	0.4310	0.9710	0.7081	20,20	Single,Stat	0
Ocean.30N.60N	105E.150E	1988	0.54	0.009	0.010	0.100	0.010	0.019	0.010	0.010	0.010	0.010	1988	0.0002	0.3034	0.0000	25,25	Single,Stat	4
Ocean.60S.30N	105E.150E	1975	0.40	0.006	0.010	0.100	0.010	0.010	0.050	0.010	0.010	0.010	1923	0.0000	0.0530	0.0000	25,25	Single,Stat	6

Ocean.60N.90N	105E.150E	1980	0.31	0.012	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1954	0.3070	0.5520	0.0783	26,26	Single,Stat	0
Ocean.60N.90N	105E.150E	2004	1.42	-0.038	0.010	0.100	0.071	0.100	0.050	0.010	0.010	0.010	2004	0.0108	0.6137	0.0145	20,20	Single,Stat	4
Ocean.90S.60S	105E.150E	1970	0.69	-0.006	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1970	0.0000	0.1171	0.0000	25,26	Single,Stat	4
Ocean.00N.30N	150E.165W	1939	0.39	0.002	0.010	0.100	0.018	0.100	0.010	0.010	0.010	0.010	1939	0.0000	0.2770	0.0000	25,26	Single,Stat	4
Ocean.00N.30N	150E.165W	1993	0.20	0.011	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1970	0.0208	0.0712	0.0000	25,26	Single,Stat	6
Ocean.30S.00N	150E.165W	1941	0.48	-0.008	0.033	0.100	0.010	0.094	0.010	0.010	0.010	0.010	1939	0.0000	0.0804	0.0000	25,26	Single,Stat	7
Ocean.30S.00N	150E.165W	1968	0.25	0.013	0.024	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1946	0.0005	0.0036	0.0000	19,20	Single,Stat	5
Ocean.30S.00N	150E.165W	1994	0.20	0.009	0.010	0.100	0.023	0.100	0.010	0.010	0.010	0.010	1994	0.0048	0.1191	0.0017	18,20	Single,Stat	6
Ocean.30N.60N	150E.165W	1896	-0.67	-0.023	0.014	0.100	0.034	0.100	0.050	0.010	0.050	0.010	1894	0.0001	0.1186	0.0002	25,26	Single,Stat	4
Ocean.30N.60N	150E.165W	1941	0.37	-0.005	0.010	0.100	0.036	0.100	0.010	0.010	0.010	0.010	1941	0.0004	0.1568	0.0003	25,26	Single,Stat	4
Ocean.30N.60N	150E.165W	1998	0.35	0.005	0.100	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1988	0.0765	0.8773	0.0059	25,26	Single,Stat	4
Ocean.30N.60N	150E.165W	2007	0.27	0.001	0.010	0.100	0.100	0.100	0.100	0.050	0.010	0.050	2010	0.0533	0.9853	0.1177	0,0	Single, N/A	0
Ocean.60S.30N	150E.165W	1969	0.26	0.004	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1934	0.0047	0.2202	0.0002	25,26	Single,Stat	4
Ocean.60N.90N	150E.165W	1976	0.35	0.009	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1962	0.1630	0.5562	0.0258	26,26	Single,Stat	4
Ocean.60N.90N	150E.165W	2001	0.85	-0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	2001	0.0482	0.9716	0.0722	20,20	Single,Stat	0
Ocean.90S.60S	150E.165W	1919	-0.15	0.004	0.029	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1957	0.0212	0.0900	0.0084	26,26	Single,Stat	4
Ocean.90S.60S	150E.165W	1969	0.32	0.001	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1992	0.0124	0.8194	0.0381	26,26	Single,Stat	4
Ocean.00N.30N	165W.120W	1935	0.23	0.003	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1976	0.0537	0.3855	0.1451	26,26	Single,Stat	2
Ocean.30S.00N	165W.120W	1981	0.30	0.003	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1906	0.0065	0.4970	0.0001	25,26	Single,Stat	4
Ocean.30N.60N	165W.120W	1902	-0.65	-0.001	0.010	0.100	0.070	0.100	0.050	0.010	0.010	0.010	1902	0.0048	0.9816	0.0077	20,20	Single,Stat	4
Ocean.30N.60N	165W.120W	1916	0.70	0.011	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1916	0.0000	0.5703	0.0000	25,26	Single,Stat	6
Ocean.60S.30N	165W.120W	1989	0.28	0.006	0.010	0.100	0.010	0.081	0.010	0.010	0.010	0.010	1906	0.0066	0.3241	0.0000	25,26	Single,Stat	4
Ocean.60N.90N	165W.120W	1918	0.93	0.009	0.010	0.100	0.064	0.100	0.010	0.010	0.010	0.010	1901	0.0031	0.4791	0.0096	26,26	Single,Stat	4
Ocean.60N.90N	165W.120W	1992	0.94	0.039	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1963	0.0096	0.1089	0.0000	25,26	Single,Stat	4
Ocean.90S.60S	165W.120W	1907	-0.27	0.001	0.016	0.100	0.028	0.100	0.050	0.050	0.010	0.010	1971	0.0011	0.8741	0.0002	25,26	Single,Stat	4
Ocean.90S.60S	165W.120W	1979	0.29	-0.001	0.010	0.100	0.035	0.100	0.010	0.010	0.010	0.010	1986	0.0021	0.7794	0.0025	25,26	Single,Stat	4
Ocean.00N.30N	120W.75W	1981	0.24	0.002	0.010	0.100	0.017	0.100	0.010	0.010	0.010	0.010	1905	0.0408	0.6934	0.0075	25,26	Single,Stat	4
Ocean.30S.00N	120W.75W	1976	0.62	-0.003	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1975	0.0001	0.6500	0.0000	25,26	Single,Stat	4

Ocean.30N.60N	120W.75W	1929	0.32	-0.007	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1961	0.0987	0.2421	0.0711	26,26	Single,Stat	0
Ocean.30N.60N	120W.75W	1997	1.16	-0.019	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1997	0.0005	0.5073	0.0001	25,26	Single,Stat	4
Ocean.60S.30N	120W.75W	1890	-0.34	0.028	0.036	0.100	0.010	0.010	0.050	0.050	0.010	0.010	1937	0.0175	0.2141	0.0000	25,25	Single,Stat	4
Ocean.60S.30N	120W.75W	1978	0.43	-0.005	0.010	0.100	0.010	0.016	0.010	0.010	0.010	0.010	1937	0.0000	0.1810	0.0000	25,25	Single,Stat	4
Ocean.60N.90N	120W.75W	1926	0.47	-0.004	0.010	0.100	0.047	0.100	0.010	0.010	0.010	0.010	1960	0.0252	0.5926	0.0389	25,26	Single,Stat	4
Ocean.60N.90N	120W.75W	1994	1.01	0.053	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1993	0.0064	0.0572	0.0000	25,26	Single,Stat	4
Ocean.90S.60S	120W.75W	1905	-0.09	0.013	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1956	0.2228	0.0038	0.0001	25,26	Single,Stat	5
Ocean.90S.60S	120W.75W	1969	0.71	-0.031	0.065	0.100	0.010	0.100	0.050	0.010	0.050	0.050	1969	0.0000	0.0282	0.0000	25,26	Single,Stat	4
Ocean.90S.60S	120W.75W	1982	0.81	0.033	0.010	0.100	0.019	0.100	0.100	0.010	0.050	0.010	1982	0.0000	0.1230	0.0001	21,20	Single,Stat	4
Ocean.00N.30N	75W.30W	1903	-0.27	0.008	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1914	0.0135	0.3274	0.0369	20,20	Single,Stat	4
Ocean.00N.30N	75W.30W	1925	0.41	-0.003	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1939	0.0001	0.6160	0.0000	25,26	Single,Stat	4
Ocean.00N.30N	75W.30W	1977	0.08	0.007	0.021	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1970	0.4260	0.2560	0.1318	26,26	Single,Stat	0
Ocean.00N.30N	75W.30W	2002	0.32	-0.006	0.010	0.100	0.045	0.100	0.100	0.010	0.050	0.010	1994	0.0134	0.6690	0.0233	18,20	Single,Stat	4
Ocean.30S.00N	75W.30W	1902	-0.28	0.012	0.052	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1902	0.0010	0.0243	0.0000	19,20	Single,Stat	6
Ocean.30S.00N	75W.30W	1938	0.23	-0.011	0.010	0.100	0.100	0.100	0.010	0.010	0.050	0.010	1938	0.0112	0.0281	0.0093	26,26	Single,Stat	5
Ocean.30S.00N	75W.30W	1968	0.27	0.004	0.019	0.100	0.027	0.100	0.050	0.010	0.010	0.010	1946	0.0438	0.7518	0.0295	18,20	Single,Stat	4
Ocean.30S.00N	75W.30W	1982	0.26	-0.013	0.051	0.100	0.100	0.100	0.010	0.010	1.000	0.050	1982	0.0176	0.2925	0.0360	11,20	Multiple,Stat	4
Ocean.30S.00N	75W.30W	1997	0.30	0.009	0.023	0.100	0.100	0.100	0.050	0.010	0.010	0.010	1997	0.0052	0.3919	0.0156	20,20	Single,Stat	4
Ocean.30N.60N	75W.30W	1926	0.54	0.003	0.010	0.100	0.010	0.100	0.100	0.010	0.050	0.050	1926	0.0000	0.6308	0.0000	22,26	Single,Stat	4
Ocean.30N.60N	75W.30W	1955	-0.38	0.001	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1961	0.0002	0.8518	0.0004	25,26	Single,Stat	4
Ocean.30N.60N	75W.30W	1997	0.60	0.006	0.010	0.100	0.010	0.100	1.000	0.050	0.010	0.010	1997	0.0000	0.5330	0.0000	22,26	Single,Stat	4
Ocean.60S.30N	75W.30W	1939	0.42	-0.000	0.018	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1939	0.0000	0.9690	0.0000	25,26	Single,Stat	5
Ocean.60S.30N	75W.30W	1967	0.27	0.011	0.067	0.100	0.035	0.100	1.000	0.050	1.000	0.050	1957	0.0650	0.6300	0.0105	9,20	Multiple,Stat	4
Ocean.60S.30N	75W.30W	1976	0.12	-0.003	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1983	0.1616	0.8582	0.1135	20,20	Single,Stat	2
Ocean.60N.90N	75W.30W	1922	1.44	-0.019	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1922	0.0000	0.0267	0.0000	25,26	Single,Stat	6
Ocean.60N.90N	75W.30W	1997	1.17	0.051	0.033	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1994	0.0008	0.0921	0.0000	25,26	Single,Stat	5
Ocean.90S.60S	75W.30W	1967	0.63	0.014	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1935	0.0030	0.0427	0.0000	25,26	Single,Stat	4
Ocean.00N.30N	30W.15E	1902	-0.40	0.007	0.018	0.100	0.100	0.100	0.050	0.010	0.010	0.010	1902	0.0022	0.4553	0.0074	20,20	Single,Stat	4

Ocean.00N.30N	30W.15E	1925	0.38	-0.009	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1936	0.0007	0.2037	0.0000	25,26	Single,Stat	4
Ocean.00N.30N	30W.15E	1978	0.32	-0.005	0.036	0.100	0.098	0.100	0.010	0.010	0.010	0.010	1945	0.0164	0.6939	0.0110	26,26	Single,Stat	4
Ocean.00N.30N	30W.15E	1994	0.30	0.015	0.010	0.100	0.100	0.100	0.050	0.050	0.010	0.010	2009	0.0375	0.2893	0.0854	20,20	Single,Stat	0
Ocean.30S.00N	30W.15E	1919	0.33	0.006	0.010	0.100	0.062	0.100	0.010	0.010	0.010	0.010	1911	0.0017	0.1474	0.0045	26,26	Single,Stat	4
Ocean.30S.00N	30W.15E	1967	0.27	-0.011	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1945	0.0628	0.4300	0.1346	26,26	Single,Stat	0
Ocean.30S.00N	30W.15E	1982	0.31	0.011	0.010	0.100	0.100	0.100	0.010	0.010	0.100	0.010	1982	0.0188	0.4149	0.0561	11,20	Multiple,Stat	0
Ocean.30N.60N	30W.15E	1930	0.37	-0.003	0.010	0.100	0.026	0.100	0.010	0.010	0.010	0.010	1900	0.0003	0.5108	0.0011	25,26	Single,Stat	4
Ocean.30N.60N	30W.15E	1961	-0.38	0.010	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1968	0.0001	0.0417	0.0001	25,26	Single,Stat	6
Ocean.30N.60N	30W.15E	1994	0.51	-0.009	0.010	0.100	0.022	0.100	1.000	0.050	0.010	0.010	1994	0.0000	0.2420	0.0000	22,20	Single,Stat	5
Ocean.60S.30N	30W.15E	1968	0.53	0.002	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1967	0.0000	0.4359	0.0000	25,26	Single,Stat	4
Ocean.60N.90N	30W.15E	1922	0.53	-0.005	0.010	0.100	0.058	0.100	0.010	0.010	0.010	0.010	1923	0.0001	0.3950	0.0006	26,26	Single,Stat	4
Ocean.60N.90N	30W.15E	1961	-0.56	0.015	0.023	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1965	0.0000	0.0082	0.0000	25,26	Single,Stat	6
Ocean.60N.90N	30W.15E	2001	0.72	0.001	0.010	0.100	0.010	0.100	1.000	0.050	0.010	0.010	2001	0.0002	0.9424	0.0000	22,20	Single,Stat	5
Ocean.90S.60S	30W.15E	1972	0.80	-0.014	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1973	0.0000	0.0026	0.0000	25,26	Single,Stat	6
Ocean.00N.30N	15E.60E	1934	0.32	-0.002	0.010	0.100	0.024	0.100	0.010	0.010	0.010	0.010	1913	0.0003	0.5535	0.0008	25,26	Single,Stat	4
Ocean.00N.30N	15E.60E	1968	0.18	0.005	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1945	0.0732	0.3951	0.0976	26,26	Single,Stat	0
Ocean.00N.30N	15E.60E	1996	0.23	0.003	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1996	0.0329	0.7715	0.0503	20,20	Single,Stat	0
Ocean.30S.00N	15E.60E	1968	0.31	0.007	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1903	0.0008	0.1488	0.0000	25,26	Single,Stat	4
Ocean.30S.00N	15E.60E	1997	0.02	0.014	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1992	0.8141	0.1098	0.1590	20,20	Single,Stat	2
Ocean.30N.60N	15E.60E	1933	0.18	-0.003	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1913	0.1130	0.4270	0.1712	26,26	Single,Stat	0
Ocean.30N.60N	15E.60E	1997	0.70	0.018	0.010	0.100	0.010	0.100	0.050	0.010	0.010	0.010	1993	0.0002	0.2616	0.0000	25,26	Single,Stat	4
Ocean.60S.30N	15E.60E	1911	0.44	-0.003	0.092	0.100	0.046	0.100	1.000	0.050	1.000	0.050	1905	0.0012	0.7130	0.0039	13,20	Multiple,Stat	4
Ocean.60S.30N	15E.60E	1933	0.55	0.008	0.010	0.100	0.010	0.100	0.100	0.010	0.010	0.010	1929	0.0001	0.3780	0.0002	22,26	Single,Stat	4
Ocean.60S.30N	15E.60E	1970	0.15	0.018	0.069	0.100	0.043	0.100	0.050	0.010	0.010	0.010	1963	0.2600	0.2420	0.0222	19,20	Single,Stat	6
Ocean.60S.30N	15E.60E	1983	0.16	-0.015	0.010	0.100	0.010	0.100	0.050	0.010	0.050	0.050	2001	0.0515	0.1310	0.0027	18,20	Single,Stat	5
Ocean.60N.90N	15E.60E	1919	0.78	-0.006	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1961	0.0005	0.4978	0.0001	25,26	Single,Stat	4
Ocean.60N.90N	15E.60E	1999	0.85	0.052	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1962	0.0105	0.1205	0.0000	25,26	Single,Stat	4
Ocean.90S.60S	15E.60E	1937	0.35	0.004	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1935	0.0042	0.2412	0.0143	26,26	Single,Stat	6

Ocean.00N.30N	60E.105E	1939	0.20	0.001	0.010	0.100	0.080	0.100	0.010	0.010	0.010	0.010	1911	0.0094	0.8607	0.0147	26,26	Single,Stat	4
Ocean.00N.30N	60E.105E	1976	0.28	0.002	0.010	0.100	0.016	0.100	0.010	0.010	0.010	0.010	1956	0.0121	0.8437	0.0123	19,20	Single,Stat	4
Ocean.00N.30N	60E.105E	1996	0.17	0.002	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1996	0.0920	0.7920	0.2105	20,20	Single,Stat	0
Ocean.30S.00N	60E.105E	1976	0.28	0.006	0.010	0.100	0.010	0.100	0.010	0.010	0.010	0.010	1945	0.0037	0.3669	0.0000	25,26	Single,Stat	4
Ocean.30S.00N	60E.105E	2000	0.11	0.012	0.010	0.100	0.082	0.100	0.010	0.010	0.010	0.010	1991	0.2402	0.1945	0.0667	20,20	Single,Stat	2
Ocean.60S.30N	60E.105E	1897	-0.24	-0.001	0.010	0.100	0.100	0.100	0.100	0.050	0.010	0.010	1913	0.0064	0.9090	0.0175	23,20	Single,Stat	4
Ocean.60S.30N	60E.105E	1925	-0.39	0.018	0.022	0.100	0.015	0.100	1.000	0.010	0.010	0.010	1913	0.0033	0.4736	0.0010	18,20	Single,Stat	4
Ocean.60S.30N	60E.105E	1932	0.49	-0.021	0.033	0.100	0.010	0.100	1.000	0.100	0.050	0.010	1941	0.0005	0.5351	0.0000	22,23	Single,Stat	4
Ocean.60S.30N	60E.105E	1975	0.42	0.008	0.010	0.100	0.010	0.100	1.000	0.010	0.010	0.010	1974	0.0000	0.0059	0.0000	22,26	Single,Stat	6
Ocean.60N.90N	60E.105E	1919	0.92	0.012	0.010	0.100	0.091	0.100	0.010	0.010	0.010	0.010	1962	0.0019	0.2634	0.0068	26,26	Single,Stat	4
Ocean.60N.90N	60E.105E	2004	1.96	-0.048	0.010	0.100	0.010	0.025	0.010	0.010	0.010	0.010	1962	0.0009	0.5895	0.0000	25,25	Single,Stat	4
Ocean.90S.60S	60E.105E	1896	-0.20	0.004	0.021	0.100	0.100	0.100	1.000	0.050	0.010	0.010	1897	0.0116	0.6452	0.0353	2,2	Single, N/A	4
Ocean.90S.60S	60E.105E	1908	0.26	-0.010	0.010	0.100	0.010	0.100	1.000	0.100	0.050	0.010	1926	0.0010	0.2990	0.0001	21,23	Single,Stat	4
Ocean.90S.60S	60E.105E	1936	0.32	0.008	0.010	0.100	0.100	0.100	0.010	0.010	0.010	0.010	1991	0.1202	0.5040	0.2776	26,26	Single,Stat	0

Table A5.1.34 RCP2.6 Mean/median return intervals for statistically significant (ANCOVA p <=0.05) shift-like regime changes in each zone from an ensemble of 25 global climate models.

1850 to 1950	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	182	182	114	136	119	195
Ocean	124	94	101	97	76	182
Land/Ocean	144	124	97	101	78	227

1951 to 1975	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	84	61	84	61	96	135
Ocean	68	48	68	68	48	75
Land/Ocean	61	68	96	61	48	68

1976 to 2040	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	26	28	25	30	30	46
Ocean	25	21	25	24	23	41
Land/Ocean	25	22	25	29	22	43

2041 to 2100	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	162	231	180	231	180	203
Ocean	90	62	116	231	52	108
Land/Ocean	108	180	135	231	52	147

1850 to 2100	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	68	71	61	69	69	103
Ocean	56	44	55	58	42	81
Land/Ocean	59	56	57	63	42	92

Table A5.1.35 RCP8.5 Mean return intervals for significant (ANCOVA p<=0.05 or CPindex=0) shift-like regime changes in each zone from an ensemble of 28 global climate models..

1850 to 1950	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	130	114	88	105	88	160
Ocean	76	51	74	70	43	83
Land/Ocean	80	72	76	78	42	114

1951 to 1975	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	68	56	68	52	96	113
Ocean	48	45	56	61	38	61
Land/Ocean	42	68	75	56	40	56

1976 to 2040	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
21/21	22	22	24	27	23	34
21/21	20	17	22	22	20	31
18/18	18	19	20	23	20	28

2041 to 2100	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	25	29	27	25	26	23
Ocean	18	18	27	27	20	25
Land/Ocean	22	23	28	26	20	25

1850 to 2100	60N.90N	30N.60N	00N.30N	30S.00N	60S.30S	90S.60S
Land	38	39	39	40	38	46
Ocean	30	25	35	35	26	40
Land/Ocean	30	32	35	36	27	41

Table A5.1.36; matching. Table A5.1.32. Results of the main findings, break years, shifts and diagnostics for NCDC observed zonal annual average temperatures from the MSBV (adapted from Ricketts 2015.) The bivariate test shows the first changed year, the internal shift and change of trend at that time. The unit root tests show the statistical significance of each test under two conditions. The A column shows the test applied to the segment of data containing a change-point and, the B column shows the same test on the residuals after the implicit shift and trend-changes are removed. The exogenous year for the ZA test and its time difference from the MSBV change-point are also shown. ANOVA tests are provided for the significance of a change of trend, and independently a change of intercept where the time is relative to the year of change (both should be considered in the context of the ANCOVA). The ANCOVA tests the two segment regression at the change-point against the single regression and is equivalent to a Chow test. The segment classification is as per Table A4.1.29. Pink fill indicates a finding of stationarity or exogenous change. Red text indicates findings for which the false determination rate exceeds 5% and which were ignores in classifying the change-points. Green shading indicates probabilities of 5% or less.

Zone	Bivariate			KPSS				ADF		Zivot-Andrews(ZA)				ANOVA/ANVOVA			Segment Classification	
				Level		Trend						Change Year		Level Change ANOVA	Trend Change ANOVA	Level or Trend Change ANCOVA		
	Year	^{0}C	^{0}C/Year	Pr A	Pr B	Pr A	Pr B	Pr A	Pr B	Pr A	Pr B	Year		Pr	Pr	Pr	Code	Class
land.60N.90N	1920	0.61	-0.0056	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.05	1949	29	0.44	0.00	0.00	25,26	Single, Stat
land.60N.90N	1988	0.56	0.0373	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1963	-25	0.00	<0.01	<0.01	25,26	Single, Stat
land.30N.60N	1894	0.16	-0.0134	0.01	0.10	0.10	0.10	0.05	0.01	0.05	0.01	1902	8	0.34	0.11	0.07	20,20	Single, Stat
land.30N.60N	1921	0.28	0.0005	0.01	0.10	0.04	0.10	0.01	0.01	0.01	0.01	1913	-8	0.94	0.02	0.01	25,26	Single, Stat
land.30N.60N	1981	0.27	0.0162	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1963	-18	0.29	0.00	0.03	25,26	Single, Stat
land.30N.60N	1997	0.54	-0.0195	0.01	0.10	0.10	0.10	0.10	0.01	0.01	0.01	1997	0	0.31	0.01	0.00	20,20	Single, Stat
land.00N.30N	1903	-0.36	0.0006	0.01	0.10	0.10	0.10	0.05	0.01	0.01	0.01	1902	-1	0.90	<0.01	<0.01	20,20	Single, Stat
land.00N.30N	1926	0.16	-0.0041	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1942	16	0.42	0.01	0.02	25,26	Single, Stat
land.00N.30N	1979	0.16	0.0101	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1977	-2	0.17	0.00	0.02	25,26	Single, Stat
land.00N.30N	1998	0.3	0.0006	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1997	-1	0.96	0.03	0.01	20,20	Single, Stat
land.30S.00N	1937	0.33	-0.0092	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1939	2	0.11	0.00	<0.01	25,26	Single, Stat
land.30S.00N	1957	0.26	0.0131	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1957	0	0.09	0.01	0.00	20,20	Single, Stat
land.30S.00N	1979	0.2	0.0091	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1976	-3	0.20	0.05	0.01	20,20	Single, Stat
land.30S.00N	2002	0.17	-0.0086	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	2010	8	0.43	0.23	0.04	20,20	Single, Stat
land.60S.30S	1938	0.09	0.0011	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1902	-36	0.67	0.26	0.12	26,26	Single, Stat

land.60S.30S	1977	0.28	0.0005	0.01	0.10	0.03	0.10	0.01	0.01	0.01	0.05	1976	-1	0.91	0.00	0.00	25,26	Single, Stat
land.60S.30S	2003	0.07	0.0225	0.01	0.10	0.04	0.10	0.05	0.01	0.01	0.01	1991	-12	0.10	0.05	0.06	18,20	Single, Stat
land.60S.60N	1921	0.1	0.0031	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1902	-19	0.65	0.16	0.15	20,20	Single, Stat
land.60S.60N	1938	0.12	-0.0038	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1944	6	0.60	0.08	0.07	20,20	Single, Stat
land.60S.60N	1979	0.22	0.015	0.01	0.10	0.01	0.10	0.05	0.01	0.01	0.01	1976	-3	0.04	<0.01	0.00	25,26	Single, Stat
land.60S.60N	1997	0.29	-0.0065	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1997	0	0.47	0.01	0.00	20,20	Single, Stat
land.20N.90N	1921	0.23	-0.003	0.01	0.10	0.02	0.10	0.01	0.01	0.01	0.01	1963	42	0.34	0.00	0.00	25,26	Single, Stat
land.20N.90N	1988	0.47	0.0008	0.01	0.10	0.01	0.10	0.05	0.01	0.01	0.01	1963	-25	0.98	0.00	0.01	25,26	Single, Stat
land.20N.90N	1998	0.42	0.004	0.01	0.10	0.09	0.10	0.10	0.01	0.05	0.01	1997	-1	0.86	0.02	0.01	2,2	Single, N/A
land.20S.20N	1904	-0.17	0.0021	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1903	-1	0.57	0.01	0.00	20,20	Single, Stat
land.20S.20N	1926	0.22	0.0046	0.01	0.10	0.03	0.10	0.05	0.01	0.01	0.01	1939	13	0.30	0.00	0.00	19,20	Single, Stat
land.20S.20N	1957	0.13	-0.0004	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1942	-15	0.94	0.27	0.12	20,20	Single, Stat
land.20S.20N	1979	0.19	0.0106	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1976	-3	0.23	0.06	0.02	20,20	Single, Stat
land.20S.20N	1997	0.12	0.0022	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	2001	4	0.82	0.45	0.18	20,20	Single, Stat
land.90S.20S	1926	0.24	-0.0014	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1943	17	0.76	0.03	0.01	26,26	Single, Stat
land.90S.20S	1957	0.23	0.0014	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1957	0	0.85	0.04	0.02	20,20	Single, Stat
land.90S.20S	1977	0.26	0.0071	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1976	-1	0.35	0.03	0.00	20,20	Single, Stat
land.90S.20S	2002	0.15	0.0083	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	2010	8	0.48	0.10	0.08	20,20	Single, Stat
land.00N.90N	1921	0.25	-0.0019	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1936	15	0.46	0.00	0.00	25,26	Single, Stat
land.00N.90N	1980	0.2	0.019	0.01	0.10	0.01	0.10	0.05	0.01	0.01	0.01	1963	-17	0.05	<0.01	0.00	25,26	Single, Stat
land.00N.90N	1997	0.32	-0.0074	0.01	0.10	0.10	0.10	0.05	0.01	0.01	0.01	1997	0	0.54	0.03	0.01	20,20	Single, Stat
land.90S.00N	1937	0.32	-0.0095	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1939	2	0.09	0.00	<0.01	25,26	Single, Stat
land.90S.00N	1957	0.24	0.0141	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1957	0	0.06	0.01	0.00	20,20	Single, Stat
land.90S.00N	1979	0.19	0.007	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1976	-3	0.28	0.05	0.01	20,20	Single, Stat
land.90S.00N	2002	0.14	-0.0036	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	2010	8	0.71	0.23	0.07	20,20	Single, Stat
land.90S.90N	1925	0.18	0	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1936	11	0.99	0.01	0.00	26,26	Single, Stat
land.90S.90N	1980	0.2	0.015	0.01	0.10	0.01	0.10	0.05	0.01	0.01	0.01	1963	-17	0.07	<0.01	0.00	25,26	Single, Stat
land.90S.90N	1997	0.25	-0.0042	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1997	0	0.67	0.05	0.01	20,20	Single, Stat

land_ocean.60N.90N	1920	0.58	0.0023	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1919	-1	0.66	0.00	<0.01	25,26	Single, Stat
land_ocean.60N.90N	1988	0.53	0.0115	0.03	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1963	-25	0.63	0.00	0.01	25,26	Single, Stat
land_ocean.60N.90N	2002	0.34	0.0105	0.01	0.10	0.10	0.10	0.05	0.01	0.01	0.01	2001	-1	0.73	0.31	0.11	2,2	Single, N/A
land_ocean.30N.60N	1921	0.34	0.0011	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1920	-1	0.62	<0.01	<0.01	25,26	Single, Stat
land_ocean.30N.60N	1988	0.37	-0.0141	0.04	0.10	0.03	0.10	0.01	0.01	0.01	0.01	1963	-25	0.51	0.00	0.00	25,26	Single, Stat
land_ocean.30N.60N	1997	0.43	0.0188	0.01	0.10	0.10	0.10	0.10	0.01	0.10	0.01	1996	-1	0.34	0.00	0.00	2,2	Single, N/A
land_ocean.00N.30N	1904	-0.23	0.0033	0.01	0.10	0.10	0.10	0.01	0.01	0.10	0.01	1906	2	0.54	0.01	0.00	11,20	Multiple, Stat
land_ocean.00N.30N	1926	0.25	-0.0025	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1942	16	0.59	<0.01	0.00	25,26	Single, Stat
land_ocean.00N.30N	1979	0.11	0.0075	0.01	0.10	0.05	0.10	0.01	0.01	0.01	0.01	1970	-9	0.24	0.01	0.04	25,26	Single, Stat
land_ocean.00N.30N	1997	0.14	-0.0022	0.01	0.10	0.10	0.10	0.01	0.01	0.05	0.05	1986	-11	0.76	0.19	0.05	20,20	Single, Stat
land_ocean.30S.00N	1937	0.16	0.0044	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1902	-35	0.04	0.00	0.00	25,26	Single, Stat
land_ocean.30S.00N	1979	0.18	0.0044	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1945	-34	0.49	0.00	0.01	25,26	Single, Stat
land_ocean.30S.00N	1997	0.11	-0.0042	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.05	2001	4	0.52	0.23	0.05	20,20	Single, Stat
land_ocean.60S.30S	1887	-0.19	0.0168	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.01	1913	26	0.36	<0.01	0.00	13,20	Multiple, Stat
land_ocean.60S.30S	1937	0.24	0.0015	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.01	1931	-6	0.48	<0.01	<0.01	13,26	Multiple, Stat
land_ocean.60S.30S	1968	0.09	0.0174	0.01	0.10	0.02	0.10	0.10	0.01	>0.1	0.05	1962	-6	0.08	0.00	0.01	12,20	Single, Stat
land_ocean.60S.30S	1977	0.05	-0.0176	0.01	0.10	0.02	0.10	>0.1	0.01	>0.1	0.01	1981	4	0.01	0.00	0.00	0,2	Single, N/A
land_ocean.60S.30S	1996	0.1	0.0014	0.01	0.10	0.10	0.10	>0.1	0.01	>0.1	0.01	1995	-1	0.61	0.00	0.00	11,20	Multiple, Stat
land_ocean.60S.60N	1903	-0.14	-0.0041	0.01	0.10	0.08	0.10	0.10	0.01	>0.1	0.01	1894	-9	0.57	0.01	0.01	11,20	Multiple, Stat
land_ocean.60S.60N	1914	0.2	0.0023	0.02	0.10	0.10	0.10	>0.1	0.01	>0.1	0.01	1913	-1	0.79	0.01	0.00	2,2	Single, N/A
land_ocean.60S.60N	1925	0.1	0.0092	0.01	0.10	0.10	0.10	0.01	0.01	0.05	0.01	1915	-10	0.26	0.11	0.02	2,2	Single, N/A
land_ocean.60S.60N	1937	0.11	-0.0048	0.01	0.10	0.10	0.10	0.05	0.01	0.01	0.01	1945	8	0.53	0.01	0.03	20,20	Single, Stat
land_ocean.60S.60N	1979	0.14	0.0091	0.01	0.10	0.01	0.04	0.05	>0.1	0.01	>0.1	1945	-34	0.04	<0.01	0.00	25,13	Single, Non-Stat
land_ocean.60S.60N	1997	0.17	-0.0054	0.01	0.10	0.10	0.10	0.01	0.01	0.05	0.05	1996	-1	0.25	0.00	0.00	20,20	Single, Stat
land_ocean.20N.90N	1925	0.33	0.0003	0.01	0.10	0.01	0.10	0.01	0.01	0.01	>0.1	1920	-5	0.85	<0.01	<0.01	25,17	Non-Stat
land_ocean.20N.90N	1988	0.29	0.0079	0.01	0.10	0.01	0.10	0.10	>0.1	0.01	0.10	1963	-25	0.63	<0.01	0.01	22,14	Single, Non-Stat
land_ocean.20N.90N	1998	0.28	0.0004	0.01	0.10	0.07	0.10	>0.1	0.10	>0.1	>0.1	1996	-2	0.97	0.01	0.00	2,2	Single, N/A
land_ocean.20S.20N	1936	0.21	0.0027	0.01	0.10	0.01	0.10	0.01	0.05	0.01	0.10	1925	-11	0.25	0.00	0.00	25,17	Non-Stat

land_ocean.20S.20N	1979	0.18	0.0067	0.01	0.10	0.02	0.10	0.01	0.01	0.01	0.05	1945	-34	0.35	0.00	0.01	25,26	Single, Stat
land_ocean.20S.20N	1997	0.1	-0.0044	0.01	0.10	0.10	0.10	0.01	0.01	0.05	0.01	2001	4	0.59	0.44	0.14	20,20	Single, Stat
land_ocean.90S.20S	1887	-0.17	0.0121	0.01	0.10	0.01	0.10	>0.1	0.01	0.05	0.01	1911	24	0.40	<0.01	0.00	22,20	Single, Stat
land_ocean.90S.20S	1937	0.23	-0.0001	0.01	0.10	0.01	0.10	0.10	0.01	>0.1	0.01	1931	-6	0.96	<0.01	<0.01	13,26	Multiple, Stat
land_ocean.90S.20S	1969	0.19	0.0046	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.01	1945	-24	0.71	0.00	0.01	12,20	Single, Stat
land_ocean.90S.20S	1977	0.08	-0.0013	0.01	0.10	0.10	0.01	>0.1	0.05	0.10	0.01	1976	-1	0.85	0.03	0.03	2,0	Single, Stat
land_ocean.90S.20S	1997	0.1	-0.0003	0.01	0.10	0.10	0.10	>0.1	0.01	>0.1	0.01	1995	-2	0.89	0.00	0.00	11,20	Multiple, Stat
land_ocean.00N.90N	1925	0.31	0.0034	0.01	0.10	0.01	0.10	0.01	0.05	0.01	0.05	1923	-2	0.03	<0.01	<0.01	25,26	Single, Stat
land_ocean.00N.90N	1987	0.22	0.005	0.01	0.10	0.01	0.10	0.05	0.01	0.01	0.01	1963	-24	0.70	<0.01	0.01	25,26	Single, Stat
land_ocean.00N.90N	1997	0.22	0.0031	0.01	0.10	0.08	0.10	0.10	0.01	0.05	0.01	1996	-1	0.77	0.01	0.00	2,2	Single, N/A
land_ocean.90S.00N	1890	-0.12	0.0143	0.02	0.10	0.01	0.01	>0.1	>0.1	0.01	0.05	1911	21	0.11	<0.01	0.00	22,22	Single, Stat
land_ocean.90S.00N	1937	0.2	-0.001	0.01	0.10	0.03	0.10	0.05	0.01	0.10	0.01	1911	-26	0.60	<0.01	<0.01	16,26	Multiple, Stat
land_ocean.90S.00N	1969	0.18	-0.0005	0.01	0.10	0.02	0.10	0.05	0.10	0.05	0.10	1945	-24	0.96	0.01	0.02	18,14	Single, Non-Stat
land_ocean.90S.00N	1979	0.12	0.0058	0.01	0.10	0.10	0.10	0.01	>0.1	0.01	>0.1	1976	-3	0.48	0.10	0.02	2,2	Single, N/A
land_ocean.90S.00N	1997	0.11	-0.0032	0.01	0.10	0.10	0.10	0.01	0.05	0.05	0.05	1996	-1	0.38	0.02	0.00	20,20	Single, Stat
land_ocean.90S.90N	1930	0.25	0.0032	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1913	-17	0.02	<0.01	<0.01	25,26	Single, Stat
land_ocean.90S.90N	1979	0.12	0.0089	0.01	0.10	0.01	0.10	0.05	0.01	0.01	0.01	1945	-34	0.04	<0.01	0.00	25,26	Single, Stat
land_ocean.90S.90N	1997	0.16	-0.0049	0.01	0.10	0.10	0.10	0.05	0.01	0.01	0.01	1996	-1	0.29	0.01	0.00	20,20	Single, Stat
ocean.60N.90N	1926	0.4	0.002	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1925	-1	0.24	<0.01	<0.01	25,26	Single, Stat
ocean.60N.90N	2000	0.48	0.0249	0.01	0.10	0.01	0.10	>0.1	0.01	0.01	0.01	1981	-19	0.01	<0.01	<0.01	22,26	Single, Stat
ocean.30N.60N	1902	-0.34	0.0031	0.01	0.10	0.04	0.10	0.10	0.01	0.10	0.01	1901	-1	0.62	<0.01	<0.01	9,20	Multiple, Stat
ocean.30N.60N	1915	0.21	0.0135	0.01	0.10	0.10	0.10	0.05	0.05	0.10	0.10	1914	-1	0.06	0.00	<0.01	2,2	Single, N/A
ocean.30N.60N	1930	0.13	-0.0113	0.01	0.10	0.02	0.10	0.01	0.01	0.10	0.01	1938	8	0.10	0.00	0.01	9,20	Multiple, Stat
ocean.30N.60N	1964	-0.18	-0.0063	0.01	0.10	0.02	0.10	0.01	0.01	0.01	0.05	1963	-1	0.09	0.00	0.00	25,26	Single, Stat
ocean.30N.60N	1989	0.28	0.0068	0.02	0.10	0.01	0.10	>0.1	0.01	0.05	>0.1	1988	-1	0.62	0.00	0.00	18,11	Single, Non-Stat
ocean.30N.60N	1998	0.21	0.0074	0.01	0.10	0.10	0.10	0.05	0.01	0.10	>0.1	1993	-5	0.59	0.03	0.01	2,2	Single, N/A
ocean.00N.30N	1907	-0.23	0.01	0.01	0.10	0.10	0.10	0.01	0.01	0.10	>0.1	1906	-1	0.13	0.02	0.00	11,11	Single, Non-Stat
ocean.00N.30N	1926	0.24	-0.0061	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.05	1945	19	0.28	<0.01	0.00	25,26	Single, Stat

ocean.00N.30N	1987	0.15	0.0051	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.01	1969	-18	0.12	0.00	0.00	25,26	Single, Stat
ocean.30S.00N	1937	0.15	0.0029	0.01	0.10	0.02	0.10	0.01	0.01	0.01	0.10	1911	-26	0.18	0.01	0.00	25,17	Non-Stat
ocean.30S.00N	1979	0.18	0.0036	0.01	0.10	0.01	0.10	0.01	0.01	0.01	0.10	1945	-34	0.58	0.01	0.01	25,17	Non-Stat
ocean.30S.00N	1997	0.13	-0.0056	0.01	0.10	0.10	0.10	0.01	0.01	0.05	0.01	2001	4	0.36	0.09	0.02	20,20	Single, Stat
ocean.60S.30S	1887	-0.2	0.0171	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.01	1913	26	0.37	<0.01	0.00	13,20	Multiple, Stat
ocean.60S.30S	1937	0.24	0.0029	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.05	1931	-6	0.15	<0.01	<0.01	13,26	Multiple, Stat
ocean.60S.30S	1970	0.14	0.0111	0.01	0.10	0.03	0.10	0.10	0.01	>0.1	0.10	1962	-8	0.46	0.00	0.03	12,11	Non-Stat
ocean.60S.30S	1977	0.05	-0.0122	0.01	0.10	0.08	0.10	>0.1	0.01	>0.1	0.05	1981	4	0.15	0.02	0.02	2,2	Single, N/A
ocean.60S.30S	1996	0.11	0.0001	0.01	0.10	0.10	0.10	>0.1	0.01	>0.1	0.10	1995	-1	0.97	0.00	0.00	11,11	Single, Non-Stat
ocean.60S.60N	1890	-0.14	0.008	0.02	0.10	0.01	0.10	>0.1	0.01	>0.1	0.01	1913	23	0.45	0.00	0.02	12,20	Single, Stat
ocean.60S.60N	1930	0.22	0.0011	0.01	0.10	0.02	0.10	0.01	0.01	>0.1	0.01	1913	-17	0.54	<0.01	<0.01	16,26	Multiple, Stat
ocean.60S.60N	1977	0.11	-0.0017	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1945	-32	0.87	0.07	0.10	20,20	Single, Stat
ocean.60S.60N	1987	0.1	0.0013	0.01	0.10	0.10	0.10	0.01	0.01	>0.1	0.01	1986	-1	0.85	0.08	0.02	2,2	Single, N/A
ocean.60S.60N	1997	0.14	0.0013	0.01	0.10	0.10	0.10	0.05	0.01	>0.1	0.01	1996	-1	0.85	0.01	0.00	2,2	Single, N/A
ocean.20N.90N	1902	-0.21	-0.0035	0.01	0.10	0.04	0.10	0.10	0.01	>0.1	0.01	1901	-1	0.51	<0.01	<0.01	9,20	Multiple, Stat
ocean.20N.90N	1915	0.14	0.0245	0.01	0.10	0.02	0.04	>0.1	0.01	>0.1	0.01	1907	-8	0.00	0.00	<0.01	0,0	Single, N/A
ocean.20N.90N	1930	0.13	-0.0154	0.01	0.10	0.01	0.01	>0.1	0.01	0.10	0.01	1945	15	0.00	<0.01	<0.01	12,18	Multiple, Stat
ocean.20N.90N	1964	-0.16	-0.0036	0.01	0.10	0.02	0.04	0.01	>0.1	0.05	>0.1	1963	-1	0.18	0.00	0.00	19,13	Single, Non-Stat
ocean.20N.90N	1988	0.2	0.0056	0.01	0.10	0.02	0.10	>0.1	0.01	>0.1	0.05	1987	-1	0.60	0.00	0.00	9,20	Multiple, Stat
ocean.20N.90N	1997	0.2	0.0067	0.01	0.10	0.09	0.10	0.10	0.01	>0.1	>0.1	1993	-4	0.51	0.01	0.00	2,2	Single, N/A
ocean.20S.20N	1926	0.26	0.006	0.01	0.10	0.03	0.10	0.01	>0.1	0.01	>0.1	1924	-2	0.01	0.00	<0.01	25,14	Single, Non-Stat
ocean.20S.20N	1979	0.16	0.0049	0.01	0.10	0.03	0.10	0.01	0.10	0.01	>0.1	1945	-34	0.48	0.01	0.03	25,14	Single, Non-Stat
ocean.20S.20N	1997	0.1	-0.0064	0.01	0.10	0.10	0.01	0.01	>0.1	0.05	0.10	2001	4	0.43	0.40	0.11	20,9	Single, Stat
ocean.90S.20S	1887	-0.18	0.013	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.05	1911	24	0.39	<0.01	0.00	13,20	Multiple, Stat
ocean.90S.20S	1937	0.25	-0.0001	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	0.01	1933	-4	0.98	<0.01	<0.01	13,26	Multiple, Stat
ocean.90S.20S	1969	0.17	0.0121	0.01	0.10	0.01	0.10	>0.1	0.01	>0.1	>0.1	1944	-25	0.32	<0.01	0.01	12,11	Non-Stat
ocean.90S.20S	1977	0.04	-0.0096	0.01	0.10	0.07	0.10	>0.1	0.01	>0.1	0.10	1987	10	0.12	0.02	0.02	2,2	Single, N/A
ocean.90S.20S	1996	0.11	-0.0002	0.01	0.10	0.10	0.01	>0.1	>0.1	>0.1	0.05	1995	-1	0.92	0.00	0.00	11,18	Single, Non-Stat

ocean.00N.90N	1903	-0.25	0.0106	0.01	0.10	0.02	0.10	>0.1	0.01	>0.1	0.05	1906	3	0.01	<0.01	<0.01	12,20	Single, Stat
ocean.00N.90N	1926	0.27	-0.0079	0.01	0.10	0.01	0.10	0.10	0.05	0.01	0.10	1945	19	0.01	<0.01	<0.01	22,17	Single, Non-Stat
ocean.00N.90N	1987	0.13	0.0071	0.04	0.10	0.02	0.10	0.01	0.01	0.05	0.01	1969	-18	0.50	0.00	0.04	25,26	Single, Stat
ocean.00N.90N	1997	0.13	0.0001	0.01	0.10	0.10	0.10	0.05	0.01	>0.1	0.01	1996	-1	0.99	0.07	0.03	2,2	Single, N/A
ocean.90S.00N	1890	-0.12	0.0161	0.02	0.10	0.01	0.01	>0.1	>0.1	0.01	0.10	1911	21	0.07	<0.01	0.00	22,13	Single, Non-Stat
ocean.90S.00N	1937	0.2	-0.0023	0.01	0.10	0.04	0.10	0.05	0.01	>0.1	0.05	1911	-26	0.25	<0.01	<0.01	16,26	Multiple, Stat
ocean.90S.00N	1969	0.18	0.0025	0.02	0.10	0.01	0.10	0.10	0.05	0.05	>0.1	1945	-24	0.82	0.00	0.01	21,11	Single, Stat
ocean.90S.00N	1979	0.1	0.0027	0.01	0.10	0.10	0.10	0.01	>0.1	0.05	>0.1	1976	-3	0.70	0.09	0.02	2,2	Single, N/A
ocean.90S.00N	1997	0.12	-0.0044	0.01	0.10	0.10	0.10	0.01	0.10	0.05	>0.1	1996	-1	0.16	0.00	0.00	20,11	Single, Non-Stat
ocean.90S.90N	1890	-0.13	0.0083	0.01	0.10	0.01	0.01	>0.1	>0.1	>0.1	0.10	1913	23	0.43	0.00	0.02	12,12	Non-Stat
ocean.90S.90N	1930	0.22	0.001	0.01	0.10	0.02	0.10	0.05	0.01	>0.1	0.05	1913	-17	0.56	<0.01	<0.01	16,26	Multiple, Stat
ocean.90S.90N	1977	0.11	-0.0017	0.01	0.10	0.10	0.10	0.01	0.01	0.01	0.01	1945	-32	0.87	0.07	0.10	20,20	Single, Stat
ocean.90S.90N	1987	0.1	0.0021	0.01	0.10	0.10	0.10	0.05	0.01	>0.1	>0.1	1986	-1	0.77	0.07	0.02	2,2	Single, N/A
ocean.90S.90N	1997	0.14	0.0011	0.01	0.10	0.10	0.10	0.05	0.01	>0.1	0.10	1996	-1	0.88	0.01	0.00	2,2	Single, N/A

Appendix 5.2: Additional figures

Figure A5.1.54: Pre-industrial Control ensemble Internal shifts (left) and change of internal trend (right) for all significant change-points (ANCOVA p<=0.05).

Figure A5.1.55: RCP2.6 ensemble, Internal shifts (left) and change of internal trend (right) for all significant change-points (ANCOVA p<=0.05)

Figure A5.1.56: RCP8.5 ensemble, Internal shifts (left) and change of internal trend (right) for all significant change-points (ANCOVA p<=0.0.5)

Appendix 6.1: Quadrant correlations of climate models

Table A6.1.37: Sample shift correlations for Africa and for Northern Pacific Ocean. Note considerably higher correlation indices North relative to South for Africa

AFRICA									
Model	AFRIC A	AFRICA N	AFRICA NE	AFRICAN W	AFRICA S	AFRICA SE	AFRICAS W	Mean Rank	Over all
ACCESS1-0	0.53	0.53	0.80	0.29	0.16	0.13	0.09		
ACCESS1-3	0.40	0.47	0.29	0.31	0.03	0.15	0.00		
bcc-csm1-1	0.47	0.47	0.05	0.50	0.11	0.05	0.09		
bcc-csm1-1-m	0.17	0.11	0.03	0.12	0.00	0.04	0.00		
BNU-ESM	0.30	0.30	0.05	0.32	0.00	0.02	0.00		
CanESM2	0.01	0.01	0.03	0.00	0.01	0.09	0.00		
CESM1-BGC	0.26	0.26	0.15	0.15	0.06	0.07	0.03		
CESM1-CAM5	0.62	0.64	0.35	0.55	0.17	0.33	0.07		
IPSL-CM5A-LR	0.52	0.51	0.82	0.36	0.16	0.08	0.18		
IPSL-CM5A-MR	0.46	0.31	0.04	0.37	0.02	0.07	0.00		
IPSL-CM5B-LR	0.30	0.35	0.45	0.15	0.01	0.07	0.00		
MIROC5	0.06	0.13	0.32	0.04	0.00	0.00	0.00		
MIROC-ESM	0.17	0.04	0.02	0.04	0.00	0.00	0.00		
MIROC-ESM-CHEM	0.32	0.40	0.13	0.29	0.01	0.00	0.00		
MPI-ESM-LR	0.17	0.07	0.19	0.01	0.15	0.35	0.06		
MPI-ESM-MR	0.68	0.70	0.49	0.62	0.08	0.06	0.05		
MRI-CGCM3	0.24	0.16	0.01	0.15	0.00	0.04	0.00		
NorESM1-M	0.29	0.27	0.34	0.16	0.23	0.07	0.24		
NorESM1-ME	0.26	0.17	0.22	0.02	0.04	0.06	0.00		
Ranks	AFRIC A	AFRICA N	AFRICA NE	AFRICAN W	AFRICA S	AFRICA SE	AFRICASW		
ACCESS1-0	3	3	2	9	4	4	3	3.6	3
ACCESS1-3	7	6	8	7	10	3	10	5.6	6
bcc-csm1-1	5	5	13	3	6	13	4	6.6	7
bcc-csm1-1-m	15	16	16	14	16	15	18	12.6	19
BNU-ESM	9	10	14	6	15	16	12	9.6	12
CanESM2	19	19	17	19	13	5	19	12	16
CESM1-BGC	12	12	11	13	8	8	8	8	11
CESM1-CAM5	**2**	**2**	**5**	**2**	**2**	**2**	**5**	**2.8**	**1**
IPSL-CM5A-LR	**4**	**4**	**1**	**5**	**3**	**6**	**2**	**2.8**	**1**
IPSL-CM5A-MR	6	9	15	4	11	10	9	7.6	9
IPSL-CM5B-LR	10	8	4	12	12	7	15	7.6	9
MIROC5	18	15	7	16	18	19	14	11.2	15

MIROC-ESM	17	18	18	15	19	17	11	12.2	17
MIROC-ESM-CHEM	8	7	12	8	14	18	13	10.2	13
MPI-ESM-LR	16	17	10	18	5	1	6	7	8
MPI-ESM-MR	1	1	3	1	7	11	7	4.4	4
MRI-CGCM3	14	14	19	11	17	14	17	12.2	17
NorESM1-M	11	11	6	10	1	9	1	5.2	5
NorESM1-ME	13	13	9	17	9	12	16	10.8	14

North Pacific									
Model	NPAC	NEPAC	NEPACN	NEPACS	NWPAC	NWPACN	NWPACS	Mean Rank	Overall
ACCESS1-0	0.28	0.03	0.04	0.02	0.50	0.67	0.23		
ACCESS1-3	0.00	0.04	0.02	0.02	0.00	0.00	0.00		
bcc-csm1-1	0.22	0.00	0.03	0.01	0.34	0.45	0.03		
bcc-csm1-1-m	0.03	0.00	0.04	0.01	0.04	0.04	0.01		
BNU-ESM	0.16	0.04	0.00	0.00	0.11	0.01	0.21		
CanESM2	0.24	0.08	0.01	0.15	0.29	0.56	0.09		
CESM1-BGC	0.11	0.07	0.00	0.16	0.07	0.01	0.14		
CESM1-CAM5	0.19	0.04	0.03	0.02	0.29	0.26	0.19		
IPSL-CM5A-LR	0.14	0.02	0.01	0.03	0.23	0.13	0.08		
IPSL-CM5A-MR	0.16	0.07	0.01	0.04	0.30	0.49	0.14		
IPSL-CM5B-LR	0.07	0.00	0.07	0.04	0.01	0.00	0.00		
MIROC5	0.06	0.03	0.03	0.01	0.14	0.30	0.02		
MIROC-ESM	0.22	0.00	0.01	0.01	0.40	0.17	0.27		
MIROC-ESM-CHEM	0.00	0.00	0.01	0.00	0.01	0.05	0.00		
MPI-ESM-LR	0.07	0.05	0.06	0.01	0.06	0.02	0.03		
MPI-ESM-MR	0.21	0.05	0.05	0.04	0.48	0.42	0.24		
MRI-CGCM3	0.00	0.01	0.04	0.00	0.06	0.00	0.05		
NorESM1-M	0.19	0.01	0.00	0.01	0.28	0.29	0.08		
NorESM1-ME	0.08	0.00	0.01	0.00	0.08	0.23	0.00		
Ranks	NPAC	NEPAC	NEPACN	NEPACS	NWPAC	NWPACN	NWPACS		
ACCESS1-0	1	9	6	7	1	1	3	3.4	2
ACCESS1-3	19	8	10	8	19	19	16	10.6	16
bcc-csm1-1	4	17	7	13	4	4	12	7.2	7
bcc-csm1-1-m	16	15	4	15	16	13	15	9.4	13
BNU-ESM	9	6	17	16	11	16	4	10.6	16
CanESM2	2	1	14	2	7	2	8	5.2	3
CESM1-BGC	11	2	19	1	13	15	7	8.4	12
CESM1-CAM5	7	7	9	9	6	8	5	6.2	5
IPSL-CM5A-LR	10	11	15	6	9	11	9	8.2	11
IPSL-CM5A-MR	8	3	13	4	5	3	6	5.2	3
IPSL-CM5B-LR	13	19	1	3	18	17	17	7.6	8
MIROC5	15	10	8	12	10	6	14	8	10

MIROC-ESM	3	16	11	11	3	10	1	6.6	6
MIROC-ESM-CHEM	17	18	12	19	17	12	19	12.4	19
MPI-ESM-LR	14	5	2	10	15	14	13	7.8	9
MPI-ESM-MR	**5**	**4**	**3**	**5**	**2**	**5**	**2**	**3**	**1**
MRI-CGCM3	18	12	5	18	14	18	11	10.4	15
NorESM1-M	6	13	18	14	8	7	10	9.8	14
NorESM1-ME	12	14	16	17	12	9	18	12	18

Table A6.1.38: Preferred model, best and worst quadrants in each case for five continents and five ocean basins.

Region	GCM	Worst R^2	Best R^2	Whole Continent R^2	Comments
Africa	CESM1-CAM5	0.07 (SW)	0.35(NE)	0.62	Correlation in North greater than South
Australia	NorESM1-M	0.01 (NE)	0.19 (SE)	0.17	NE and SW bad, others not much better
Eurasia	ACCESS1-0	0.01 (NE)	0.60 (NW)	0.56	NE uncorrelated
Indian	IPSL-CM5A-MR	0.0 (NE)	0.75 (SE)	0.56	NE uncorrelated, NW 0.22
N America	bcc-csm1-1	0.02 (NW)	0.41 (NE)	0.41	SE and SW 0.13
N Atlantic	bcc-csm1-1-m	0.04 (NE)	0.56 (SW)	0.71	Combined East 0.34, Combined West 0.74, West much higher than East
S America	IPSL-CM5A-MR	0.05 (SE)	0.22 (NW)	0.45	West higher than East
S Atlantic	CanESM2	0.0 (NW)	0.47 (SE)	0.18	South higher than North
N Pacific	MPI-ESM-MR	0.04 (SE)	0.42 (NW)	0.21	No correlation East and moderate Westward
S Pacific	bcc-csm1-1-m	0.01 (NE)	0.44 (SW)	0.18	No correlation East and moderate Westward

Table A6.1.39: Ocean basin and land masses. Shown are the lowest and highest correlation coefficients (observed vs model) for the best performing model given the spatially averaged signal from each land mass or ocean basin. Also shown are the correlation coefficients for spatially averages quadrants from within the same entities.

Region	GCM	Worst Sector R^2	Best Sector R^2	Whole Continent R^2	Comments
Africa	CESM1-CAM5	0.07 (SW)	0.35 (NE)	0.62	North greater than South
Australia	NorESM1-M	0.01 (NE)	0.19 (SE)	0.17	NE and SW bad, others not much better
Eurasia	ACCESS1-0	0.01 (NE)	0.60 (NW)	0.56	NE bad
Indian	IPSL-CM5A-MR	0.0 (NE)	0.75 (SE)	0.56	NE uncorrelated, NW 0.22 Area of Gyre is high correlation especially East
N America	bcc-csm1-1	0.02 (NW)	0.41 (NE)	0.41	SE and SW 0.13
N Atlantic	bcc-csm1-1-m	0.04 (NE)	0.56 (SW)	0.71	East 0.34, West 0.74, West much higher than East
S America	IPSL-CM5A-MR	0.05 (SE)	0.22 (NW)	0.45	West higher than East
S Atlantic	CanESM2	0.0 (NW)	0.47 (SE)	0.18	South higher than North
N Pacific	MPI-ESM-MR	0.04 (SE)	0.42 (NW)	0.21	No correlation East and moderate Westward
S Pacific	bcc-csm1-1-m	0.01 (NE)	0.44 (SW)	0.18	No correlation East and moderate Westward

Appendix 6.2: Initialisation of climate models

How initialisation of a global climate model may affect a spatial analysis

Global climate models are not initialised in such a way that their internal circulation states can be expected to represent the observed ones at any specific time. They most certainly do, none the less, form modelled internal circulation states. The zonal analysis of models shows that some sort of consensus of temperature variations in zonal and global temperature averages can be extracted and found to align with observations. The findings of this work to this point suggest that there are multiple circulation systems that interact with the atmosphere and each other. The model-based zonal results therefore suggest either that a consensus of models form similar configurations; or that there are multiple circulation configurations which can affect large scale temperatures similarly so that the consensus is one of outcome, and an inference of similarity of mechanisms is not sustained. Published patterns of modelled PDO and similar indices suggest that the models overall do evolved reasonable models of individual ocean circulation systems but little is published on the interactions between such systems. Several authors have presented evidence that observed systems interact so that their state changes tend to coordinate or self organise (Tsonis et al., 2007, Wang et al., 2009, Tsonis and Swanson, 2012).

A global climate model is initialised with set boundary conditions including a "pre-industrial atmosphere" and then run for hundreds to thousands of model years ("spun-up") until it is judged to be stable and then it is switched from a stable atmosphere to an estimate of the observed atmosphere. To explore possible futures, projected future atmospheres representing various outcomes are prescribed at the end of the observational period. A "RIP" code designates runs from a single climate model that share a common ancestry.

r = "realization": simulations started from equally likely initial conditions – often branching from the same spin-up at different times

i = "initialization": different initialization procedures

p = "physics": ensemble members or simulations forced by slightly modified parameterizations

E.g R1I1P1 means realization 1, initialization method 1, physics 1.

Switching to observed conditions does not involve any matching with observations at this time. While the model runs with a prescribed atmosphere, this dictates the amount of heat retained in the model atmosphere. This heat must eventually find its way out to space, but it first interacts with the Earth surface and must be distributed based on the internal state the model has evolved and continues to evolve. Internal circulation systems must therefore distribute heat according to their own states, and under the hypothesis of interaction with forcing, will each evolve under this action, possibly constrained by other sub-systems.

Box A6.2.1: How initialisation of a global climate model may affect a spatial analysis

,

Samples of patterns of surface covariation as a function of GMST

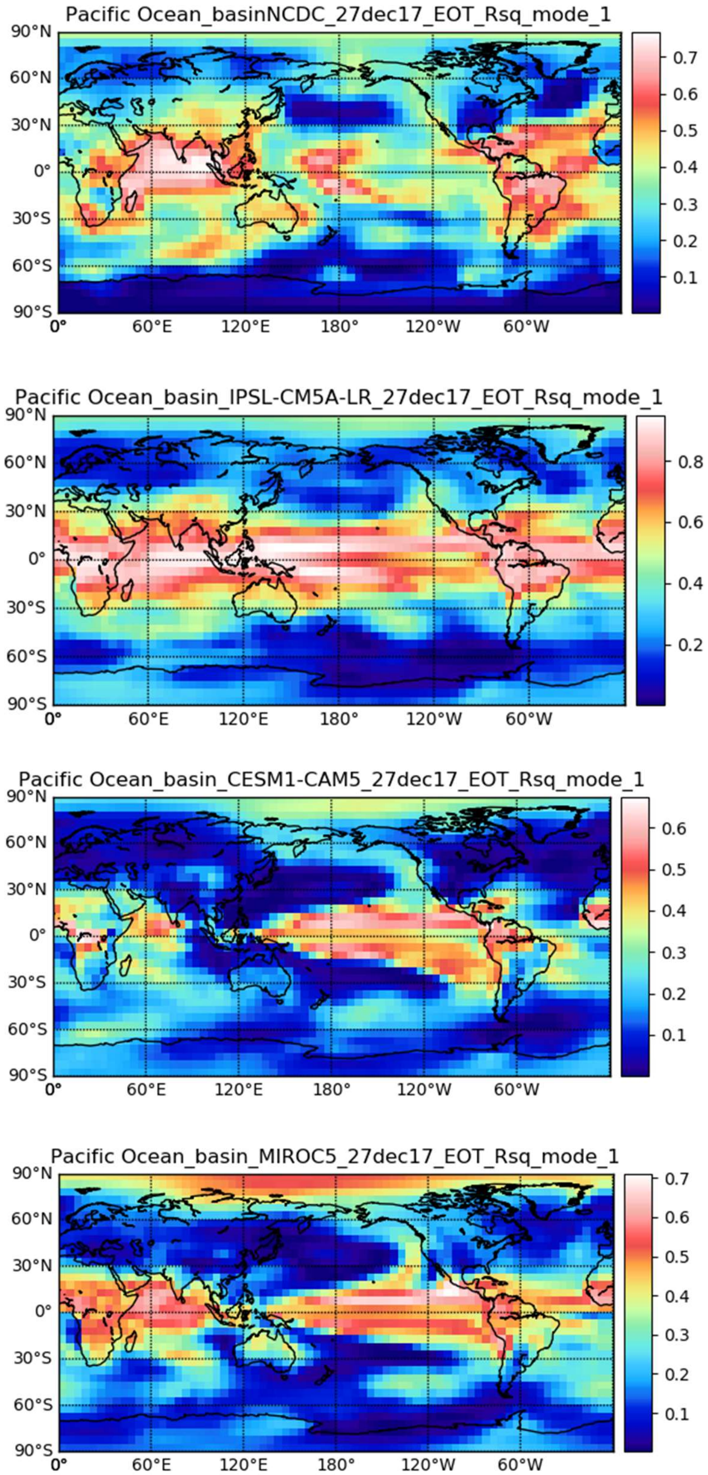


Figure A6.2.57: Three sample GCM EOT plots compared to observations (top). These demonstrate differences between observations and models in the covariation between GMST and surface temperatures at grid-scale. In particular East and West tropical Pacific regions do not co-vary, and yet this result and the Quadrant correlation both suggest they do.

Empirical Orthogonal Teleconnections (EOT) are conceptually simple. The first mode is computed by computing a time-series of the average signal over an area of interest, and then performing a correlation analysis between that and the spatial data. Higher modes then use the R^2 value to remove a correlation weighted part of the signal, then produce a new time-series from the means of the remnant and repeat. These are the R^2 values returned from a sample of climate models when tested against the signal over the Pacific. That is, how does the surface of the planet co-vary with the mean temperature of the Pacific? The observational data is the top pane, and shows that in fact the mean Pacific signal is more correlated with the Northern Indian. The basin itself shows quite confined regions of variability. The models tend to show covariation across the Tropical Pacific which is not observed.

The second EOT is IPSL-CM5A-LR which was the model selected for further analysis. All three models show more connectivity between the West and the East sides of the Pacific ocean, and this is typical across the range of models.

Sample patterns of the Pacific Decadal Oscillation for 41 GCMs, taken from the literature

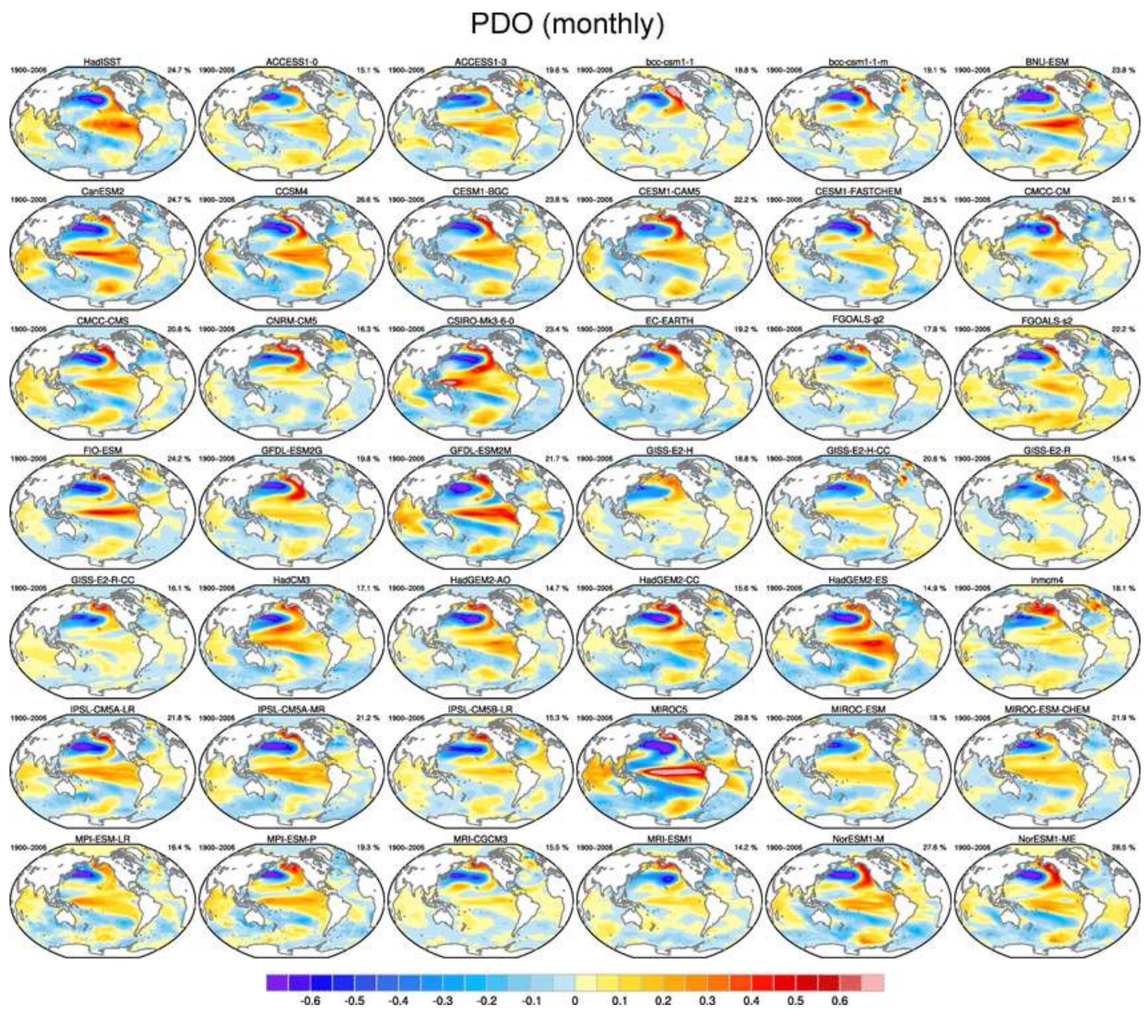
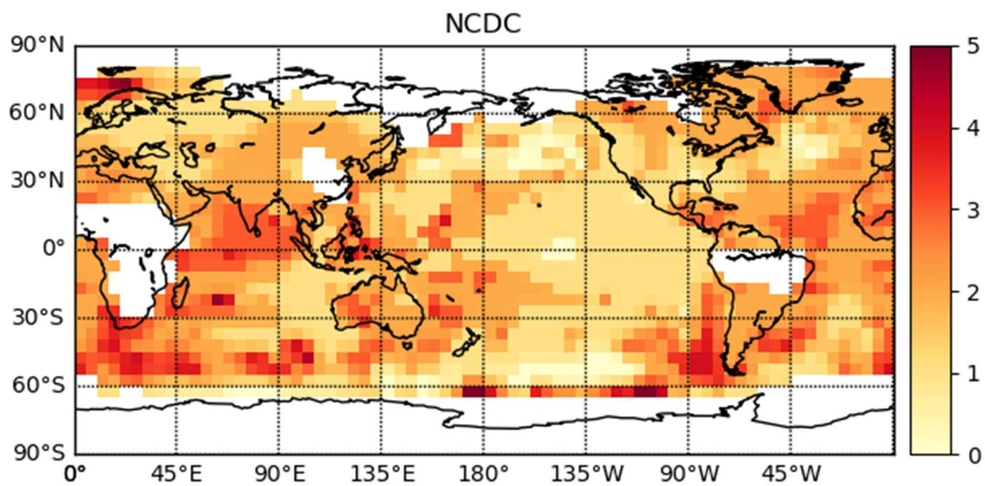
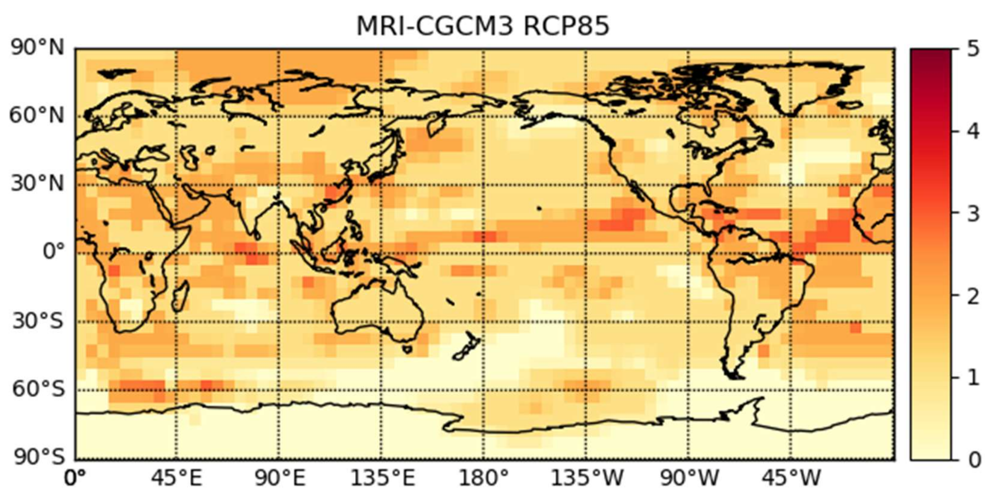


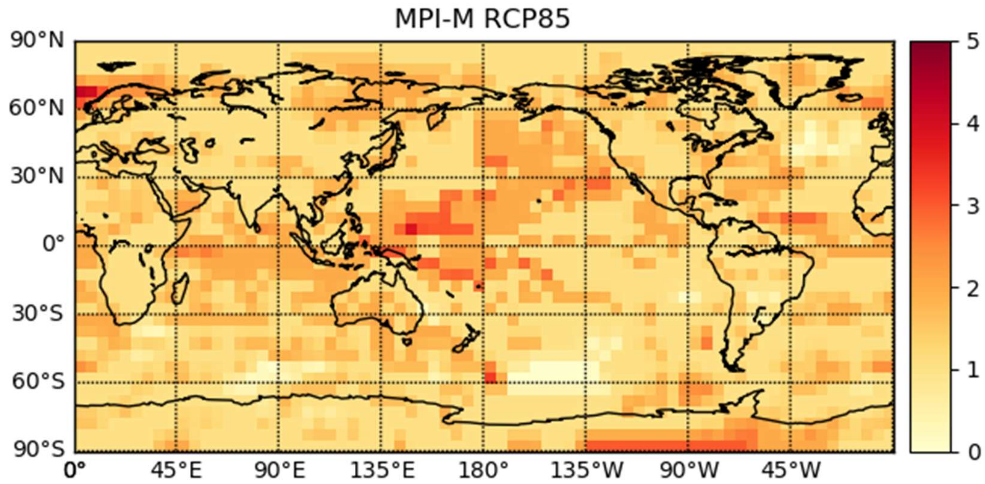
Figure A6.2.58: Figure and caption from (Eyring et al., 2016, their Figure 8) "The PDO as simulated by 41 CMIP5 models (individual panels labelled by model name) and observations (upper left panel) for the historical period 1900–2005. These patterns show the global SST anomalies (_C) associated with a one standard deviation change in the normalized principal component (PC) time series. The percent variance accounted by the PDO is given in the upper right of each panel. The PDO is defined as the leading empirical orthogonal function of monthly SST anomalies (minus the global mean SST) over the North Pacific (20–70_ N, 110_ E–100_ W). The global patterns (_C) are formed by regressing monthly SST anomalies at each grid point onto the PC time series. Most CMIP5 models show realistic patterns in the North Pacific. However, linkages with the tropics and the tropical Pacific in particular, vary across models. The lack of a strong tropical expression of the PDO is a major shortcoming in many CMIP5 models (Flato et al., 2013). Figure produced with namelist_CVDP.xml."

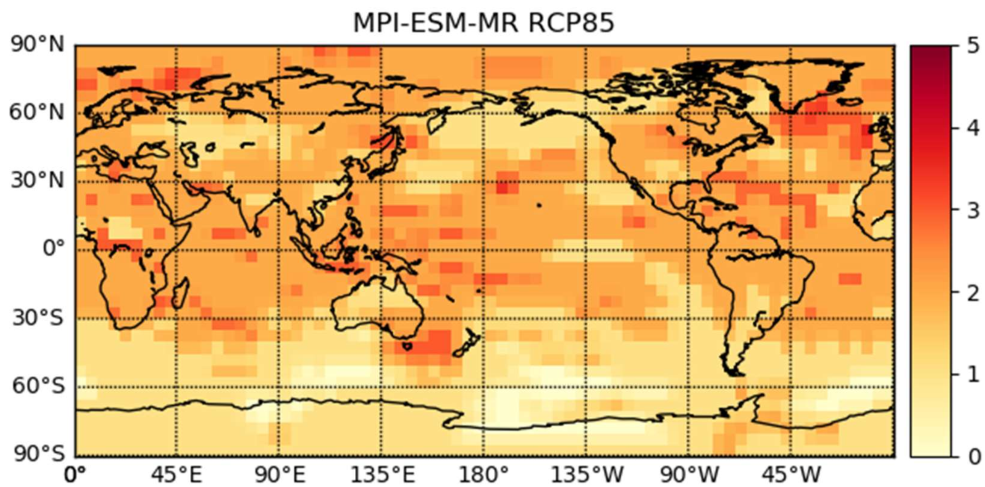
Appendix 6.3: "Hotspots", Patterns of change-point frequencies 1880-2016

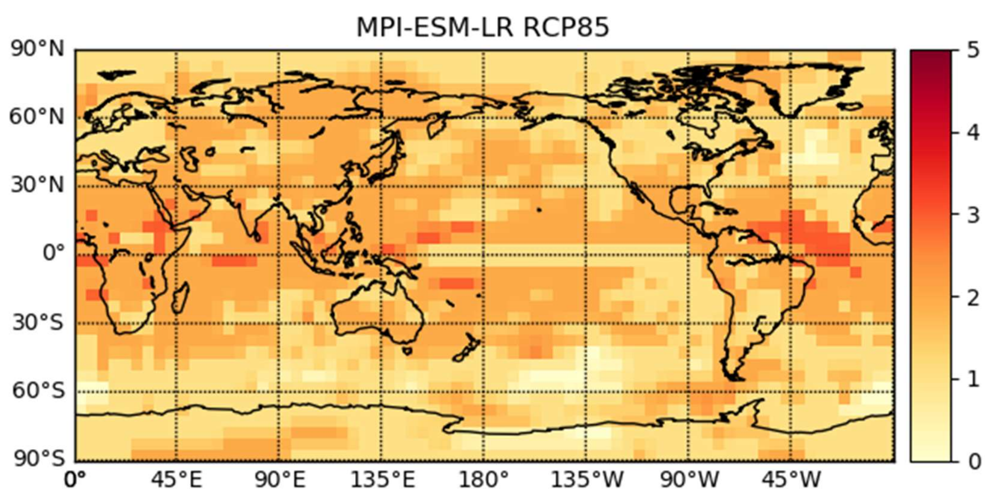
A comparison of the numbers of shift-like change-points over the period of observations for observations (GISSTEM3 obtained from NCDC) and a selection of climate models. These illustrate the variation across climate models in the locations of step-like changes.

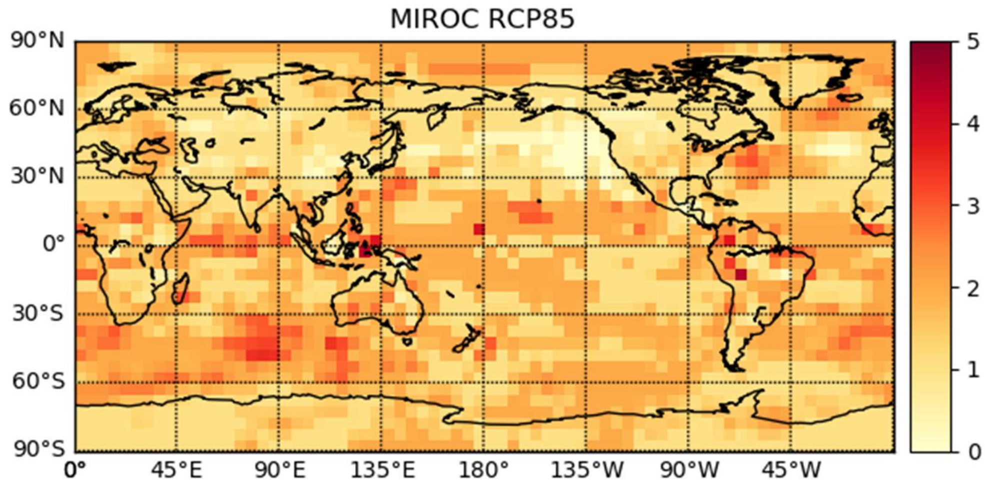


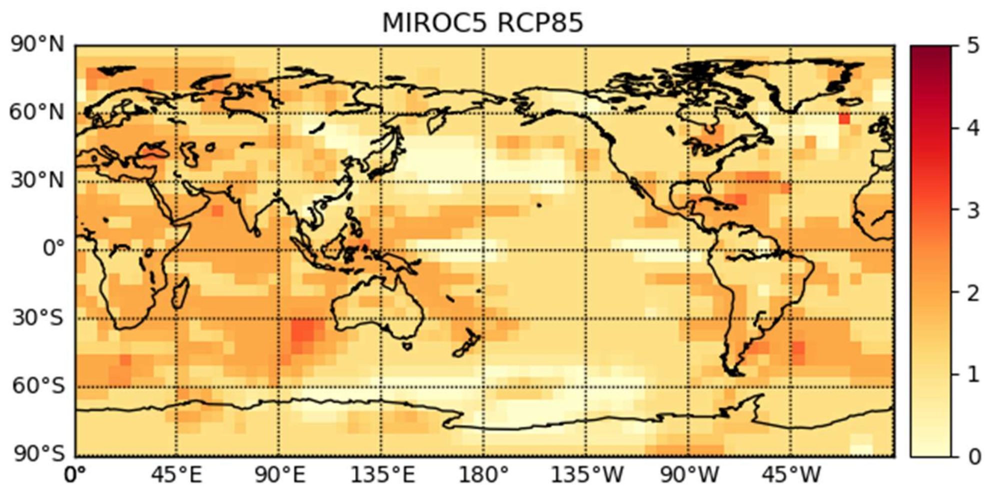


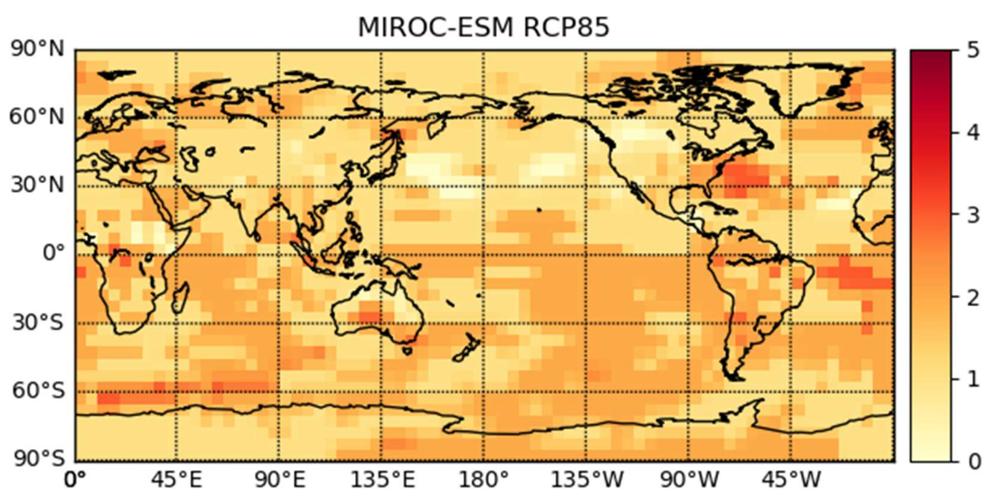

MPI-M RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

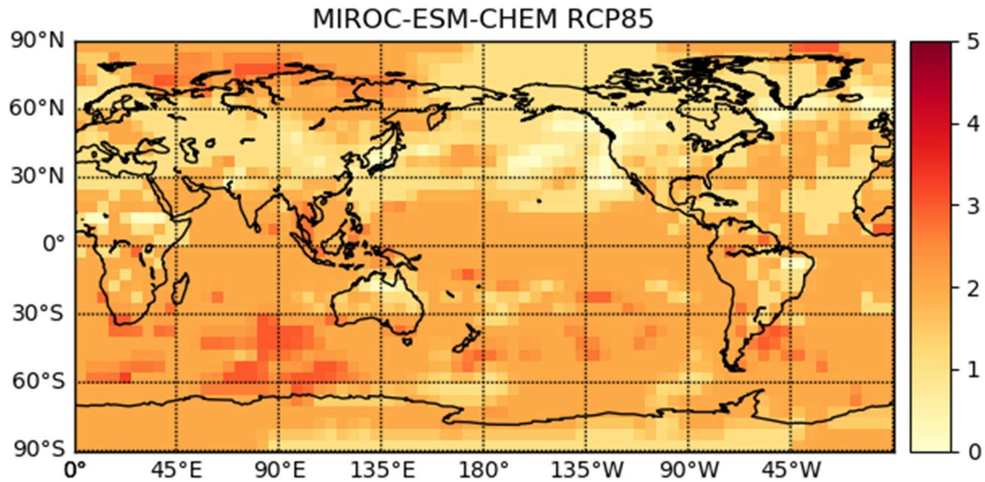

MPI-ESM-MR RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

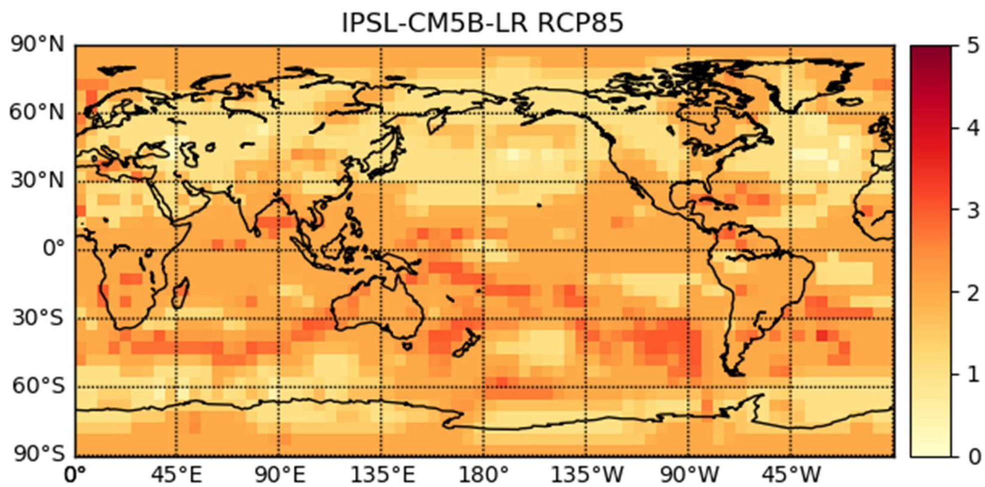

MPI-ESM-LR RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

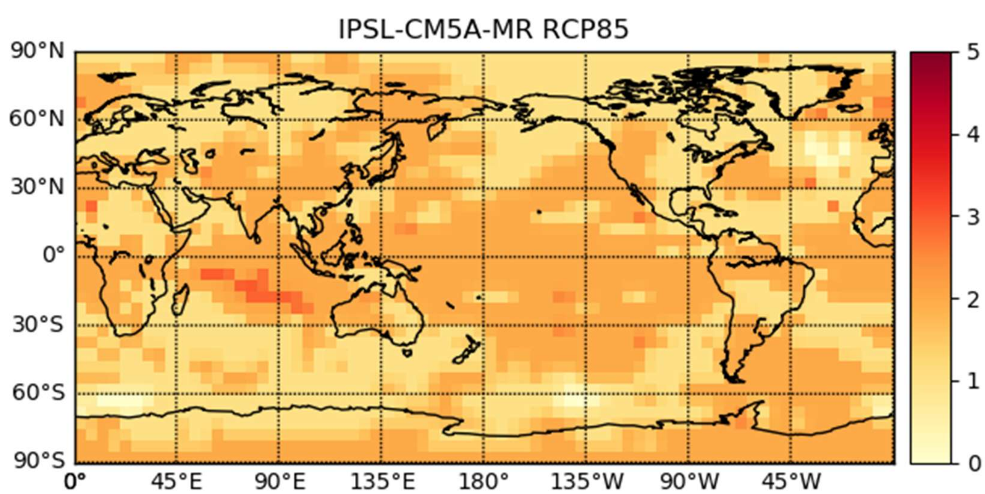

MIROC RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

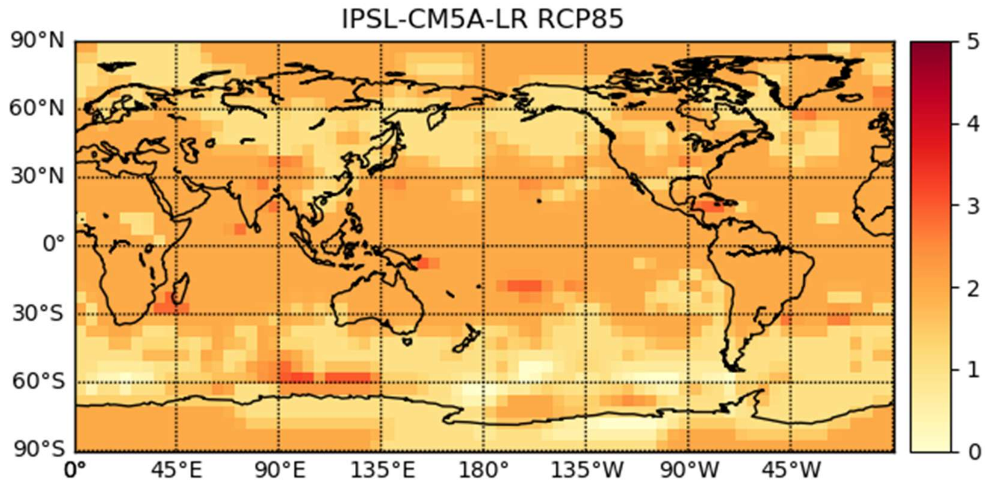

MIROC5 RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

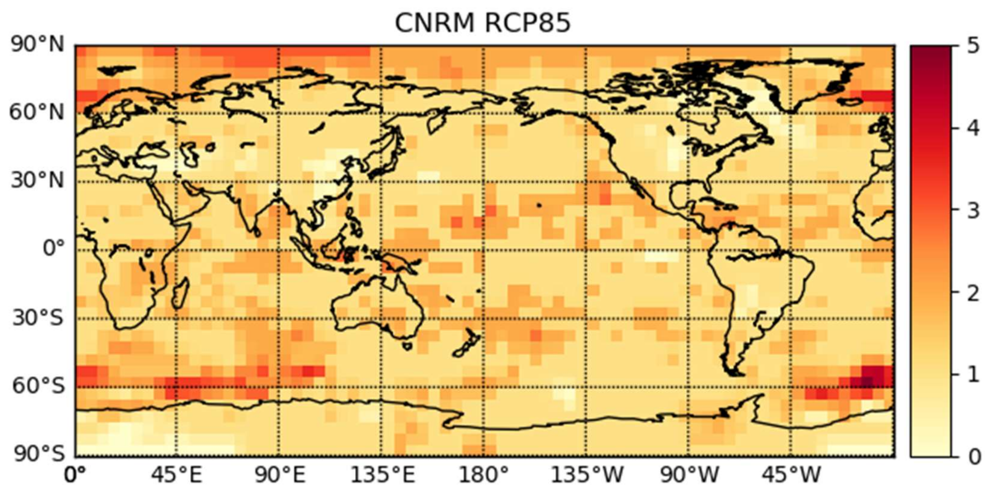

MIROC-ESM RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

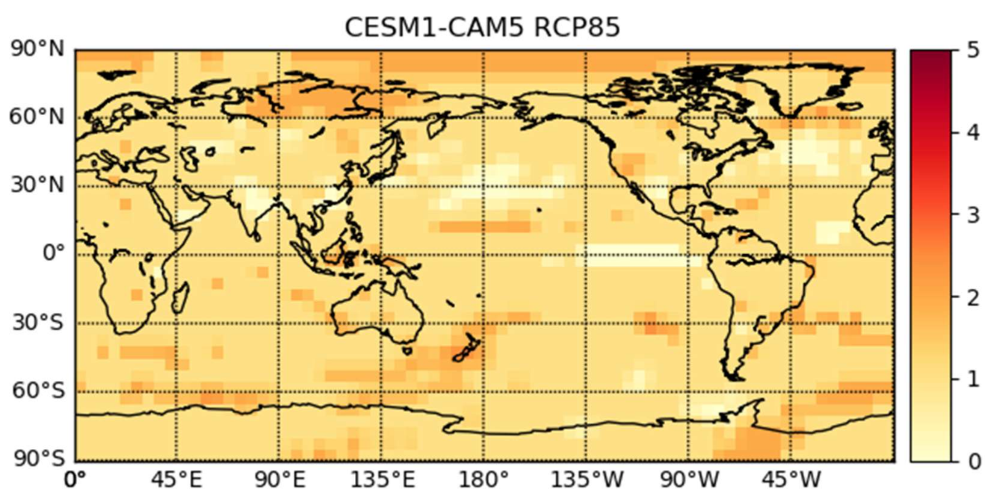

MIROC-ESM-CHEM RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

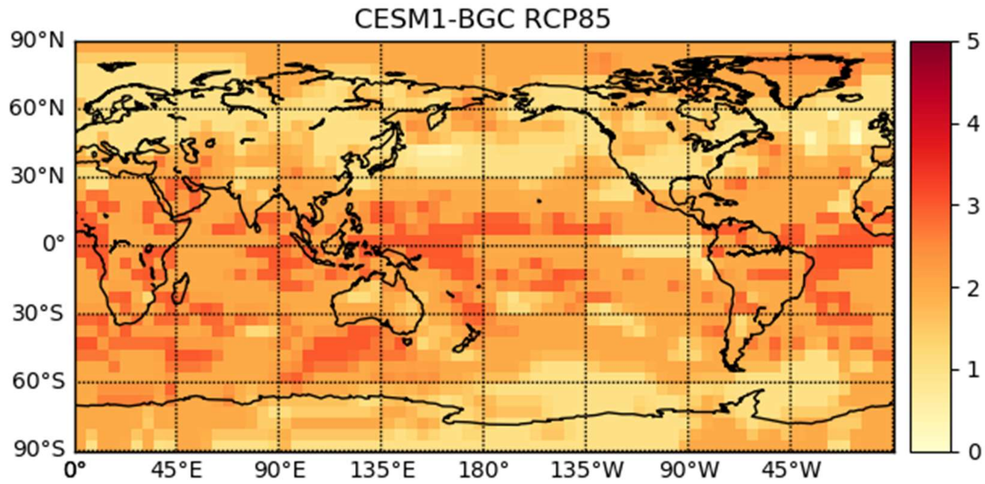

IPSL-CM5B-LR RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0


IPSL-CM5A-MR RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

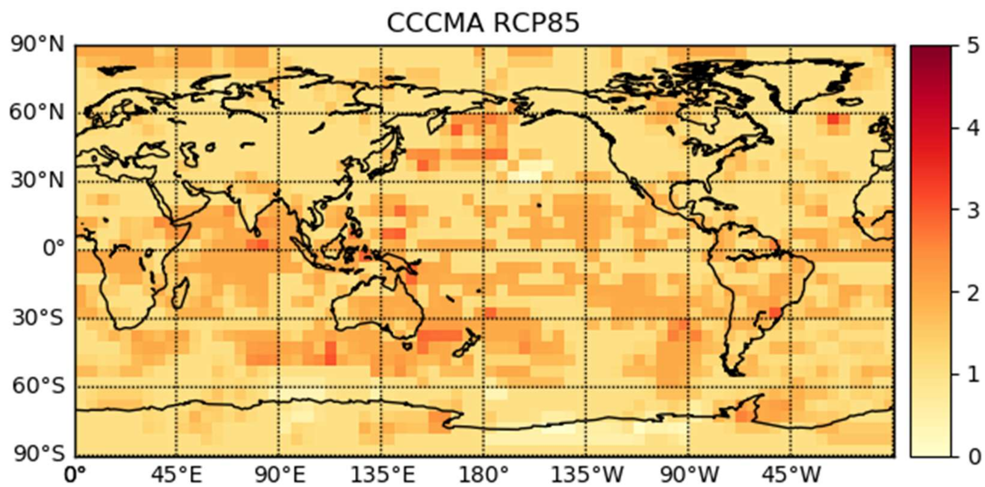

IPSL-CM5A-LR RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0


CNRM RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

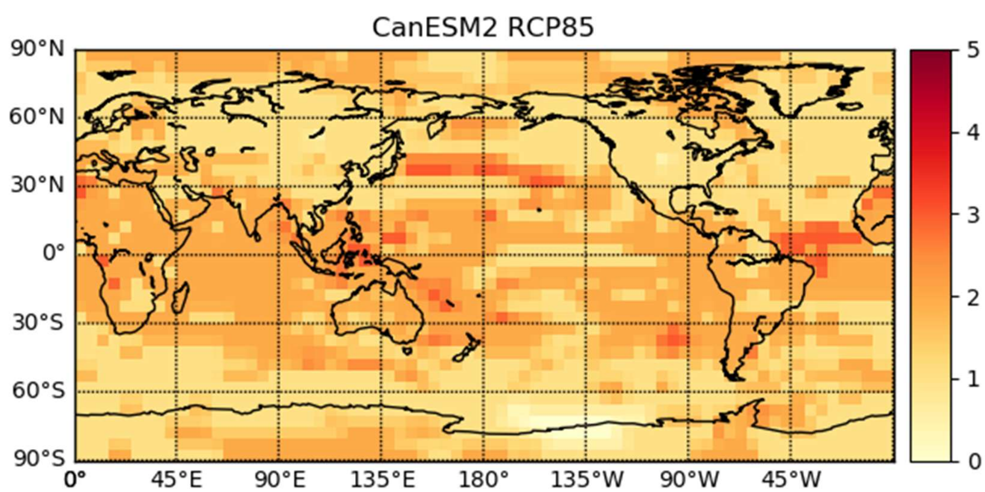

CESM1-CAM5 RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0



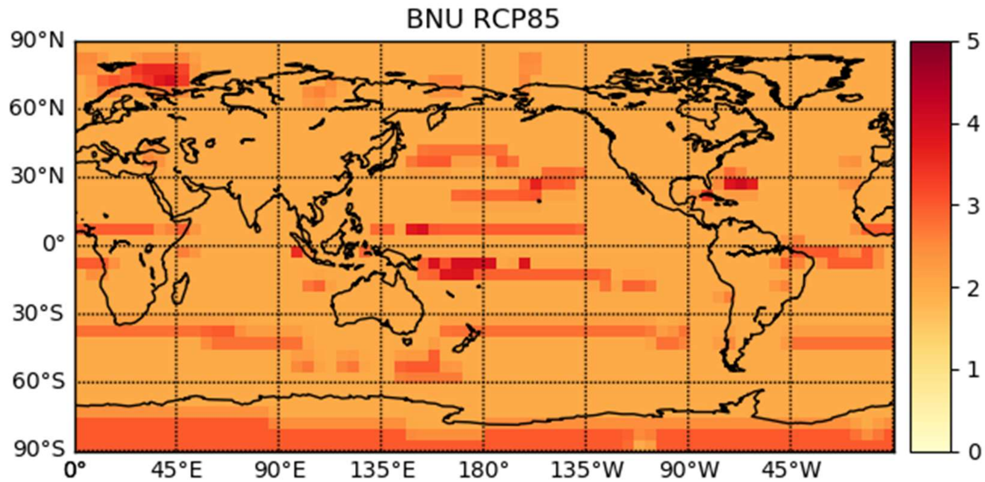
CESM1-BGC RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

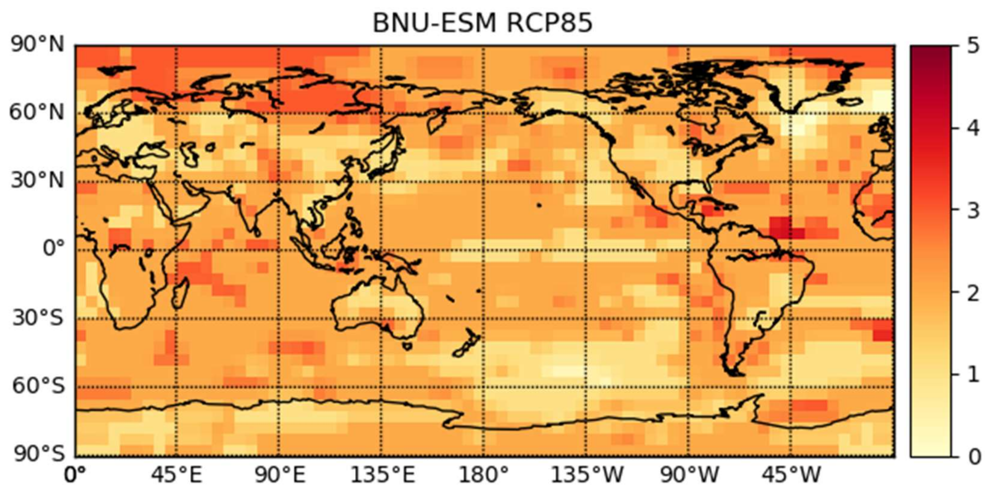


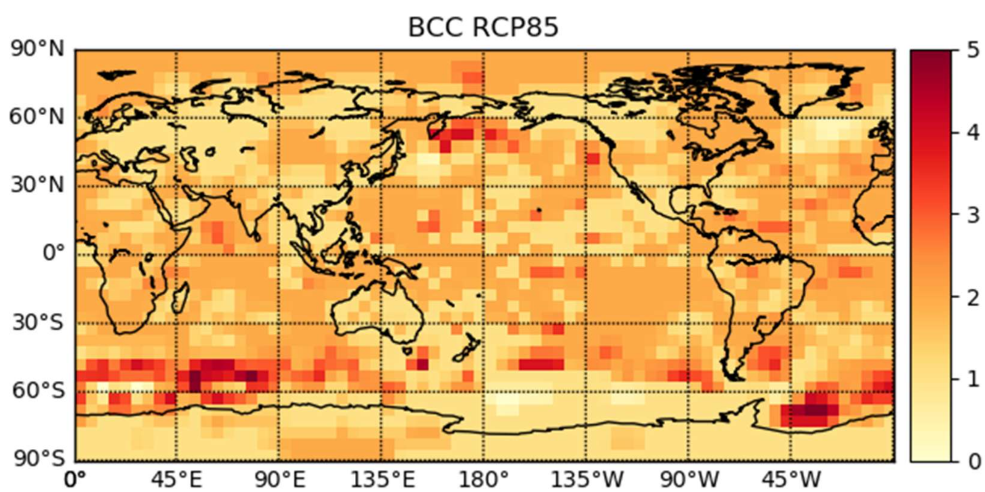
CCCMA RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

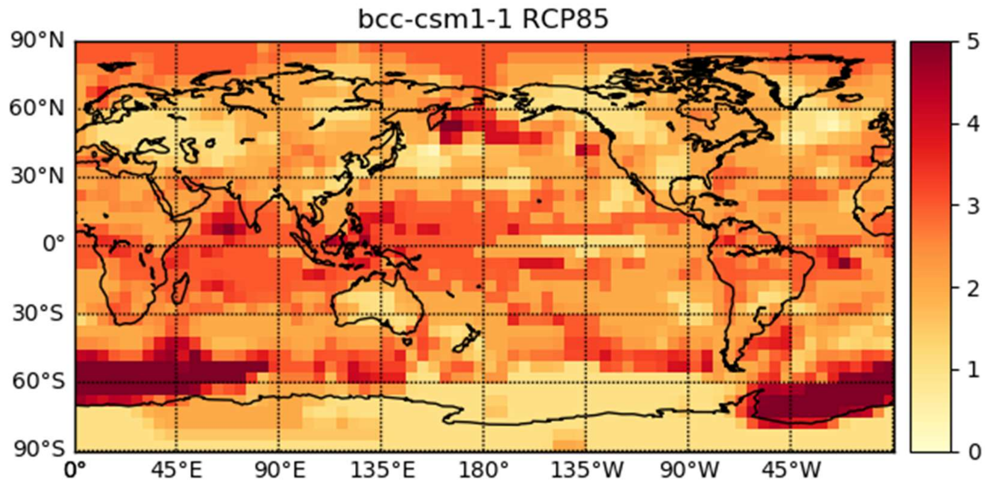


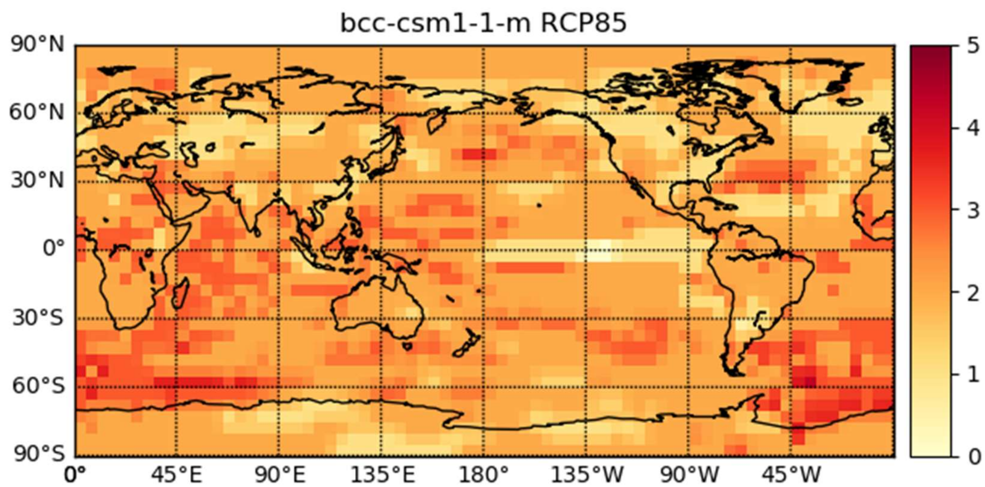
CanESM2 RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

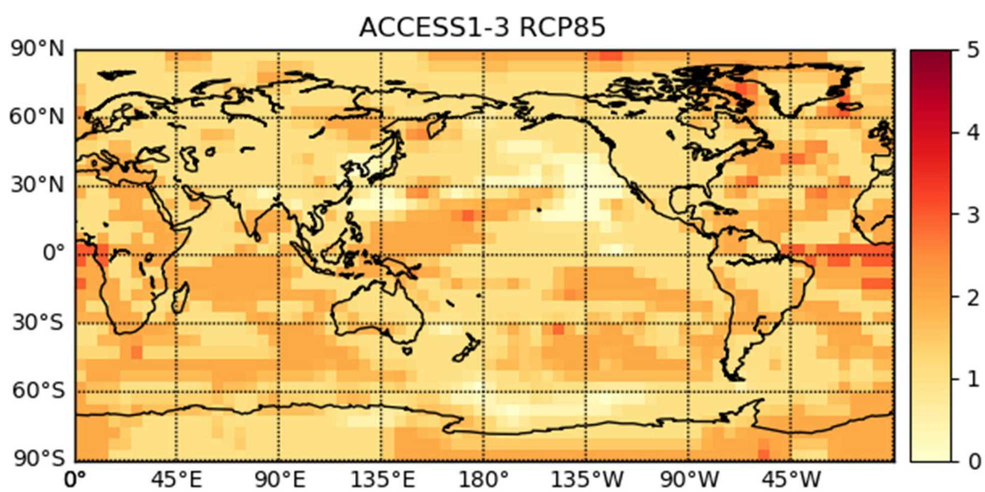

BNU RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
0
1
2
3
4
5

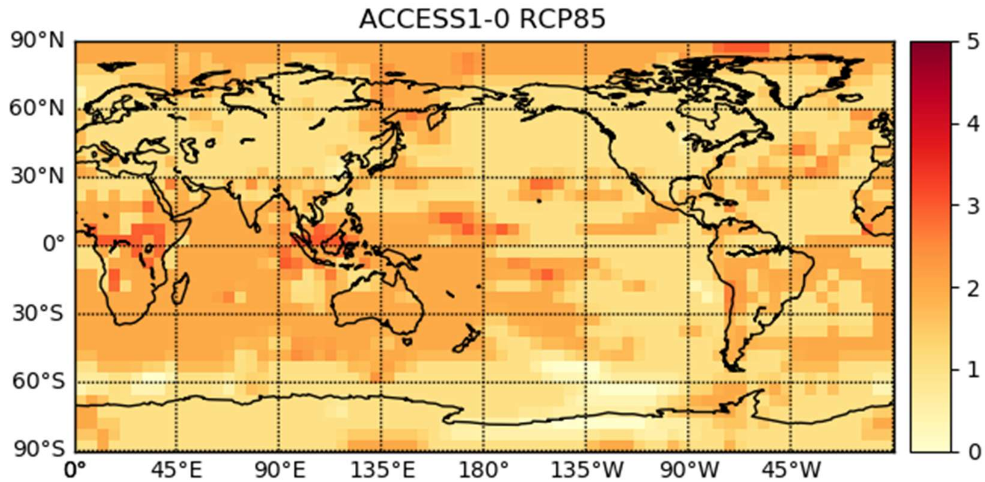

BNU-ESM RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
0
1
2
3
4
5

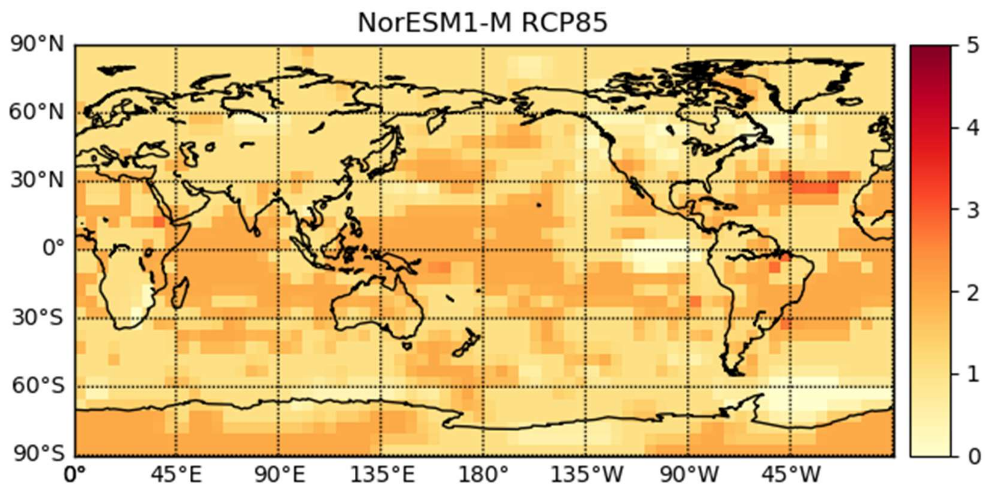

BCC RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
0
1
2
3
4
5

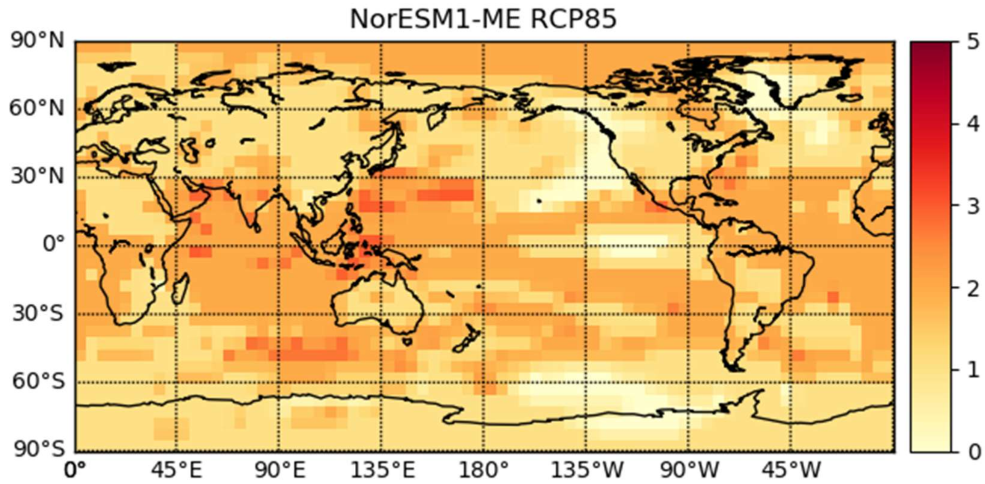

bcc-csm1-1 RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
0
1
2
3
4
5

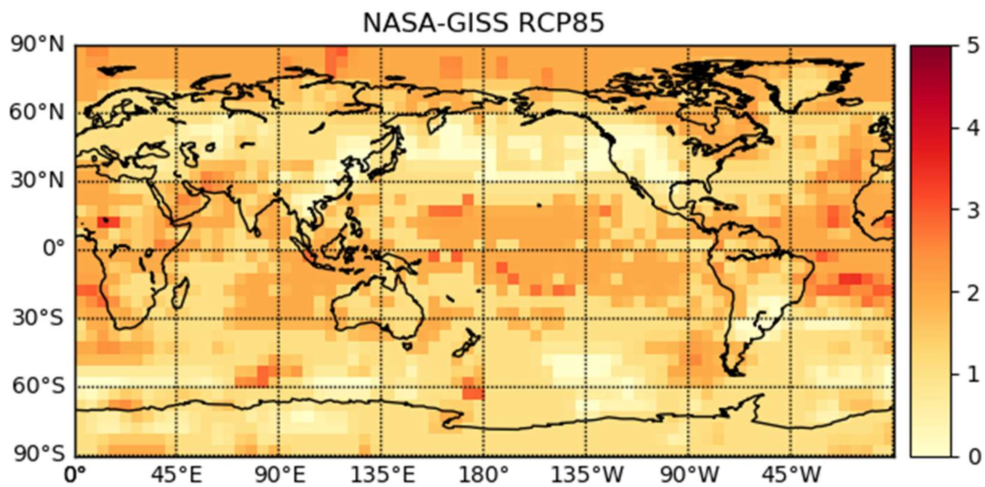

bcc-csm1-1-m RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
0
1
2
3
4
5


ACCESS1-3 RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
0
1
2
3
4
5


ACCESS1-0 RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0


NorESM1-M RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0


NorESM1-ME RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0


NASA-GISS RCP85
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
45°E
90°E
135°E
180°
135°W
90°W
45°W
5
4
3
2
1
0

References

ABRAHAM, J. P., BARINGER, M., BINDOFF, N. L., BOYER, T., CHENG, L. J., CHURCH, J. A., CONROY, J. L., DOMINGUES, C. M., FASULLO, J. T., GILSON, J., GONI, G., GOOD, S. A., GORMAN, J. M., GOURETSKI, V., ISHII, M., JOHNSON, G. C., KIZU, S., LYMAN, J. M., MACDONALD, A. M., MINKOWYCZ, W. J., MOFFITT, S. E., PALMER, M. D., PIOLA, A. R., RESEGHETTI, F., SCHUCKMANN, K., TRENBERTH, K. E., VELICOGNA, I. & WILLIS, J. K. 2013. A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics,* 51**,** 450-483.

AKAIKE, H. 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on,* 19**,** 716-723.

ALEXANDERSSON, H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology,* 6**,** 661-675.

ALEXANDERSSON, H. & MOBERG, A. 1997. Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *International Journal of climatology,* 17**,** 25-34.

ALHEIT, J., MOLLMANN, C., DUTZ, J., KORNILOVS, G., LOEWE, P., MOHRHOLZ, V. & WASMUND, N. 2005. Synchronous ecological regime shifts in the central Baltic and the North Sea in the late 1980s. *ICES Journal of Marine Science,* 62**,** 1205-1215.

ALLEN, M. R. & SMITH, L. A. 1994. Investigating the origins and significance of low-frequency modes of climate variability. *Geophysical Research Letters,* 21**,** 883-886.

ALLEY, R. B., MAROTZKE, J., NORDHAUS, W. D., OVERPECK, J. T., PETEET, D. M., PIELKE, R. A., PIERREHUMBERT, R., RHINES, P., STOCKER, T. & TALLEY, L. 2003. Abrupt climate change. *science,* 299**,** 2005-2010.

AMBAUM, M. H., HOSKINS, B. J. & STEPHENSON, D. B. 2001. Arctic oscillation or North Atlantic oscillation? *Journal of Climate,* 14**,** 3495-3507.

ANDERSEN, T., CARSTENSEN, J., HERNANDEZ-GARCIA, E. & DUARTE, C. M. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution,* 24**,** 49 - 57.

ANDREWS, D. W. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society***,** 821-856.

ANDREWS, D. W. K. & FAIR, R. C. 1988. Inference in Nonlinear Econometric Models with Structural Change. *The Review of Economic Studies,* 55**,** 615-639.

ANDREWS, T., GREGORY, J. M. & WEBB, M. J. 2015. The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate,* 28**,** 1630-1648.

ARRHENIUS, S. 1896. XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 41**,** 237-276.

BAI, J. 1997. Estimation of a change point in multiple regression models. *Review of Economics and Statistics,* 79**,** 551-563.

BARTSEV, S., BELOLIPETSKII, P. & DEGERMENDZHI, A. Multistable states in the biosphere-climate system: towards conceptual models. IOP Conference Series: Materials Science and Engineering, 2017. IOP Publishing, 012005.

BEAUGRAND, G. 2004. The North Sea regime shift: Evidence, causes, mechanisms and consequences. *Progress in Oceanography,* 60**,** 245-262.

BEAULIEU, C., CHEN, J. & SARMIENTO, J. L. 2012. Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 370**,** 1228-1249.

BEAULIEU, C. & KILLICK, R. 2018. Distinguishing trends and shifts from memory in climate data. *Journal of Climate,* 31**,** 9519-9543.

BEAULIEU, C., SEIDOU, O., OUARDA, T. B. & ZHANG, X. 2009. Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models. *Water Resources Research,* 45.

BEAULIEU, C., SEIDOU, O., OUARDA, T. B., ZHANG, X., BOULET, G. & YAGOUTI, A. 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research,* 44.

BELCU, M., STEFAN, D. S., STEFAN, M., UNTEA, I. & DANCILA, A. M. 2015. Climate change: some insights from mean surface temperature statistics. *International Journal of Sustainable Development & World Ecology,* 22**,** 437-444.

BELOLIPETSKY, P. 2014. The Shifts Hypothesis-an alternative view of global climate change. *arXiv preprint arXiv:1406.5805*.

BELOLIPETSKY, P., BARTSEV, S., IVANOVA, Y. & SALTYKOV, M. 2015. Hidden staircase signal in recent climate dynamic. *Asia-Pacific Journal of Atmospheric Sciences,* 51**,** 323-330.

BENESTAD, R. E. 2016. A mental picture of the greenhouse effect. *Theoretical and Applied Climatology***,** 1-10.

BENZI, R., PARISI, G., SUTERA, A. & VULPIANI, A. 1982. Stochastic resonance in climatic change. *Tellus,* 34**,** 10-16.

BENZI, R., SUTERA, A. & VULPIANI, A. 1981. The mechanism of stochastic resonance. *Journal of Physics A: mathematical and general,* 14**,** L453.

BINDOFF, N. L., STOTT, P. A., ACHUTARAO, K. M., ALLEN, M. R., GILLETT, N., GUTZLER, D., HANSINGO, K., HEGERL, G., HU, Y., JAIN, S., MOKHOV, I. I., OVERLAND, J., PERLWITZ, J., SEBBARI, R. & ZHANG, X. 2013. Detection and Attribution of Climate Change: from Global to Regional. *In:* STOCKER, T. F., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

BJERKNES, J. 1969. Atmospheric teleconnections from the equatorial pacific 1. *Monthly Weather Review,* 97**,** 163-172.

BOISEAU, M., GHIL, M. & JUILLET-LECLERC, A. 1999. Climatic trends and interdecadal variability from south-central Pacific coral records. *Geophysical Research Letters,* 26**,** 2881-2884.

BOND, N. A., CRONIN, M. F., FREELAND, H. & MANTUA, N. 2015. Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophysical Research Letters,* 42**,** 3414-3420.

BOUCHAREL, J., DEWITTE, B., DU PENHOAT, Y., GAREL, B., YEH, S. W. & KUG, J. S. 2011. ENSO nonlinearity in a warming climate. *Climate Dynamics,* 37**,** 2045-2065.

BOX, J. E. 2002. Survey of Greenland instrumental temperature records: 1873–2001. *International Journal of Climatology,* 22**,** 1829-1847.

BRANSTATOR, G. & SELTEN, F. 2009. “Modes of variability” and climate change. *Journal of Climate,* 22**,** 2639-2658.

BREUSCH, T. S. & PAGAN, A. R. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society***,** 1287-1294.

BROWN, R. L., DURBIN, J. & EVANS, J. M. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 149-192.

BÜCHER, A. & DESSENS, J. 1991. Secular trend of surface temperature at an elevated observatory in the Pyrenees. *Journal of Climate,* 4**,** 859-868.

BUISHAND, T. 1984. Tests for detecting a shift in the mean of hydrological time series. *Journal of Hydrology,* 73**,** 51-69.

BUISHAND, T. A. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of hydrology,* 58**,** 11-27.

BYRNE, J. P. & PERMAN, R. 2006. Unit roots and structural breaks: a survey of the literature. *Paper provided by Business School-Economics, University of Glasgow in its series Working Papers with*.

CAHILL, N., RAHMSTORF, S. & PARNELL, A. C. 2015. Change points of global temperature. *Environmental Research Letters,* 10**,** 084002.

CAI, W. & JONES, R. Response of potential evaporation to climate variability and change: what GCMs simulate. Pan evaporation: an example of the detection and attribution of trends in climate variables: proceedings of a workshop, Australian Academy of Science, Canberra, 2005. 71-78.

CAI, W. & WHETTON, P. 2001. A time-varying greenhouse warming pattern and the tropical–extratropical circulation linkage in the Pacific Ocean. *Journal of Climate,* 14**,** 3337-3355.

CALLENDAR, G. S. 1938. The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society,* 64**,** 223-240.

CARLIN, B. P., GELFAND, A. E. & SMITH, A. F. 1992. Hierarchical Bayesian analysis of changepoint problems. *Applied statistics***,** 389-405.

CARTER, B. 2006. When science fails, just use the Precautionary Principle.

CHANG, M. & LEE, R. 1974. Objective double-mass analysis. *Water resources research,* 10**,** 1123-1126.

CHANG, Y., KAUFMANN, R. K., KIM, C. S., MILLER, J. I., PARK, J. Y. & PARK, S. 2016. Time series analysis of global temperature distributions: Identifying and estimating persistent features in temperature anomalies.

CHOW, G. C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society***,** 591-605.

COGGIN, T. D. 2012. Using econometric methods to test for trends in the HadCRUT3 global and hemispheric data. *International journal of climatology,* 32**,** 315-320.

CONRAD, V. 1925. Homogenitätsbestimmung meteorologischer Beobachtungsreihen. *Meteorol. Z,* 42**,** 482-485.

CORTI, S., MOLTENI, F. & PALMER, T. 1999. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature,* 398**,** 799-802.

CRADDOCK, J. 1979. Methods of comparing annual rainfall records for climatic purposes. *Weather,* 34**,** 332-346.

DAVIS, R. A., LEE, T. C. M. & RODRIGUEZ-YAM, G. A. 2006. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association,* 101**,** 223-239.

DEGAETANO, A. T. 2006. Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate,* 19**,** 838-853.

DELVAUX, C., INGELS, R., VRÁBEĹ, V., JOURNÉE, M. & BERTRAND, C. 2019. Quality control and homogenization of the Belgian historical temperature data. *International Journal of Climatology,* 39**,** 157-171.

DELWORTH, T. L., ZENG, F., ROSATI, A., VECCHI, G. A. & WITTENBERG, A. T. 2015. A link between the hiatus in global warming and North American drought. *Journal of Climate,* 28**,** 3834-3845.

DESER, C. & BLACKMON, M. L. 1993. Surface Climate Variations over the North Atlantic Ocean during Winter: 1900–1989. *Journal of Climate,* 6**,** 1743-1753.

DEYOUNG, B., HARRIS, R., ALHEIT, J., BEAUGRAND, G., MANTUA, N. & SHANNON, L. 2004. Detecting regime shifts in the ocean: Data considerations. *Progress in Oceanography,* 60**,** 143-164.

DI LENA, B., SILVESTRONI, O., MARIANI, L., PARISI, S. & ANTENUCCI, F. European climate variability effects on grapevine harvest date time series in the Abruzzi (Italy). XXVIII International Horticultural Congress on Science and Horticulture for People (IHC2010): International Symposium on the 931, 2010. 63-69.

DICKEY, D. A. & FULLER, W. A. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057-1072.

DOMONKOS, P., VENEMA, V., AUER, I., MESTRE, O. & BRUNETTI, M. 2012. The historical pathway towards more accurate homogenisation. *Advances in Science and Research,* 8**,** 45-52.

DOOL, H. M. V. D., SAHA, S. & JOHANSSON, Å. 2000. Empirical Orthogonal Teleconnections. *Journal of Climate,* 13**,** 1421-1435.

DRIJFHOUT, S. 2018. The relation between natural variations in ocean heat uptake and global mean surface temperature anomalies in CMIP5. *Scientific Reports,* 8**,** 7402.

DRIJFHOUT, S., BATHIANY, S., BEAULIEU, C., BROVKIN, V., CLAUSSEN, M., HUNTINGFORD, C., SCHEFFER, M., SGUBIN, G. & SWINGEDOUW, D. 2015. Catalogue of abrupt shifts in Intergovernmental Panel on Climate Change climate models. *Proceedings of the National Academy of Sciences,* 112**,** 5777–5786.

DUCRÉ-ROBITAILLE, J. F., VINCENT, L. A. & BOULET, G. 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology,* 23**,** 1087-1101.

DYKE, J. G. & WEAVER, I. S. 2013. The Emergence of Environmental Homeostasis in Complex Ecosystems. *PLOS Computational Biology,* 9**,** e1003050.

EASTERLING, D. R. & PETERSON, T. C. Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review. Fifth International Meeting on Statistical Climatology, June, 1992. 22-26.

EASTERLING, D. R. & PETERSON, T. C. 1995. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology,* 15**,** 369-377.

ECKLEY, I. A., FEARNHEAD, P. & KILLICK, R. 2011. Analysis of changepoint models. *Bayesian Time Series Models***,** 205-224.

ELLIOTT, G., ROTHENBERG, T. J. & STOCK, J. H. 1992. Efficient tests for an autoregressive unit root. National Bureau of Economic Research Cambridge, Mass., USA.

EPSTEIN, E. S. 1982. Detecting climate change. *Journal of Applied Meteorology,* 21**,** 1172-1182.

ESTRADA, F. & PERRON, P. 2014. Detection and attribution of climate change through econometric methods. *Boletín de la Sociedad Matemática Mexicana,* 20**,** 107-136.

ESTRADA, F., PERRON, P., GAY-GARCÍA, C. & MARTÍNEZ-LÓPEZ, B. 2013. A Time-Series Analysis of the 20th Century Climate Simulations Produced for the IPCC's Fourth Assessment Report. *PloS one,* 8**,** e60017.

EYRING, V., RIGHI, M., LAUER, A., EVALDSSON, M., WENZEL, S., JONES, C., ANAV, A., ANDREWS, O., CIONNI, I. & DAVIN, E. L. 2016. ESMValTool (v1. 0)–a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development,* 9**,** 1747-1802.

FEARNHEAD, P. 2006. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and computing,* 16**,** 203-213.

FISCHER, J. W., WALTER, W. D. & AVERY, M. L. 2013. Brownian Bridge Movement Models to Characterize Birds' Home Ranges: Modelos de Movimiento de Puente Browniano Para Caracterizar el Rango de Hogar de las Aves. *The Condor,* 115**,** 298-305.

FLEMING, J. R. 1999. Joseph Fourier, the 'greenhouse effect', and the quest for a universal theory of terrestrial temperatures. *Endeavour,* 23**,** 72-75.

FOLLAND, C., PARKER, D. & KATES, F. 1984. Worldwide marine temperature fluctuations 1856–1981. *Nature,* 310**,** 670-673.

FOSTER, G. & ABRAHAM, J. 2015. Lack of evidence for a slowdown in global temperature. *US CLIVAR*, 6.
FOSTER, G. & RAHMSTORF, S. 2011. Global temperature evolution 1979-2010. *Environmental Research Letters,* 6, 044022.
FRANKS, S. W. 1999. Identification of a change in climate state using regional flood data. *Hydrology and Earth System Sciences,* 6, 11-16.
FREDERIKSEN, J., FREDERIKSEN, C., OSBROUGH, S. & SISSON, J. Changes in Southern Hemisphere rainfall, circulation and weather systems. 19th International Congress on Modelling and Simulation, 2011.
FREDERIKSEN, J. S. & FREDERIKSEN, C. S. 2007. Interdecadal changes in southern hemisphere winter storm track modes. *Tellus A,* 59, 599-617.
FREITAS, A. C., FREDERIKSEN, J. S., WHELAN, J., O'KANE, T. J. & AMBRIZZI, T. 2015. Observed and simulated inter-decadal changes in the structure of Southern Hemisphere large-scale circulation. *Climate Dynamics*, 1-25.
FUKAC, M. 2005. Inflation Expectations in the Czech Interbank Market.
GAN, T. Y. 1998. Hydroclimatic trends and possible climatic warming in the Canadian Prairies. *Water Resources Research,* 34, 3009-3015.
GARNAUT, R. 2008. The Garnaut climate change review. *Global Environmental change,* 13, 1-5.
GAY-GARCIA, C., ESTRADA, F. & SÁNCHEZ, A. 2009. Global and hemispheric temperatures revisited. *Climatic Change,* 94, 333-349.
GÉRARD-MARCHANT, P. G., STOOKSBURY, D. E. & SEYMOUR, L. 2008. Methods for starting the detection of undocumented multiple changepoints. *Journal of Climate,* 21, 4887-4899.
GIAVANTE, S., DI GIUSEPPE, E. & ESPOSITO, S. Flat steps models for the analysis of temperature and precipitation italian time series from 1961 to 2007. Statistical Methods for the analysis of large data-sets-Book of short papers, 2009. Cleup, 427-430.
GISSTEMP TEAM 2015. *GISS Surface Temperature Analysis (GISTEMP)*, NASA Goddard Institute for Space Studies. .
GLYNN, J., PERERA, N. & VERMA, R. 2007. Unit root tests and structural breaks: a survey with applications. *Faculty of Commerce-Papers*, 455.
GOETZ, S. J., BUNN, A. G., FISKE, G. J. & HOUGHTON, R. 2005. Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 13521-13525.
GRANGER, C. W. & MORRIS, M. J. 1976. Time series modelling and interpretation. *Journal of the Royal Statistical Society. Series A (General)*, 246-257.
GRIGGS, D. J. & NOGUER, M. 2002. Climate change 2001: the scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. *Weather,* 57, 267-269.
GU, G., ADLER, R. F. & HUFFMAN, G. J. 2016. Long-term changes/trends in surface temperature and precipitation during the satellite era (1979–2012). *Climate Dynamics,* 46, 1091-1105.
GUEMAS, V., DOBLAS-REYES, F. J., ANDREU-BURILLO, I. & ASIF, M. 2013. Retrospective prediction of the global warming slowdown in the past decade. *Nature Climate Change,* 3, 649.
GULLETT, D., VINCENT, L. & SAJECKI, P. 1990. *Testing for homogeneity in temperature time series at Canadian climate stations*, Atmospheric Environment Service, Canadian Climate Centre.
HACKER, R. S. 2010. The Effectiveness of Information Criteria in Determining Unit Root and Trend Status. Royal Institute of Technology, CESIS-Centre of Excellence for Science and Innovation Studies.

HAIG, B. D. 2016. Tests of Statistical Significance Made Sound. *Educational and Psychological Measurement***,** 0013164416667981.

HANSEN, J., FUNG, I., LACIS, A., RIND, D., LEBEDEFF, S., RUEDY, R., RUSSELL, G. & STONE, P. 1988. Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *Journal of Geophysical Research: Atmospheres (1984–2012),* 93**,** 9341-9364.

HARE, S. R. & MANTUA, N. J. 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Progress in Oceanography,* 47**,** 103-145.

HARPER, D. 2014. *Dictionary of Etymology* [Online]. Available: http://www.etymonline.com/index.php?term=climate [Accessed 27/11/2014].

HARVEY, D. I., LEYBOURNE, S. J. & TAYLOR, A. R. 2013. Testing for unit roots in the possible presence of multiple trend breaks using minimum Dickey–Fuller statistics. *Journal of Econometrics,* 177**,** 265-284.

HASSELMANN, K. 2002. Is climate predictable? *In:* BUNDE, A., KROPP, J. & SCHELLNHUBER, H. J. (eds.) *The science of disasters: climate disruptions, heart attacks, and market crashes.* Berlin Heidelbeg: Springer Science & Business Media.

HE, W., FENG, G., WU, Q., HE, T., WAN, S. & CHOU, J. 2012. A new method for abrupt dynamic change detection of correlated time series. *International Journal of climatology,* 32**,** 1604-1614.

HE, Y., HUANG, J., SHUGART, H. H., GUAN, X., WANG, B. & YU, K. 2017. Unexpected evergreen expansion in the Siberian forest under warming hiatus. *Journal of Climate,* 30**,** 5021-5039.

HEGERL, G. C., HOEGH-GULDBERG, O., CASASSA, G., HOERLING, M. P., KOVATS, R., PARMESAN, C., PIERCE, D. W. & STOTT, P. A. Good practice guidance paper on detection and attribution related to anthropogenic climate change. Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change, 2010. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland.

HENLEY, B. J. 2017. Pacific decadal climate variability: Indices, patterns and tropical-extratropical interactions. *Global and Planetary Change,* 155**,** 42-55.

HENNESSY, K., CLARKE, J. & RICKETTS, J. 2011. *Methods for producing extreme temperature projections for Australia,* Melbourne, CSIRO.

HENNESSY, K. J., C, L., N, N., BATHOLS, J. M., SUPPIAH, R. & RICKETTS, J. 2005. Climate change impacts on fire-weather in south-east Australia © CSIRO 2005. . CSIRO.

HOPE, P., DROSDOWSKY, W. & NICHOLLS, N. 2006. Shifts in the synoptic systems influencing southwest Western Australia. *Climate Dynamics,* 26**,** 751-764.

HOPE, P., TIMBAL, B. & FAWCETT, R. 2010. Associations between rainfall variability in the southwest and southeast of Australia and their evolution through time. *International Journal of Climatology,* 30**,** 1360-1371.

HOTHORN, T., ZEILEIS, A., FAREBROTHER, R. W., CUMMINS, C., MILLO, G. & MITCHELL, D. 2015. lmtest: Testing linear regression models. *R package version 0.9-34, URL https://cran. r-project. org/package= lmtest*.

HOUGHTON, J. 2009. *Global warming: the complete briefing*, Cambridge University Press.

HOUGHTON, J. T. 1990. *Climate change: The IPCC scientific assessment,* Cambridge, Great Britain, New York, NY, USA and Melbourne, Australia Cambridge University Press.

HOUGHTON, J. T. 1996. *Climate change 1995: The science of climate change: contribution of working group I to the second assessment report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

HOY, A., FESKE, N., ŠTĚPÁNEK, P., SKALÁK, P., SCHMITT, A. & SCHNEIDER, P. 2018. Climatic Changes and Their Relation to Weather Types in a Transboundary Mountainous Region in Central Europe. *Sustainability,* 10**,** 2049.

HUANG, B., BANZON, V. F., FREEMAN, E., LAWRIMORE, J., LIU, W., PETERSON, T. C., SMITH, T. M., THORNE, P. W., WOODRUFF, S. D. & ZHANG, H.-M. 2015. Extended reconstructed sea surface temperature version 4 (ERSST. v4). Part I: upgrades and intercomparisons. *Journal of climate,* 28**,** 911-930.

HURRELL, J., DELWORTH, T., DANABASOGLU, G., DRANGE, H., GRIFFIES, S., HOLBROOK, N., KIRTMAN, B., KEENLYSIDE, N., LATIF, M. & MAROTZKE, J. Decadal climate prediction: Opportunities and challenges. OceanObs' 09: Sustained Ocean Observations and Information for Society, 2010. 521-533.

HURRELL, J. W. 1995. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science,* 269**,** 676-679.

HURRELL, J. W. & DESER, C. 2010. North Atlantic climate variability: the role of the North Atlantic Oscillation. *Journal of Marine Systems,* 79**,** 231-244.

HURRELL, J. W. & VAN LOON, H. 1997. Decadal variations in climate associated with the North Atlantic Oscillation. *Climatic change at high elevation sites.* Springer.

HUŠKOVÁ, M. & KIRCH, C. 2008. Bootstrapping confidence intervals for the change-point of time series. *Journal of Time Series Analysis,* 29**,** 947-972.

IPCC 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press.

IPCC 2014a. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)],* Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press.

IPCC 2014b. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)],* Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press.

ISAACS, J. D. 1976. Some ideas and frustrations about fishery science. *Calif. Coop. Oceanic Fish. Invest Rep,* 18**,** 34-43.

JARUŠKOVÁ, D. 1996. Change-point detection in meteorological measurement. *Monthly Weather Review,* 124**,** 1535-1543.

JARUŠKOVÁ, D. 1997. Some problems with application of change-point detection methods to environmental data. *Environmetrics,* 8**,** 469-484.

JONES, G. S., GREGORY, J. M., STOTT, P. A., TETT, S. F. B. & THORPE, R. B. 2005. An AOGCM simulation of the climate response to a volcanic super-eruption. *Climate Dynamics,* 25**,** 725-738.

JONES, R. North central Victorian climate: Past, present and future. Proc. R. Soc. Vic, 2010. 147-160.

JONES, R., HENNESSY, K. J. & ABBS, D. J. 1999. *Climate change analysis relevant to Jabiluka*, Environment Australia.

JONES, R., MCMAHON, T. & BOWLER, J. 2001. Modelling historical lake levels and recent climate change at three closed lakes, Western Victoria, Australia (c. 1840–1990). *Journal of Hydrology,* 246**,** 159-180.

JONES, R. & PAGE, C. Assessing the risk of climate change on the water resources of the Macquarie River Catchment. Integrating models for natural resources management across disciplines, issues and scales, 2001. 673-678.

JONES, R., YOUNG, C., HANDMER, J., KEATING, A., MEKALA, G. & SHEEHAN, P. 2013. Valuing adaptation under rapid change. National Climate Change Adaptation Research Facility, Gold Coast, Australia.

JONES, R. N. 1995. *Modelling hydrologic and climatic controls of closed lakes, western Victoria.* University of Melbourne.

JONES, R. N. 2012. Detecting and attributing nonlinear anthropogenic regional warming in southeastern Australia. *Journal of Geophysical Research: Atmospheres (1984--2012),* 117.

JONES, R. N. & RICKETTS, J. H. 2016a. Atmospheric warming 1997–2014: hiatus, pause or regime? Climate Change Working Paper No. 39. Victoria University, Melbourne: Victoria Institute of Strategic Economic Studies.

JONES, R. N. & RICKETTS, J. H. 2016b. The climate wars and "the pause" – are both sides wrong? Climate Change Working Paper No. 37. Victoria University, Melbourne: Victoria Institute of Strategic Economic Studies.

JONES, R. N. & RICKETTS, J. H. 2017a. Has the step-change trigger for global warming been found? . *In preparation.*

JONES, R. N. & RICKETTS, J. H. 2017b. Reconciling the signal and noise of atmospheric warming on decadal timescales. *Earth Syst. Dynam.,* 8**,** 177-210.

JONES, R. N. & RICKETTS, J. H. 2019. The Pacific Ocean heat engine: global climate's regulator. *Earth System Dynamics (for open review)*.

KAISER, R. & MARAVALL, A. 1999. Seasonal outliers in time series.

KATZAV, J. 2011. Should we assess climate model predictions in light of severe tests? *EOS, Transactions American Geophysical Union,* 92**,** 195-195.

KATZAV, J. 2013. Severe testing of climate change hypotheses. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* 44**,** 433-441.

KATZAV, J., DIJKSTRA, H. A. & DE LAAT, A. T. J. 2012. Assessing climate model projections: State of the art and philosophical reflections. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* 43**,** 258-276.

KAUFMANN, R., KAUPPI, H. & STOCK, J. 2006. Emissions, Concentrations, & Temperature: A Time Series Analysis. *Climatic Change,* 77**,** 249-278.

KAUFMANN, R. K. & STERN, D. I. 1997. Evidence for human influence on climate from hemispheric temperature relations. *Nature,* 388**,** 39.

KEJRIWAL, M. & PERRON, P. 2010. A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. *Journal of Time Series Analysis,* 31**,** 305-328.

KILLICK, R. 2012. *Novel methods for changepoint problems.* Ph.D, Lancaster University.

KILLICK, R. & ECKLEY, I. 2014. changepoint: a comprehensive changepoint analysis package for R. *Journal of Statistical Software*.

KILLICK, R. & ECKLEY, I. A. 2011. Changepoint: an R package for changepoint analysis. *R package version 0.6, URL http://CRAN. R-project. org/package= changepoint*.

KILLICK, R., ECKLEY, I. A., EWANS, K. & JONATHAN, P. 2010. Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Engineering,* 37**,** 1120-1126.

KILLICK, R., FEARNHEAD, P. & ECKLEY, I. 2012. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association,* 107**,** 1590-1598.

KIRBY, M., LUND, S., PATTERSON, W., ANDERSON, M., BIRD, B., IVANOVICI, L., MONARREZ, P. & NIELSEN, S. 2010. A Holocene record of Pacific decadal oscillation (PDO)-related hydrologic variability in southern California (Lake Elsinore, CA). *Journal of Paleolimnology,* 44**,** 819-839.

KIRONO, D. G., JONES, R., KENT, D. & LEAHY, P. Modelling lake levels under climate change conditions: three closed lakes in Western Victoria. 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, 2009. 4312-4318.

KIRONO, D. G. & JONES, R. N. 2007. A bivariate test for detecting inhomogeneities in pan evaporation time series. *Australian Meteorological Magazine,* 56**,** 93-103.

KIRTMAN, B., POWER, S. B., ADEDOYIN, J. A., BOER, G. J., BOJARIU, R., CAMILLONI, I., DOBLAS-REYES, F. J., FIORE, A. M., KIMOTO, M., MEEHL, G. A., PRATHER, M., SARR, A., SCHÄR, C., SUTTON, R., VAN OLDENBORGH, G. J., VECCHI, G. & WANG, H. J. 2013. Near-term Climate Change: Projections and Predictability. *In:* STOCKER, T. F., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

KNUDSEN, M. F., SEIDENKRANTZ, M.-S., JACOBSEN, B. H. & KUIJPERS, A. 2011. Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nature communications,* 2**,** 178.

KOČENDA, E. & ČERNÝ, A. 2015. *Elements of time series econometrics: An applied approach*, Charles University in Prague, Karolinum Press.

KÖHLER, M. 1949. On the use of double mass analysis for testing the consistency of meteorological records and for making required adjustments. *Bull. Am. Meteorol. Soc,* 30**,** 188-189.

KOSAKA, Y. & XIE, S.-P. 2013. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature,* 501**,** 403-407.

KOSAKA, Y. & XIE, S.-P. 2016. The tropical Pacific as a key pacemaker of the variable rates of global warming. *Nature Geoscience,* 9**,** 669.

KREIL, K. 1848. Mehrjährige Beobachtungen in Wien vom Jahre 1775 bis 1850, Jahrbücher der kk Central-Anstalt für Meteorologie und Erdmagnetismus, I. Band–Jg.

KWIATKOWSKI, D., PHILLIPS, P. C., SCHMIDT, P. & SHIN, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics,* 54**,** 159-178.

LABITZKE, K. & NAUJOKAT, B. 1984. On the effect of the volcanic eruptions of Mount Agung and El Chichón on the temperature of the stratosphere. *Geofisica Internacional,* 23.

LEE, S.-K., PARK, W., BARINGER, M. O., GORDON, A. L., HUBER, B. & LIU, Y. 2015. Pacific origin of the abrupt increase in Indian Ocean heat content during the warming hiatus. *Nature Geoscience,* 8**,** 445.

LENTON, T. M., HELD, H., KRIEGLER, E., HALL, J. W., LUCHT, W., RAHMSTORF, S. & SCHELLNHUBER, H. J. 2008. Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences,* 105**,** 1786-1793.

LEPAGE, Y. 1971. A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika,* 58**,** 213-217.

LEVITUS, S., ANTONOV, J., BOYER, T., BARANOVA, O., GARCIA, H., LOCARNINI, R., MISHONOV, A., REAGAN, J., SEIDOV, D. & YAROSH, E. 2010. World Ocean Heat Content and Thermosteric Sea Level change (0-2000 m), 1955.

LEVITUS, S., JI, A., OK, B., TP, B., CL, C., HE, G., AI, G., RA, L., AV, M. & CL, S. 2013. The world ocean database. *Data Science Journal,* 12**,** WDS229-WDS234.

LEWANDOWSKY, S., RISBEY, J. S. & ORESKES, N. 2015. The “Pause” in Global Warming: Turning a Routine Fluctuation into a Problem for Science. *Bulletin of the American Meteorological Society*.

LIDDLE, B. & MESSINIS, G. 2015. Revisiting sulfur Kuznets curves with endogenous breaks modeling: Substantial evidence of inverted-Us/Vs for individual OECD countries. *Economic Modelling,* 49**,** 278-285.

LIU, J., WU, S. & ZIDEK, J. V. 1997. On segmented multivariate regression. *Statistica Sinica***,** 497-525.

LIU, W., HUANG, B., THORNE, P. W., BANZON, V. F., ZHANG, H.-M., FREEMAN, E., LAWRIMORE, J., PETERSON, T. C., SMITH, T. M. & WOODRUFF, S. D. 2015. Extended reconstructed sea surface temperature version 4 (ERSST. v4): Part II. Parametric and structural uncertainty estimations. *Journal of Climate,* 28**,** 931-951.

LLOYD, J., MOROS, M., PERNER, K., TELFORD, R. J., KUIJPERS, A., JANSEN, E. & MCCARTHY, D. 2011. A 100 yr record of ocean temperature control on the stability of Jakobshavn Isbrae, West Greenland. *Geology,* 39**,** 867-870.

LO, T. T. & HSU, H. H. 2010. Change in the dominant decadal patterns and the late 1980s abrupt warming in the extratropical Northern Hemisphere. *Atmospheric Science Letters,* 11**,** 210-215.

LU, J., CHEN, G. & FRIERSON, D. M. 2008. Response of the zonal mean atmospheric circulation to El Niño versus global warming. *Journal of Climate,* 21**,** 5835-5851.

LUDDEN, T. M., BEAL, S. L. & SHEINER, L. B. 1994. Comparison of the Akaike Information Criterion, the Schwarz criterion and the F test as guides to model selection. *Journal of pharmacokinetics and biopharmaceutics,* 22**,** 431-445.

LUMSDAINE, R. L. & PAPELL, D. H. 1997. Multiple trend breaks and the unit-root hypothesis. *Review of economics and Statistics,* 79**,** 212-218.

LYNAM, C., LILLEY, M., BASTIAN, T., DOYLE, T., BEGGS, S. & HAYS, G. 2011. Have jellyfish in the Irish Sea benefited from climate change and overfishing? *Global Change Biology,* 17**,** 767-782.

MAHER, N., GUPTA, A. S. & ENGLAND, M. H. 2014. Drivers of decadal hiatus periods in the 20th and 21st centuries. *Geophysical Research Letters,* 41**,** 5978-5986.

MAHMOOD, R. & JIA, S. 2017. Spatial and temporal hydro-climatic trends in the transboundary Jhelum River basin. *Journal of Water and Climate Change,* 8**,** 423-440.

MANABE, S. & STRICKLER, R. F. 1964. Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of the Atmospheric Sciences,* 21**,** 361-385.

MANN, H. B. & WHITNEY, D. R. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics,* 18**,** 50-60.

MANN, M. E., STEINMAN, B. A. & MILLER, S. K. 2014. On forced temperature changes, internal variability, and the AMO. *Geophysical Research Letters,* 41**,** 3211-3219.

MANTUA, N. 2004. Methods for detecting regime shifts in large marine ecosystems: a review with approaches applied to North Pacific data. *Progress in Oceanography,* 60**,** 165-182.

MANTUA, N. J., HARE, S. R., ZHANG, Y., WALLACE, J. M. & FRANCIS, R. C. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society,* 78**,** 1069-1079.

MARONNA, R. & YOHAI, V. J. 1978. A bivariate test for the detection of a systematic change in mean. *Journal of the American Statistical Association,* 73**,** 640-645.

MARTÍNEZ, J. M. P., MEDINA-ELIZALDE, M., BURNS, S. J., JIANG, X. & SHEN, C.-C. 2015. Climate regime shifts in paleoclimate time series from the Yucatán Peninsula: from the Preclassic to Classic period.

MAYO, D. G. 1996. *Error and the growth of experimental knowledge*, University of Chicago Press.

MAYO, D. G. 2004. An error-statistical philosophy of evidence. *The nature of scientific evidence: Statistical, philosophical and empirical considerations***,** 79-96.

MAYO, D. G. & COX, D. R. 2006. Frequentist statistics as a theory of inductive inference. *Lecture Notes-Monograph Series***,** 77-97.

MAYO, D. G. & SPANOS, A. 2004. Methodology in Practice: Statistical Misspecification Testing. *Philosophy of Science,* 71**,** 1007-1025.

MAYO, D. G. & SPANOS, A. 2006. Severe Testing as a Basic Concept in a Neyman–Pearson Philosophy of Induction. *The British Journal for the Philosophy of Science,* 57**,** 323-357.

MAYO, D. G. & SPANOS, A. 2011. Error statistics. *Handbook of the philosophy of science,* 7**,** 153-198.

MCCARTHY, G. D., HAIGH, I. D., HIRSCHI, J. J. M., GRIST, J. P. & SMEED, D. A. 2015. Ocean impact on decadal Atlantic climate variability revealed by sea-level observations. *Nature,* 521**,** 508-510.

MCDONNELL, M. D. & ABBOTT, D. 2009. What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. *PLoS Computational Biology,* 5**,** e1000348.

MCDOWALL, D. 1980. *Interrupted time series analysis*, Sage.

MCFARLANE, G. A., KING, J. R. & BEAMISH, R. J. 2000. Have there been recent changes in climate? Ask the fish. *Progress in Oceanography,* 47**,** 147-169.

MCGREGOR, S., TIMMERMANN, A. & TIMM, O. 2010. A unified proxy for ENSO and PDO variability since 1650. *Climate of the Past,* 6**,** 1-17.

MEEHL, G. A., ARBLASTER, J. M., FASULLO, J. T., HU, A. & TRENBERTH, K. E. 2011. Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change,* 1**,** 360-364.

MEEHL, G. A., HU, A. & SANTER, B. D. 2009. The mid-1970s climate shift in the Pacific and the relative roles of forced versus inherent decadal variability. *Journal of Climate,* 22**,** 780-792.

MEEHL, G. A., TENG, H. & ARBLASTER, J. M. 2014. Climate model simulations of the observed early-2000s hiatus of global warming. *Nature Clim. Change,* 4**,** 898-902.

MEINSHAUSEN, M., SMITH, S. J., CALVIN, K., DANIEL, J. S., KAINUMA, M., LAMARQUE, J., MATSUMOTO, K., MONTZKA, S., RAPER, S. & RIAHI, K. 2011. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change,* 109**,** 213-241.

MILLS, B. J., SCOTESE, C. R., WALDING, N. G., SHIELDS, G. A. & LENTON, T. M. 2017. Elevated CO 2 degassing rates prevented the return of Snowball Earth during the Phanerozoic. *Nature communications,* 8**,** 1110.

MINOBE, S. 1997. A 50–70 year climatic oscillation over the North Pacific and North America. *Geophysical Research Letters,* 24**,** 683-686.

MINOBE, S. 2002. Interannual to interdecadal changes in the Bering Sea and concurrent 1998/99 changes over the North Pacific. *Progress in Oceanography,* 55**,** 45-64.

MITCHELL, J. F. B., JOHNS, T. C., EAGLES, M., INGRAM, W. J. & DAVIS, R. A. 1999. Towards the Construction of Climate Change Scenarios. *Climatic Change,* 41**,** 547-581.

MITCHELL, T. D. 2003. Pattern Scaling. An Examination of the Accuracy of the Technique for Describing Future Climates. *Climatic Change,* 60**,** 217-242.

MIZON, G. E. 1995. A simple message for autocorrelation correctors: Don't. *Journal of Econometrics,* 69**,** 267-288.

MOBERG, A. & ALEXANDERSSON, H. 1997. Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology,* 17**,** 35-54.

MOORE, G., HALFAR, J., MAJEED, H., ADEY, W. & KRONZ, A. 2017. Amplification of the Atlantic Multidecadal Oscillation associated with the onset of the industrial-era warming. *Scientific reports,* 7**,** 40861.

MOSLEY-THOMPSON, E., READINGER, C., CRAIGMILE, P., THOMPSON, L. & CALDER, C. 2005. Regional sensitivity of Greenland precipitation to NAO variability. *Geophysical Research Letters,* 32.

MOSS, F. & WIESENFELD, K. 1995. The benefits of background noise. *Scientific American,* 273**,** 66-69.
NAKAMURA, M. 2013. Greenland Sea Surface Temperature Change and Accompanying Changes in the Northern Hemispheric Climate. *Journal of Climate,* 26**,** 8576-8596.
NELSON, C. R. & PLOSSER, C. R. 1982. Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics,* 10**,** 139-162.
NEWEY, W. K. & WEST, K. D. 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies,* 61**,** 631-653.
NEWMAN, M., ALEXANDER, M. A., AULT, T. R., COBB, K. M., DESER, C., LORENZO, E. D., MANTUA, N. J., MILLER, A. J., MINOBE, S., NAKAMURA, H., SCHNEIDER, N., VIMONT, D. J., PHILLIPS, A. S., SCOTT, J. D. & SMITH, C. A. 2016. The Pacific Decadal Oscillation, Revisited. *Journal of Climate,* 29**,** 4399-4427.
NICHOLLS, N., DELLA-MARTA, P. & COLLINS, D. 2004. 20 th century changes in temperature and rainfall in New South Wales. *Australian Meteorological Magazine,* 53**,** 263-268.
NORTH, G. R., KIM, K.-Y., SHEN, S. S. P. & HARDIN, J. W. 1995. Detection of Forced Climate Signals. Part 1: Filter Theory. *Journal of Climate,* 8**,** 401-408.
O'KANE, T., MATEAR, R., CHAMBERLAIN, M. & OKE, P. 2013. ENSO regimes and the late 1970's climate regime shift: The role of synoptic weather and South Pacific ocean spiciness. *Journal of Computational Physics*.
OVERLAND, J., RODIONOV, S., MINOBE, S. & BOND, N. 2008. North Pacific regime shifts: Definitions, issues and recent transitions. *Progress in Oceanography,* 77**,** 92-102.
PACHAURI, R. & MEYER, L. 2014. Climate change 2014: Synthesis report. fifth assessment report of the intergovernmental panel on climate change. *Tech. Rep.*
PAGE, E. 1954. Continuous inspection schemes. *Biometrika,* 41**,** 100-115.
PARKER, J. A. 2018. Fundamental Concepts of Time-Series Econometrics. *Theory and Practice of Econometrics Spring 2018.* Reed College.
PEARSON, E. S. 1955. Statistical concepts in their relation to reality. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 204-207.
PERCIVAL, D. B., OVERLAND, J. E. & MOFJELD, H. O. 2001. Interpretation of North Pacific variability as a short-and long-memory process. *Journal of Climate,* 14**,** 4545-4559.
PERNEGER, T. V. 1998. What's wrong with Bonferroni adjustments. *British Medical Journal,* 316**,** 1236.
PERREAULT, L., BERNIER, J., BOBÉE, B. & PARENT, E. 2000. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology,* 235**,** 221-241.
PERRON, P. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society***,** 1361-1401.
PERRON, P. & YABU, T. 2009. Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics,* 151**,** 56-69.
PETERSON, T. C. & EASTERLING, D. R. 1994. Creation of homogeneous composite climatological reference series. *International journal of climatology,* 14**,** 671-679.
PETTITT, A. 1979. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 28**,** 126-135.
PEYSER, C. E., YIN, J., LANDERER, F. W. & COLE, J. E. 2016. Pacific sea level rise patterns and global surface temperature variability. *Geophysical Research Letters*.
PFAFF, B., ZIVOT, E. & STIGLER, M. 2016. Unit Root and Cointegration Tests for Time Series Data.
PIERCE, D. W. 2001. Distinguishing coupled ocean–atmosphere interactions from background noise in the North Pacific. *Progress in Oceanography,* 49**,** 331-352.
PLASS, G. N. 1956. The Carbon Dioxide Theory of Climatic Change. *Tellus,* 8**,** 140-154.

POTTER, K. W. 1981. Illustration of a New Test for Detecting a Shift in Mean in Precipitation Series. *Mon. Wea. Rev.,* 109**,** 2040-2045.
RAHMSTORF, S. 2002. Ocean circulation and climate during the past 120,000 years. *Nature,* 419**,** 207.
RAHMSTORF, S., FOSTER, G. & CAHILL, N. 2017. Global temperature evolution: recent trends and some pitfalls. *Environmental Research Letters,* 12**,** 054001.
RAHMSTORF, S., FOSTER, G. & CAZENAVE, A. 2012. Comparing climate projections to observations up to 2011. *Environmental Research Letters,* 7**,** 044035.
RAJARATNAM, B., ROMANO, J., TSIANG, M. & DIFFENBAUGH, N. 2015. Debunking the climate hiatus. *Climatic Change***,** 1-12.
RAO, C., RAY, A., SARKAR, S. & YASAR, M. 2009. Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing,* 3**,** 101-114.
REASON, C., LANDMAN, W. & TENNANT, W. 2006. Seasonal to decadal prediction of southern African climate and its links with variability of the Atlantic Ocean. *Bulletin of the American Meteorological Society,* 87**,** 941-956.
REEVES, J., CHEN, J., WANG, X. L., LUND, R. & LU, Q. 2007. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology \& Climatology,* 46.
REID, P. C. 2016. Ocean warming: setting the scene. *Explaining ocean warming: Causes, scale, effects and consequences.*
REID, P. C. & BEAUGRAND, G. 2012. Global synchrony of an accelerating rise in sea surface temperature. *Journal of the Marine Biological Association of the United Kingdom,* 92**,** 1435-1450.
REID, P. C., HARI, R. E., BEAUGRAND, G., LIVINGSTONE, D. M., MARTY, C., STRAILE, D., BARICHIVICH, J., GOBERVILLE, E., ADRIAN, R. & AONO, Y. 2015. Global impacts of the 1980s regime shift. *Global change biology*.
RICHARD, Y., TRZASKA, S., ROUCOU, P. & ROUAULT, M. 2000. Modification of the southern African rainfall variability/ENSO relationship since the late 1960s. *Climate Dynamics,* 16**,** 883-895.
RICKETTS, J., KOKIC, P. & CARTER, J. Consistent Climate Scenarios: projecting representative future daily climate from global climate models based on historical climate data. *In:* PIANTADOSI, J., ANDERSSEN, R. S. & BOLAND, J. E., eds. MODSIM2013, 20th International Congress on Modelling and Simulation, 2013 Adelaide, Australia. Modelling and Simulation Society of Australia and New Zealand.
RICKETTS, J. H. OzClim for the MTSRF region. The 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation., 13-17 July 2009 2009 Cairns, Australia. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, 2049-2055.
RICKETTS, J. H. Using genetic programming for symbolic regression to detect climate change signatures. *In:* PIANTADOSI, J., ANDERSSEN, R. S. & BOLAND, J. E., eds. MODSIM2013, 20th International Congress on Modelling and Simulation, December 1-6 2013 Adelaide, Australia. Modelling and Simulation Society of Australia and New Zealand.
RICKETTS, J. H. A probabilistic approach to climate shift detection based on the Maronna-Yohai bivariate test. *In:* WEBER, T., MCPHEE, M.J. AND ANDERSSEN, R.S. , ed. MODSIM2015, The 21st International Congress on Modelling and Simulation, December 2015 2015a Gold Coast, Queensland, Australia. Modelling and Simulation Society of Australia and New Zealand, 1310-1316.
RICKETTS, J. H. Stepwise symbolic regression compared to a probabilistic bivariate test for step-change detection. (Invited paper). *In:* WEBER, T., MCPHEE, M.J. AND ANDERSSEN, R.S. , ed. MODSIM2015, The 21st International Congress on Modelling and Simulation,

December 2015 2015b Gold Coast, Queensland, Australia Modelling and Simulation Society of Australia and New Zealand, 592-598.
RICKETTS, J. H. & JONES, R. N. Characterizing change-points in climate series with a severe approach. The 22nd International Congress on Modelling and Simulation (MODSIM2017), 3-8 December 2017 2017 Hobart. The Modelling and Simulation Society of Australia and New Zealand Inc.
RICKETTS, J. H. & PAGE, C. M. A Web Based Version of OzClim for Exploring Climate Change Impacts and Risks in the Australian Region. MODSIM 2007, December 2007 2007 Christchurch, New Zealand. 560-566.
RISBEY, J. S., LEWANDOWSKY, S., COWTAN, K., ORESKES, N., RAHMSTORF, S., JOKIMÄKI, A. & FOSTER, G. 2018. A fluctuation in surface temperature in historical context: reassessment and retrospective on the evidence. *Environmental Research Letters,* 13**,** 123008.
RISBEY, J. S., LEWANDOWSKY, S., LANGLAIS, C., MONSELESAN, D. P., O'KANE, T. J. & ORESKES, N. 2014. Well-estimated global surface warming in climate projections selected for ENSO phase. *Nature Clim. Change,* 4**,** 835-840.
RODIONOV, S. 2005. A brief overview of the regime shift detection methods. *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability***,** 17-24.
RODIONOV, S. & OVERLAND, J. E. 2005. Application of a sequential regime shift detection method to the Bering Sea ecosystem. *ICES Journal of Marine Science: Journal du Conseil,* 62**,** 328-332.
RODIONOV, S. N. 2004. A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters,* 31.
RODIONOV, S. N. 2006a. The problem of red noise in climate regime shift detection. *Geophysical Research Letters,* 31**,** L12707.
RODIONOV, S. N. 2006b. Use of prewhitening in climate regime shift detection. *Geophysical Research Letters,* 33.
ROEMMICH, D., GILSON, J., DAVIS, R., SUTTON, P., WIJFFELS, S. & RISER, S. 2007. Decadal Spinup of the South Pacific Subtropical Gyre. *Journal of Physical Oceanography,* 37**,** 162-173.
ROUSH, S. 2010. *Optimism about the pessimistic induction*, na.
RUDNICK, D. L. & DAVIS, R. E. 2003. Red noise and regime shifts. *Deep-Sea Research Part I: Oceanographic Research Papers,* 50**,** 691-699.
SAHIN, S. & CIGIZOGLU, H. K. 2010. Homogeneity analysis of Turkish meteorological data set. *Hydrological Processes,* 24**,** 981-992.
SALTYKOV, M., BELOLIPETSKY, P., HARI, R. E., REID, P. C. & BARTSEV, S. 2017. Synchronous shifts in outgoing longwave radiation and their interpretation. *In:* LEKKAS, D. F. (ed.) *Proceedings of the 15th International Conference on Environmental Science and Technology.* Rhodes, Greece: University of the Aegean.
SANTER, B., MEARS, C., DOUTRIAUX, C., CALDWELL, P., GLECKLER, P., WIGLEY, T., SOLOMON, S., GILLETT, N., IVANOVA, D. & KARL, T. 2011. Separating signal and noise in atmospheric temperature changes: The importance of timescale. *Journal of Geophysical Research (Atmospheres),* 116**,** 22105.
SANTER, B. D., WIGLEY, T. M. L., SCHLESINGER, M. E., B., J. F. & MITCHELL 1990. Developing Climate Scenarios from Equilibrium GCM Results.
SCHLESINGER, M. E. & RAMANKUTTY, N. 1994. An oscillation in the global climate system of period 65-70 years. *Nature***,** 723-726.
SCHMIDT, G. A., SHINDELL, D. T. & TSIGARIDIS, K. 2014. Reconciling warming trends. *Nature Geoscience,* 7**,** 158-160.

SCHNEIDER, S. H. 2004. Abrupt non-linear climate change, irreversibility and surprise. *Global Environmental Change,* 14**,** 245-258.

SEIDEL, D. J. & LANZANTE, J. R. 2004. An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. *Journal of Geophysical Research: Atmospheres,* 109**,** n/a-n/a.

SEIDOU, O. & OUARDA, T. B. 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research,* 43.

SGUBIN, G., SWINGEDOUW, D., DRIJFHOUT, S., MARY, Y. & BENNABI, A. 2017. Abrupt cooling over the North Atlantic in modern climate models. *Nature communications,* 8**,** 14375.

SHEN, C., WANG, W. C., GONG, W. & HAO, Z. 2006. A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. *Geophysical Research Letters,* 33.

SILLMANN, J., DONAT, M. G., FYFE, J. C. & ZWIERS, F. W. 2014. Observed and simulated temperature extremes during the recent warming hiatus. *Environmental Research Letters,* 9**,** 064023.

SMITH, S. W. 1997. *The Scientist and Engineer's Guide to Digital Signal Processing,* California, California Technical Publishing San Diego.

SODEN, B. J., WETHERALD, R. T., STENCHIKOV, G. L. & ROBOCK, A. 2002. Global Cooling After the Eruption of Mount Pinatubo: A Test of Climate Feedback by Water Vapor. *Science,* 296**,** 727.

SOLOMON, S. 2007. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, Cambridge University Press.

SOLOW, A. R. 1987. Testing for climate change: An application of the two-phase regression model. *Journal of Climate and Applied Meteorology,* 26**,** 1401-1405.

SPANOS, A. 2010. Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification. *Journal of Econometrics,* 158**,** 204-220.

STERN, D. I. & KAUFMANN, R. K. 1997. Time series properties of global climate variables: detection and attribution of climate change.

STIPS, A. & LILOVER, M. Global climate change: Did we pass a tipping point? Baltic International Symposium (BALTIC), 2012 IEEE/OES, 2012. IEEE, 1-6.

STOCK, J. H. 1994. Unit roots, structural breaks and trends. *Handbook of econometrics.*

STOCKER, T. F., QIN, D., PLATTNER, G.-K., ALEXANDER, L. V., ALLEN, S. K., BINDOFF, N. L., BRÉON, F.-M., CHURCH, J. A., CUBASCH, U., EMORI, S., FORSTER, P., FRIEDLINGSTEIN, P., GILLETT, N., GREGORY, J. M., HARTMANN, D. L., JANSEN, E., KIRTMAN, B., KNUTTI, R., KRISHNA KUMAR, K., LEMKE, P., MAROTZKE, J., MASSON-DELMOTTE, V., MEEHL, G. A., MOKHOV, I. I., PIAO, S., RAMASWAMY, V., RANDALL, D., RHEIN, M., ROJAS, M., SABINE, C., SHINDELL, D., TALLEY, L. D., VAUGHAN, D. G. & XIE, S.-P. 2013a. Technical Summary. *In:* STOCKER, T. F., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

STOCKER, T. F., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. 2013b. Climate Change 2013. The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change-Abstract for decision-makers. Groupe d'experts intergouvernemental sur l'evolution du climat/Intergovernmental Panel on Climate Change-IPCC, C/O World Meteorological Organization, 7bis Avenue de la Paix, CP 2300 CH-1211 Geneva 2 (Switzerland).

STOCKER, T. F., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. 2013c. Climate change 2013: The physical science

basis. *Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)(Cambridge Univ Press, New York)*.
STOCKWELL, D. R. B. & COX, A. 2009. Structural break models of climatic regime-shifts: claims and forecasts. *ArXiv e-prints*.
STOER, J. & BULIRSCH, R. 2013. *Introduction to numerical analysis*, Springer Science & Business Media.
STOTT, P. A., GILLETT, N. P., HEGERL, G. C., KAROLY, D. J., STONE, D. A., ZHANG, X. & ZWIERS, F. 2010. Detection and attribution of climate change: a regional perspective. *Wiley Interdisciplinary Reviews: Climate Change,* 1**,** 192-211.
SWANSON, K. L. & TSONIS, A. A. 2009. Has the climate recently shifted? *Geophysical Research Letters,* 36.
TABACHNICK, B. G. & FIDELL, L. S. 2007. *Using multivariate statistics*, Allyn & Bacon/Pearson Education.
TAYANÇ, M., DALFES, H. N., KARACA, M. & YENIGÜN, O. 1998. A Comparative Assessment of Different Methods for Detecting Inhomogeneities in a Turkish Temperature Data Set. *International Journal of Climatology,* 18**,** 561-578.
TAYLOR, W. 1997. Change-point analyzer. Taylor Enterprises, Inc.
THOMPSON, D. W. & WALLACE, J. M. 1998. The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research letters,* 25**,** 1297-1300.
THOMPSON, D. W. J., WALLACE, J. M., KENNEDY, J. J. & JONES, P. D. 2010. An abrupt drop in Northern Hemisphere sea surface temperature around 1970. *Nature,* 467**,** 444-447.
THOMSON, R. E. & EMERY, W. J. 2014. *Data analysis methods in physical oceanography - Third Edition*, Elsevier.
TIMMRECK, C. 2011. *Limited Climate Response of Very Large Volcanic Eruptions* [Online]. Max Planck Institute for Meteorology. Available: https://www.mpimet.mpg.de/en/communication/news/focus-on-overview/climate-response-of-volcanic-eruptions/ [Accessed].
TOMASI, D., JONES, G. V., GIUST, M., LOVAT, L. & GAIOTTI, F. 2011. Grapevine Phenology and Climate Change: Relationships and Trends in the Veneto Region of Italy for 1964–2009. *American Journal of Enology and Viticulture*.
TOMÉ, A. & MIRANDA, P. 2004. Piecewise linear fitting and trend changing points of climate parameters. *Geophysical Research Letters,* 31.
TRAPLETTI, A., HORNICK, K. & LEBARON, B. 2017. Time series analysis and computational finance.
TRENARY, L. & DELSOLE, T. 2016. Does the Atlantic Multidecadal Oscillation Get Its Predictability from the Atlantic Meridional Overturning Circulation? *Journal of Climate,* 29**,** 5267-5280.
TRENBERTH, K. E. 1990. Recent Observed Interdecadal Climate Changes in the Northern Hemisphere. *Bulletin of the American Meteorological Society,* 71**,** 988-993.
TRENBERTH, K. E. 2015. Has there been a hiatus? *Science,* 349**,** 691-692.
TRENBERTH, K. E. & CARON, J. M. 2001. Estimates of Meridional Atmosphere and Ocean Heat Transports. *Journal of Climate,* 14**,** 3433-3443.
TRENBERTH, K. E. & FASULLO, J. T. 2013. An apparent hiatus in global warming? *Earth's Future*.
TRENBERTH, K. E., FASULLO, J. T., BRANSTATOR, G. & PHILLIPS, A. S. 2014. Seasonal aspects of the recent pause in surface warming. *Nature Clim. Change,* 4**,** 911-916.
TRENBERTH, K. E. & HURRELL, J. W. 1994. Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics,* 9**,** 303-319.
TSAY, R. S. 1988. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting,* 7**,** 1-20.

TSONIS, A. A., SWANSON, K. & KRAVTSOV, S. 2007. A new dynamical mechanism for major climate shifts. *Geophysical Research Letters,* 34.

TSONIS, A. A. & SWANSON, K. L. 2011. Climate mode covariability and climate shifts. *International Journal of Bifurcation and Chaos,* 21**,** 3549-3556.

TSONIS, A. A. & SWANSON, K. L. 2012. Review article "On the origins of decadal climate variability: a network perspective". *Nonlin. Processes Geophys.,* 19**,** 559-568.

TYNDALL, J. 1861. The Bakerian Lecture: On the Absorption and Radiation of Heat by Gases and Vapours, and on the Physical Connexion of Radiation, Absorption, and Conduction. *Philosophical Transactions of the Royal Society of London,* 151**,** 1-36.

VAROTSOS, C., EFSTATHIOU, M. & CHRISTODOULAKIS, J. 2019. Abrupt changes in global tropospheric temperature. *Atmospheric Research,* 217**,** 114-119.

VAROTSOS, C. A., FRANZKE, C. L., EFSTATHIOU, M. N. & DEGERMENDZHI, A. G. 2014. Evidence for two abrupt warming events of SST in the last century. *Theoretical and applied climatology,* 116**,** 51-60.

VERDON, D. C. & FRANKS, S. W. 2006. Long-term behaviour of ENSO: Interactions with the PDO over the past 400 years inferred from paleoclimate records. *Geophysical Research Letters,* 33**,** L06712.

VERES, M. C. & HU, Q. 2013. AMO-forced regional processes affecting summertime precipitation variations in the central United States. *Journal of Climate,* 26**,** 276-290.

VINCENT, L. A. 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate,* 11.

VIVES, B. & JONES, R. N. 2005. *Detection of abrupt changes in Australian decadal rainfall (1890-1989)*, CSIRO Atmospheric Research.

VON NEUMANN, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics,* 12**,** 367-395.

WALPOLE, R. E., MYERS, R. H., MYERS, S. L. & YE, K. 1993. *Probability and statistics for engineers and scientists*, Macmillan New York.

WANG, G., SWANSON, K. L. & TSONIS, A. A. 2009. The pacemaker of major climate shifts. *Geophysical Research Letters,* 36.

WANG, H., KILLICK, R. & FU, X. 2014. Distributional change of monthly precipitation due to climate change: comprehensive examination of dataset in southeastern United States. *Hydrological Processes,* 28**,** 5212-5219.

WANG, J.-L. & WANG, J. 1994. The Test and Confidence Interval for a Change-Point in Mean Vector of Multivariate Normal Distribution. *Lecture Notes-Monograph Series,* 24**,** 397-411.

WATANABE, M., KAMAE, Y., YOSHIMORI, M., OKA, A., SATO, M., ISHII, M., MOCHIZUKI, T. & KIMOTO, M. 2013. Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus. *Geophysical Research Letters,* 40**,** 3175-3179.

WHITE, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society***,** 817-838.

WIJNGAARD, J., KLEIN TANK, A. & KÖNNEN, G. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology: A Journal of the Royal Meteorological Society,* 23**,** 679-692.

WU, Y. 2005. Confidence Interval for Change-Point. *Inference for Change Point and Post Change Means After a CUSUM Test.* New York, NY: Springer New York.

XIAO, D., LI, J. & ZHAO, P. 2012. Four-dimensional structures and physical process of the decadal abrupt changes of the northern extratropical ocean–atmosphere system in the 1980s. *International Journal of Climatology,* 32**,** 983-994.

YAN, P., FENG, G. & HOU, W. 2015. A novel method for analyzing the process of abrupt climate change. *Nonlinear Processes in Geophysics Discussions,* 2**,** 43-67.

YAN, P., HOU, W. & FENG, G. 2016. Transition process of abrupt climate change based on global sea surface temperature over the past century. *Nonlinear Processes in Geophysics,* 2016**,** 115-126.

YASUNAKA, S. & HANAWA, K. 2002. Regime shifts found in the Northern Hemisphere SST field. *Journal of the Meteorological Society of Japan,* 80**,** 119-135.

YIN, J., OVERPECK, J., PEYSER, C. & STOUFFER, R. 2018. Big Jump of Record Warm Global Mean Surface Temperature in 2014–2016 Related to Unusually Large Oceanic Heat Releases. *Geophysical Research Letters***,** 2017GL076500.

YONETANI, T. 1993. Detection of long term trend, cyclic variation and step-like change by the Lepage test. *Journal of the Meteorological Society of Japan,* 71**,** 415-418.

YOUNG, K. C. 1993. Detecting and removing inhomogeneities from long-term monthly sea level pressure time series. *Journal of climate,* 6**,** 1205-1220.

YOZGATLIGIL, C. & YAZICI, C. 2015. Comparison of homogeneity tests for temperature using a simulation study. *International Journal of Climatology***,** n/a-n/a.

ZANG, C. S., JOCHNER-OETTE, S., CORTÉS, J., RAMMIG, A. & MENZEL, A. 2018. Regional trend changes in recent surface warming. *Climate Dynamics*.

ZEILEIS, A. 2005. A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals. *Econometric Reviews,* 24**,** 445-466.

ZEILEIS, A., LEISCH, F., HORNIK, K. & KLEIBER, C. 2001. strucchange. An R package for testing for structural change in linear regression models.

ZEILEIS, A., SHAH, A. & PATNAIK, I. 2010. Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics & Data Analysis,* 54**,** 1696-1706.

ZHANG, N. R. & SIEGMUND, D. O. 2007. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics,* 63**,** 22-32.

ZIVOT, E. & ANDREWS, D. W. 1992. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*.

ZIVOT, E. & ANDREWS, D. W. K. 2002. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics,* 20**,** 25-44.