

An Optimised Machine Learning Algorithm for Detecting Shocks in Road Vehicle Vibration

PhD Thesis

College of Engineering and Science, Victoria University

Julien Lépine, MSc

2016

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

In most countries, domestic transport of products is predominantly made on roads. Protective packaging is used to protect the freight from the shocks and vibration encountered during this type of transportation. Inefficient packaging constitutes a significant problem that costs hundreds of billions of dollars and has an important environmental impact. An insufficient level of packaging increases the occurrence of product damage during transport, whereas excessive packaging increases the packages weight and volume which is costly throughout the supply chain.

In order to reduce these costs, packaging protection is currently optimised by simulating the vibration produced by road vehicles. Despite these simulations being prescribed by many standards, it is broadly acknowledged that these methods oversimplify the Road Vehicle Vibration (RVV) which imposes a significant limit on packaging optimisation. More complex RVV models have been recently developed to enhance simulations. However, the fundamental problem still remains; none of these alternatives consider the different excitation modes contained in RVV: *i.e.* the nonstationary random vibration induced by pavement's profile and vehicle speed random variations; the shocks caused by pavement's aberrations and discontinuities; and the sinusoidal (harmonic) vibration caused by unbalanced wheels and the engine-borne vibration. These excitation modes should be included in the RVV simulation to ensure that simulations are realistic and accurate and that the packaging will protect against vibration during transport without being too excessive.

Each of these modes is represented by a different mathematical model and cannot be analysed with the same statistical tools. This means that they have to be characterised separately in order to create an accurate RVV simulation model. This task is challenging because all the excitation modes are simultaneously present in the acceleration signal recorded on a vehicle.

This PhD thesis proposes to use machine learning to separate these modes from a signal. Being a first attempt to apply this approach to index RVV, the most common classification algorithms are used to identify the two predominant modes; *i.e.* the nonstationary vibration and the shocks.

The important novelty of this approach is that algorithms integrate many RVV analysis methods such as moving statistics, the Discrete Wavelet Transform and the Hilbert-Huang Transform. A comprehensive evaluation and optimisation of the classification algorithms was performed using synthetically generated RVV signal. The best performing algorithm was applied on a real measurement dataset.

The RVV mode decomposition will greatly increase the ability to correctly optimise the level of packaging required. An accurate model comprising all the characteristics inherent to RVV will be a considerable step forward to reduce unnecessary level of protective packaging without risking damage to products.

STUDENT DECLARATION

“I, Julien Lepine, declare that the PhD thesis entitled “An Optimised Machine Learning Algorithm for Detecting Shocks in Road Vehicle Vibration” 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:



Date: 11-02-2017

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my principal supervisor, Professor Vincent Rouillard, for his knowledge and support throughout the course of my thesis. His encouragement and guidance helped me push the boundary of my research and knowledge. Vincent made me feel part of a team and he had a significant impact on my professional development.

I would also like to express my warm thanks to my co-supervisor, Associate Professor Michael Sek, for his valuable input on my research. His knowledge and perfectionism made my work better.

A special thank goes to my colleagues Dr Daniel Ainalis, Dr Matthew Lamb and Mr Michael Long for their support, ideas, coffees and help with English grammar (or should I say with Australian idioms). Your presence made my PhD much more enjoyable.

I am also grateful to Victoria University and the Natural Sciences and Engineering Research Council of Canada for their respective scholarship. Their funding helped me focus exclusively on my research.

I would also like to thank my friend Mr Peter Ganss who gave external perspective on my work.

My acknowledgments would be incomplete without thanking my wife Catherine who followed me from Canada to Australia for my PhD. She has an endless patience and her support was essential to my work.

PREFACE

I can finally see the end of my thesis and start thinking about the free time I will have afterward. I will have more time to enjoy the hobbies I have neglected in the past years. In prevision of that, I recently ordered a new lens for my camera. I thought I had received it a few days ago when the mailman left a big box at my doorstep. I was so eager to test my new lens that I rapidly unpacked it. It was definitely something fragile inside as the box was filled with a lot of cushioning material. While I was digging my way through the Styrofoam, I imagined how ironic it would be to find my lens broken since my PhD is related to transport hazard and packaging optimisation. Surprisingly my discovery was even more ironic. It turns out it wasn't my lens or any fragile product; instead I found a small pyjama ordered by my wife for our little daughter.

FOREWORD

Papers and publication resulting from this thesis at time of submission:

1. Lepine, J., V. Rouillard, and M. Sek, 2014, *Using the Hilbert-Huang Transform to Identify Harmonics and Transients within Random Signals*, Proceeding of the 8th Australasian Congress on Applied Mechanics, Melbourne, Australia.
2. Lepine, J., V. Rouillard, and M. Sek, 2015, *Review Paper on Road Vehicle Vibration Simulation for Packaging Testing Purposes*. *Packaging Technology and Science*, 2015. 28(8): p. 672-682.
3. Lepine, J., V. Rouillard, and M. Sek. 2015, *Wavelet Transform to Index Road Vehicle Vibration Mixed Mode Signal*, Proceeding in ASME 2015 Noise Control and Acoustics Division Conference at InterNoise 2015, American Society of Mechanical Engineers, San-Francisco, USA.
4. Lepine, J., M. Sek, and V. Rouillard, 2015, *Mixed-Mode Signal Detection of Road Vehicle Vibration Using Hilbert-Huang Transform*. Proceeding in ASME 2015 Noise Control and Acoustics Division Conference at InterNoise 2015, American Society of Mechanical Engineers, San-Francisco, USA.
5. Lepine, J., Rouillard, V. and Sek, M., 2016, *On the Use of Machine Learning to Detect Shocks in Road Vehicle Vibration Signals*, *Packaging Technology and Science*.

TABLE OF CONTENTS

Abstract	ii
Student Declaration	iv
Acknowledgments	v
Preface	vi
Foreword	vii
Abbreviations	xiii
Lexicon	xiv
Symbols	xv
Chapter 1: Introduction	1
Chapter 2: Literature Review of Road Vehicle Vibration Simulation Methods	8
2.1 Standardised Method	9
2.1.1 Shortcomings of Standardised Random Vibration Testing	10
2.2 Time History Replication	11
2.3 Non-Gaussian Simulation	11
2.4 Nonstationary Gaussian Simulation	12
2.4.1 Split Spectra Decomposition	13
2.4.2 Random Gaussian Sequence Decomposition	13
2.4.3 Vehicle and Road Characteristics-Based Simulation	15
2.4.4 Wavelet Based Gaussian Decomposition	16
2.4.5 Hilbert-Huang Transform	17
2.4.6 Bayesian Detector	18
2.5 Transient Events Simulation	19
2.6 Harmonic Simulation Method	20
2.7 Conclusion on Simulation Methods	21
Chapter 3: Hypothesis and Objective	22
3.1 Definition of the RVV Modes	22

3.1.1	Nonstationary Components	22
3.1.2	Transient Components.....	23
3.1.3	Harmonic Components.....	24
3.2	Sub-Objectives	24
Chapter 4:	Synthetic Road Vehicle Vibration Signals	25
4.1	Nonstationary Signal Synthesis.....	26
4.2	Transient Signal Synthesis	29
4.3	Harmonic Signal Synthesis	31
4.4	Conclusion on Synthetic RVV Signal.....	33
Chapter 5:	Time Domain Analysis.....	34
5.1	Crest Factor	34
5.2	Moving Statistics.....	35
5.2.1	Moving RMS.....	36
5.2.2	Moving Kurtosis.....	37
5.2.3	Moving Crest Factor.....	42
5.3	Conclusion on Time Domain Analysis	43
Chapter 6:	Time-Frequency Analysis	44
6.1	Hilbert-Huang Transform Assessment.....	45
6.1.1	Nonstationary Component.....	47
6.1.2	Transient Component	49
6.1.3	Harmonic Component	50
6.1.4	Energy Density/Average Period Significance Test	52
6.1.5	Fractional Resampling Technique.....	55
6.1.6	Conclusion on the Hilbert-Huang Transform.....	58
6.2	Discrete Wavelet Transform Assessment.....	58
6.2.1	Nonstationary Component.....	61
6.2.2	Transient Component	61
6.2.3	Harmonic Component	63
6.2.4	Conclusion on Wavelet Transform.....	64
6.3	Conclusion on Time-Frequency Analysis Tools	64
Chapter 7:	Machine Learning	65
7.1	Machine Learning Overview.....	65
7.2	Predictors.....	66
7.2.1	Moving RMS.....	67
7.2.2	Moving Crest Factor.....	67

7.2.3	Moving Kurtosis.....	68
7.2.4	Hilbert-Huang Transform Predictors.....	68
7.2.5	Discrete Wavelet Transform	70
7.3	Classifiers Training	70
7.4	Classifiers Validation	72
7.4.1	Detection Enhancement Algorithm.....	73
7.4.2	Detection Accuracy	74
7.5	Classifiers Evaluation and Selection	77
7.5.1	Discriminant Analysis	77
7.5.2	Naïve Bayes Classification.....	77
7.5.3	Decision Trees.....	78
7.5.4	<i>k</i> -Nearest Neighbours.....	80
7.5.5	Ensemble	83
7.5.6	Support Vector Machine	85
7.5.7	Neural Networks	89
7.6	Classifiers Comparison	90
7.7	Conclusion on Machine Learning	91
Chapter 8:	Road Vehicle Vibration Classifiers.....	93
8.1	Classifier Accuracy	93
8.1.1	Optimal Operation Point	94
8.1.2	OOP Selection.....	96
8.1.3	Shock Amplitudes Distribution.....	99
8.1.4	Pseudo Energy Ratio/Fall-Out (PERFO) curve.....	102
8.1.5	PERFO Criterion.....	103
8.2	Classifier Optimisation.....	109
8.2.1	Decision Tree	110
8.2.2	<i>k</i> -Nearest Neighbours.....	111
8.2.3	Ensemble	112
8.2.4	SVM.....	113
8.2.5	Optimal Classifiers.....	114
8.3	Conclusion on Classifiers.....	115
Chapter 9:	Model Validation.....	117
9.1	RVV Measurement.....	118
9.1.1	Training Measurement Dataset	119
9.1.2	Validation Measurement Datasets.....	120

9.2	Synthetic Training Signal.....	121
9.2.1	Nonstationary Component.....	122
9.2.2	Transient Component.....	124
9.3	Classifier Training.....	127
9.3.1	Synthetic Calibration.....	127
9.4	Validation on Real Measurements.....	130
9.5	Classifier Application.....	136
9.5.1	Spectral Analysis.....	139
9.6	Future Validation Work.....	140
9.7	Conclusion on Model Validation.....	141
Chapter 10:	Conclusions.....	142
References	145
Appendix A:	Hilbert-Huang Transform.....	152
A.1.	Hilbert Transform.....	152
A.2.	Empirical Mode Decomposition.....	154
A.3.	Time-Frequency Analysis tools.....	156
A.4.	Signal Detection and Noise Filtering.....	157
A.4.1	Energy Density/Average Period Significance Test.....	158
A.4.2	Fractional Resampling Technique.....	158
Appendix B:	Wavelet Transform.....	160
B.1.	Filter Bank.....	160
B.2.	Perfect Reconstruction.....	161
B.2.1	z Transform.....	162
B.3.	Multiresolution Synthesis.....	162
B.4.	Wavelet Transform Formulation.....	163
B.5.	Wavelets Creation.....	165
B.5.1	Haar Example.....	165
B.6.	Signal Decomposition.....	168
B.7.	Wavelet Selection.....	168
B.7.1	Vanishing Points.....	168
B.7.2	Filter Length.....	169
B.7.3	Orthogonality, Biorthogonality, Symmetry.....	169
B.8.	Signal Compression.....	169
Appendix C:	Sampling Frequency Conditioning to Enhance HHT Processing.....	170
C.1.	Context.....	170

C.2.	Effect on Mode-Mixing.....	171
C.3.	Effect on Computing Time.....	173
C.4.	RVV Example	174
C.5.	Conclusion on Sampling Frequency Conditioning.....	175
Appendix D:	Bayes' Criterion	176
Appendix E:	Minimax Criterion.....	178

ABREVIATIONS

AUC	Area Under the ROC Curve
CG	Centre of Gravity
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
FFT	Fast Fourier Transform
FRT	Fractional Resampling Technique
HHT	Hilbert-Huang Transform
IFFT	Inverse Fast Fourier Transform
IMF	Intrinsic Mode Function
kNN	<i>k</i> -Nearest Neighbours
MAP	Maximum a Posteriori
MLE	Maximum-Likelihood Estimator
OOP	Optimal Operation Point
PDF	Probability Density Function
PDS	Power Density Spectrum
PERFO	Pseudo Energy Ratio/Fall-Out
PMF	Proto Mode Function
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RUS	Random UnderSampling
RVV	Road Vehicle Vibration
SVM	Support Vector Machine
WPT	Wavelet Packet Transform
ZMNL	Zero-Memory NonLinear

LEXICON

Classifier	Algorithm predicting the class of a data point
Crest Factor	Ratio between absolute value and the intensity of a signal
Fall-out	Number of false no-detection over the number of points in a dataset without events, antonym: specificity
Filter Bank	Series of band-pass filters arranged in consecutive octave bandwidths
Fractional Gaussian Noise	Random process with a Gaussian amplitude distribution and a specific spectrum, <i>e.g.</i> white noise, red noise, Brown noise...
Kurtosis	Fourth statistic moment of a data set representing its distribution shape
Leptokurtic	Distribution with a kurtosis above 3, <i>i.e.</i> high probability of extremum values
Mode	Component of a signal, do not confuse with “mode” from modal analysis
Nonstationary	Characteristic of signal that has time variant statistical moments
Nyquist Frequency	Half of the sampling frequency, <i>i.e.</i> the high frequency that can be reproduced in a digital signal
Optimal Operation Point (OOP)	Classifier’s detection threshold that optimises the true detection rate and fall-out
Platykurtic	Distribution with a kurtosis below 3, <i>i.e.</i> low probability of extremum values
Predictor	Statistic used by classifiers to predict data class
Sensitivity	Number of true-detection sequences over the total number of event sequences in a dataset
Specificity	Number of true no-detection over the number of point in a dataset without events, antonym: fall-out
Transient	Short time finite high amplitude vibration <i>e.g.</i> shock

SYMBOLS

f	Frequency
G	Gaussian function
M	Mass
N	Number of occurrence
t	Time
U	Normalised uncertainty
Δf	Frequency variation
ε	Noise
κ	Kurtosis
μ	Mean
π	Number PI
σ	Standard deviation
∞	Infinity
f	Frequency
G	Gaussian function
M	Mass

Other symbols present in the thesis have context-specific meaning and are defined within the relevant text.

Chapter 1: INTRODUCTION

Road transportation is an important element of the product distribution industry. All products and goods are at some time transported by road to reach consumers. In Australia, as is the case in most countries, road transport vehicles are the main constituent of domestic transportation. For instance, the Bureau of Infrastructure (2010) estimated that more than 67 % of finished products and other goods are transported by road in Australia, equivalent to 243 billion tonnes per kilometres (tkm) of travel. Road transport is also essential to interconnect with other modes of transport (rail, air and sea).

As the structure of the economy is changing, going from local to global production, product distribution becomes more extensive and the demand for road transportation increases. In Australia, road transportation is growing by about 4 % in volume every year (Austroads, 2005). This measure is also supported by the Bureau of Infrastructure (2010) which forecasts a 4 % to 6 % annual increase in road freight activity from 2010 (Figure 1). This prediction can be generalised worldwide as a share of GDP (Gross Domestic Product) is predicted to shift from the Americas and Europe to Asia (Figure 2). This means that products and consumer goods will be increasingly produced and consumed in many different countries, increasing global demand for freight transport.

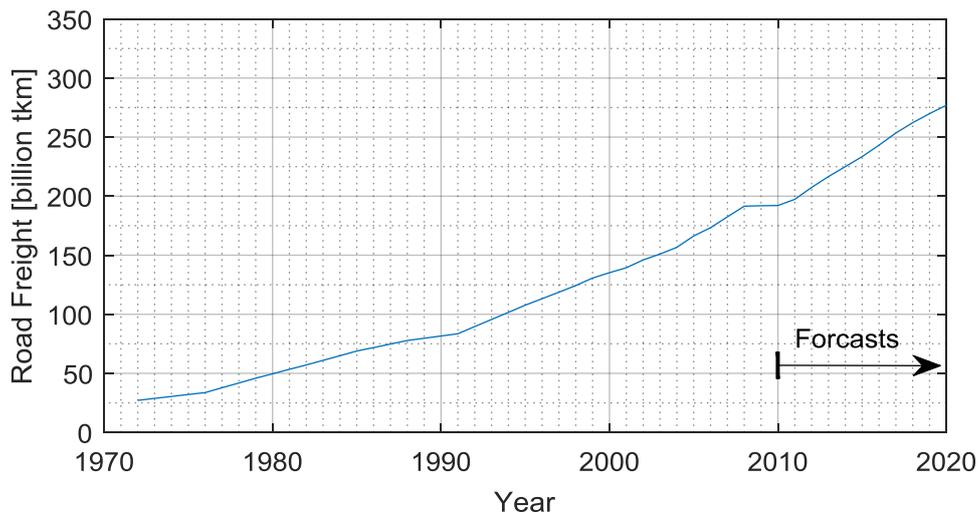


Figure 1: freight transported by road in Australia (data from Bureau of Infrastructure, 2010)

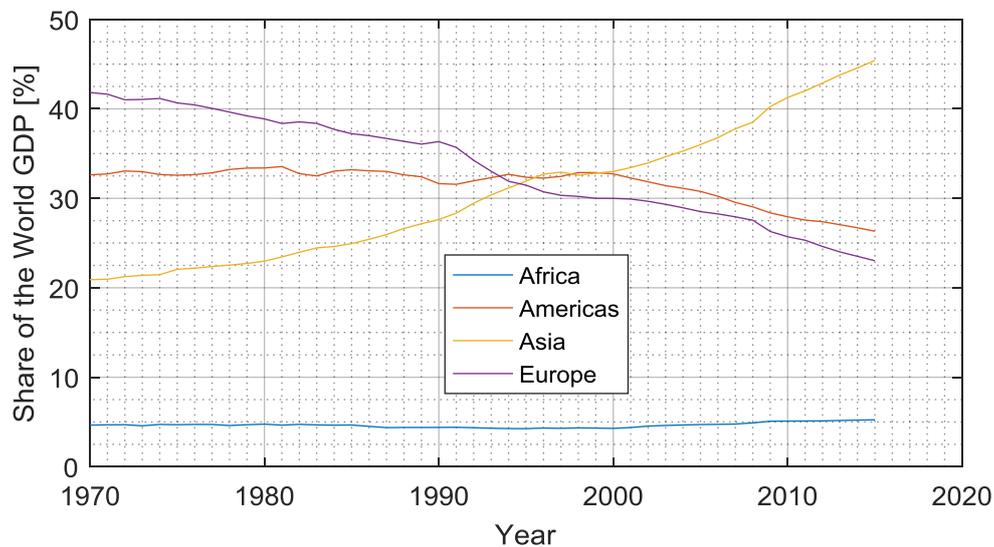


Figure 2: GDP per continent (Data from World Economics Journal, 2016, Australasia countries included in Asia)

Producers, distributors and consumers rely on an effective and reliable transport system. This is increasingly crucial to the global economy as consumer goods and products are transported all over the planet. The integrity of freight is at risk during transport. Product can be damaged and lost throughout the different stages of the transport chain, including during handling and transport. The mishandling risk such as dropping a product is principally caused by human error and can be mitigated by changing such aspects as: handling procedures, equipment and working conditions. Nonetheless, the hazards encountered during transport (*e.g.* by roads, railways, sea and air) are far more difficult to control because they depend on the randomness of the process

which makes the road transportation far from risk free. When transported by road, freight is subjected to shocks and vibration caused by the interaction between the vehicle and pavements.

The intensity of shocks and vibration depends on the pavement profile of the road and the vehicle suspension. The pavement profile generates random excitation (shocks and vibration) which varies as a function of the pavement material and level of maintenance. The quality of the material and frequency of maintenance can vary considerably between different regions even within the same country, depending on the road usage and availability of resources. There are also variations in vehicle suspension design and maintenance. Suspension systems are designed to improve ride quality by smoothing out the effect of the shocks and vibration as the vehicle travels along the road. There are two principal types of suspension, air and leaf spring suspensions, which provide different performances. Vehicles equipped with air suspension tend to have a smooth constant ride quality because the stiffness of the pneumatic spring can be adjusted for different situations. However, this type of suspension is expensive and there are still many transport vehicles equipped with leaf spring suspension, which has a constant stiffness, to the detriment of ride quality. The regular maintenance of the vehicle suspension is also essential to its good functionality. A poorly-maintained suspension increases the level of shocks and vibration induced to the freight.

Shocks and vibration occurring during road transport can be very harmful to products and consumer goods as they are sustained over a long period of time. Even if the intensity of the shocks and vibration is not sufficient to cause immediate failure, their sustained action can cause fatigue-type failure or product damage through wear (*e.g.* scuffing). Since road transportation has a pivotal role in the global economy, it is imperative to protect products and goods from the inherent hazards of this mode of transport. Protective (or distribution) packaging is used to protect freight from shocks and vibration in road transport.

The need for protective packaging has created an important market. In 2015, about 26 billion USD was spent worldwide on protective packaging and this amount is estimated to reach 36 billion USD by 2020 (Consulting, 2015). Protective packaging does not add any value to products and consumer goods, but its cost directly affects consumers. Transport costs can be reduced by optimising the use of protective packaging. Either insufficient or excessive packaging is costly; insufficient packaging increases incidences of damage and excessive packaging increases the volume and weight of packages. There is an optimal level of protection that gives the minimal cost (Figure 3).

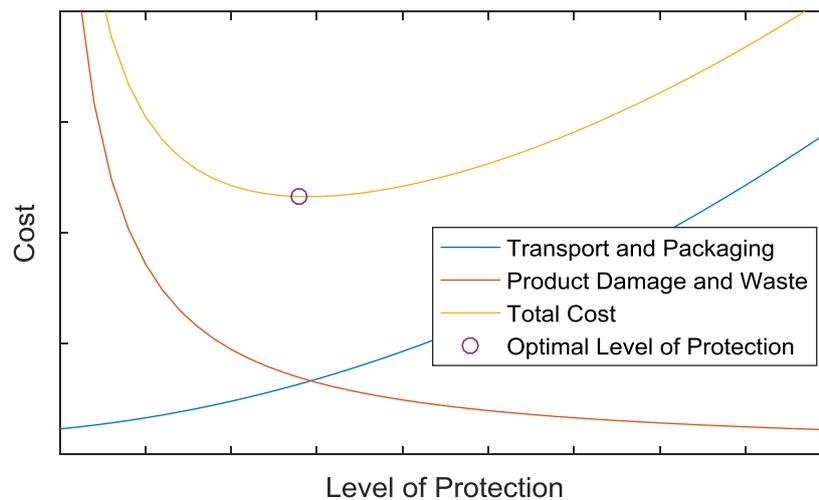


Figure 3: economic cost function for protective packaging optimisation (reproduced from Rouillard, 2007a)

In practice, transport and packaging cost minimisation is not as trivial as finding a minimum point on a curve. The optimal level of protection can be achieved by adjusting both the protective packaging and the product robustness (Figure 4).

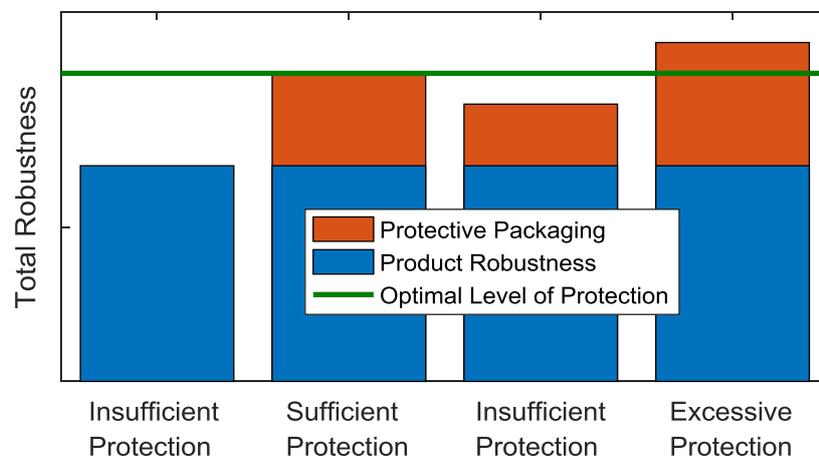


Figure 4: possible product robustness for a given level of protection (Root, 1997)

The cost of packaging has to be considered to find the optimal trade-off between protective packaging and product robustness. This is not limited to the cost of the cushioning material but also to the volume and mass of packages. Larger and heavier packages are costly throughout the supply chain process: *e.g.* they require more warehouse and vehicle space; they are difficult to handle and so on. They also have a negative environmental and social impact. For instance, the extra transport vehicle requirement increases traffic; consumes more fuel; and causes premature road deterioration. The extra cushioning material disposal also increases pollution. Over-

packaging of goods and products has therefore more than an economic impact; it also contributes to increased pollution and vehicle traffic all around the world (Cebon, 1989, Janic, 2007, Oestergaard, 1991).

Insufficient packaging also has detrimental consequences because it results in product damage and waste. The most appalling example is seen in emerging countries where in some cases, half of all food is wasted as it moves between farms and markets, which is mostly a consequence of improper transportation (Fox, 2013). In most cases, products damaged during transport are discarded and sent to waste, which directly affects the environment and the consumers' perception of the products. In addition to product waste, the problem of insufficient packaging leads to product replacement and reshipping. This increases transport costs due to repetitive deliveries (initial shipment, return to the expeditor and reshipment) which also increases traffic, fuel consumption and so forth. The issues related to insufficient packaging are more easily quantified, which explains the general tendency to over-package products.

One approach to mitigate the environmental footprint of product transport is to design sustainable packaging (Svanes et al., 2010) and create governmental initiatives to sustainable packaging, such as the Australian Packaging Covenant which targets the use 70 % sustainable materials for packaging by 2020 (APC, 2015). However, applying sustainable practices to packaging has a limited impact on the over-packaging issue as the disposal of packaging materials is a mere fraction of the whole over-packaging issue.

The economic and environmental toll of transportation can only be significantly decreased by optimising protective packaging systems. More efficient packaging has a positive effect throughout the supply chain, including product storage, handling, shipping and packaging disposal. As illustrated in Figure 4, product robustness and packaging protection are designed to reach an optimal level of protection which depends on the hazards encountered during the supply chain. The risk posed by these hazards depends on their likelihood and severity. For instance, dropping a package from the top shelf of a five metres high warehouse would result in very severe product damage. Nevertheless, for a properly operated warehouse this represents a small risk as this type of accident is very unlikely. By contrast, the vibration encountered during road transportation is far less severe, but is unavoidable. Therefore, the risk of product defects due to Road Vehicle Vibration (RVV) is greater than that of a five metres drop. In this example, the level of protection should be optimised as a function of the RVV but not of the risk of a five metres drop.

As previously stated, road transportation is omnipresent in the global economy and it constitutes an important risk as road vehicles generate shocks and vibration that can damage products. Consequently, road transportation risk corresponds to a preponderant portion of the optimal

level of protection. The level of protection required to mitigate the risks posed by road transportation is found by simulating the RVV. An optimised package should protect its content from the RVV occurring during normal transport condition without over-protecting it, which means the packaging should not protect against hazard exceeding normal RVV, *e.g.* a vehicle collisions or unexpectedly rough roads. Products are therefore over-packaged when the RVV level is poorly defined and, as stated by Sek (2001), conservative simulation leads to over-packaging.

Packaging optimisation can only be as good as the accuracy of the simulation of RVV. The characterisation of vehicle vibration is essential to optimise packaging because it can be used to produce accurate simulations of the loads and stresses that occur during transport. Due to the nature of the dynamic interaction between the road surface and road vehicles, the resulting motion is often complex and cannot be characterised by simple statistical means. As discussed by Charles (1993), the RVV can be divided into three distinct modes: (1) the nonstationary random components representing the different levels of vibration induced to freight depending on the vehicle speed and the variation in road surface; (2) the transient components composed of the shocks produced by the randomly-occurring road surface aberrations (such as large cracks, potholes, speed bumps, drains, rail crossings...); and (3) the harmonic components which are vibration produced by the vehicle's components such as its wheels, drivetrain and engine. The last-named components can also be nonstationary in time, for instance, when the vehicle speed and gear ratio change but they are fundamentally different from the nonstationary random components as they are composed of deterministic functions. A comprehensive analysis of these modes and their effect on RVV are presented in Chapter 2, 3 and 4.

Several RVV simulation methods have been developed to date but none of them can accurately simulate RVV. For instance, there are standardised methods which simulate RVV with stationary Gaussian random signals (ASTM-D4728, MIL-STD-810F, ISO-13355, ISTA). However, even if these methods are supported by the international standard organisation, they have a fundamental limitation; they do not consider any of the three RVV modes and assume that RVV consist of random Gaussian vibration. To overcome this limitation, enhanced simulation methods have been proposed but not broadly adopted by the industry yet. An extensive review of these methods is presented in detail in Chapter 2. As shown in this literature review, there have been great advances in RVV simulation in the last decade. However, even the best RVV simulation methods still have a major limitation; they considered RVV as a single mode signal. This assumption creates oversimplified models because it neglects the three distinct RVV modes. For instance, the nonstationary component can be simulated as a sequence of stationary random segments of different intensities. Segments can be characterised using

different methods such as the Random Gaussian Sequence Decomposition (Rouillard, 2007a) or the Wavelet-Based Gaussian Decomposition (Griffiths, 2012). The problem with these methods is that they cannot properly characterise the transient and harmonic modes because they will try to identify them as a Gaussian random vibration. Transient and harmonic components can occur at random times but they are fundamentally deterministic processes which cannot be modelled as a Gaussian random vibration.

The transient components have been studied using methods such as moving crest factor (Bruscella, 1997) and wavelet decomposition (Wei et al., 2005, Nei et al., 2008). These methods properly defined the transient components as the response of the road vehicle to shocks. However once again these methods consider RVV vibration as a single mode signal and their accuracy is significantly affected by nonstationary components. The nonstationary, transient and harmonic components coexist in the RVV signal and need to be identified, separated and analysed separately in order to achieve an accurate and realistic characterisation and simulation of RVV.

The objective of the thesis is to identify the different modes buried in RVV signals. Each mode needs to be well defined before being identified. The definition of each mode is presented in Chapter 3 along with the thesis' sub-objectives. With the RVV modes clearly defined, a RVV signal synthesiser was developed (Chapter 4) to assess the potential signal processing techniques in the time domain (Chapter 5) and the time-frequency domain (Chapter 6) which can be used to identify the modes. As for the current simulation methods, each technique can identify one RVV mode but none of them could accurately identify more than one mode. This is how the application of machine learning can benefit mode identification.

At the beginning of the thesis project, machine learning algorithms were not intended to be used. It was only after the assessment of the signal analysis techniques that it became essential to the RVV modes' identification. This is because machine learning allows several analysis techniques to be integrated together to classify the modes present in a signal. This greatly improves the classification of mixed-mode signals as the outputs of the best analysis techniques for each mode are used in conjunction. In other words, it combines all the techniques to work together to provide better predictions. An evaluation of several machine learning processes is presented in Chapter 7 and the best algorithms for RVV modes detection were optimised (Chapter 8) using synthetic signals before being validated on *in-situ* measurements (Chapter 9).

The merit of machine learning is that it is directly beholden to the merit of its underlying processes. So before looking into machine learning, it is essential to truly understand and master the more 'classical' RVV analysis techniques. As Newton said: "If I have seen further than others, it is by standing upon the shoulders of giants."

Chapter 2: LITERATURE REVIEW OF ROAD VEHICLE VIBRATION SIMULATION METHODS

Simulation of the complex shocks and vibration induced to products during transportation is a necessary step in the development of packaging, in order to test its effectiveness in protecting the product (Root, 1997). However, the accuracy of the simulation itself is critical for packaging optimisation. Excessively severe simulations will ensure the product's integrity but may result in the use of excessive cushioning material and have a direct impact on distribution costs, whereas moderate simulations may result in insufficient protection. Therefore, the simulation has to accurately recreate the different shock and vibration elements generated by the vehicle. These include vibration generated by road roughness, road surface aberrations (*e.g.* cracks, bumps, potholes) and by the vehicle drivetrain system (*e.g.* wheels, drivetrain, engine).

In the last few decades, various methods have been developed to improve the accuracy of Road Vehicle Vibration (RVV) simulation. Because of the complex nature of this multi-process type of vibration, these methods are varied and each method attempts to address the diverse limitations of the current simulation methods. Their frameworks are also different and include heuristic-based methods as well as more statistical and signal processing-based methods. A review of five simulation methods was presented by Richards (1990), but many new methods have been developed since then. Another review is therefore needed in order to present an overview of the different simulation approaches available today. The methods proposed by standards organisations are discussed first. A considerable amount of literature has been published on the shortcomings of these standard methods and present alternative simulation methods. These alternatives are separated into five categories: time history replication, non-Gaussian simulation, nonstationary simulation, transient events simulation and harmonic simulation.

2.1 Standardised Method

Eccentric cams fixed under a testing table were among the first devices used to simulate transport vibration. This excitation mechanism was simple. When the cams attached under a table were spinning fast enough, the product detached from the table at each rotation, creating repetitive shocks as described in the ASTM Standard D999. This method is still used today (Mahajerin and Burgess, 2010), even though it was later recognised to produce repetitive shocks rather than actual transport vibration. Hence, Kipp (2001) recommends caution in extending any conclusions drawn from the technique.

The natural evolution of the test method was to use single frequency vibration to avoid any shock events in the simulation as described in standards IEC 60068 and ASTM D3580. The excitation signal for this method consists of sweeping the sinusoidal excitation frequency during the test. Once again, this method is criticised because it does not reproduce the randomness of the road excitation. It has been suggested to use this method as an investigation and design tool rather than a packaging performance test (Kipp, 2000a, 2001, 2008).

The limitations of these standards led to the use of random vibration as a standard road vibration simulation in the early 1980s. This is described in standards: ISTA, MIL-STD-810F, ISO 13355 and ASTM D4728. These procedures produce random Gaussian vibration from various target Power Density Spectra (PDS) representing different types of vehicle. These PDS can be obtained from measurements on one or more vehicles or from road profile surveys and vehicle dynamic models. However, Rouillard and Sek (2013) concluded that the vehicle measurement-based PDS method gives better results since they do not require approximation of model parameters.

An advantage of the PDS-based method is the apparent possibility of reducing the testing time by increasing the intensity (*i.e.* Root Mean Square (RMS) value) of the simulation. The relationship between actual journey and test durations is inspired on Basquin model:

$$\frac{t_j}{t_t} = \left(\frac{a_t}{a_j} \right)^k \quad (\text{eq. 1})$$

where t_t is the test duration; t_j is the actual journey duration; a_t is the test intensity; a_j is the actual journey intensity; and k is a constant associated with the material/product tested, and where $k = 2$ (Shires, 2011, Kipp, 2000b) and $k = 5$ (Ulrich Braunmiller, 1999) are generally used.

2.1.1 Shortcomings of Standardised Random Vibration Testing

Random vibration testing based on target PDS is a common simulation method used to test packaging effectiveness. However many researchers challenge the validity of the Gaussian nature of the vibration produced by this method (even though many standards use this method).

To understand the major deficiency of this method, the process that underpins the generation of the vibration signal must be understood. The synthetic signal (Figure 5 a) is created from a target PDS of a RVV signal (Figure 5 b) to ensure that it has the specific frequency spectrum contained in the real RVV signal (Figure 5 c). However, PDS do not have any time information; they only provide the average signal power density as a function of frequency. Therefore, in order to create a random signal, the PDS is transformed into an amplitude spectrum with a uniformly-distributed random phase. This spectrum is then transformed into the time domain via an Inverse Fast Fourier Transform (IFFT). This produces a random Gaussian signal that corresponds to the specific frequency spectrum.

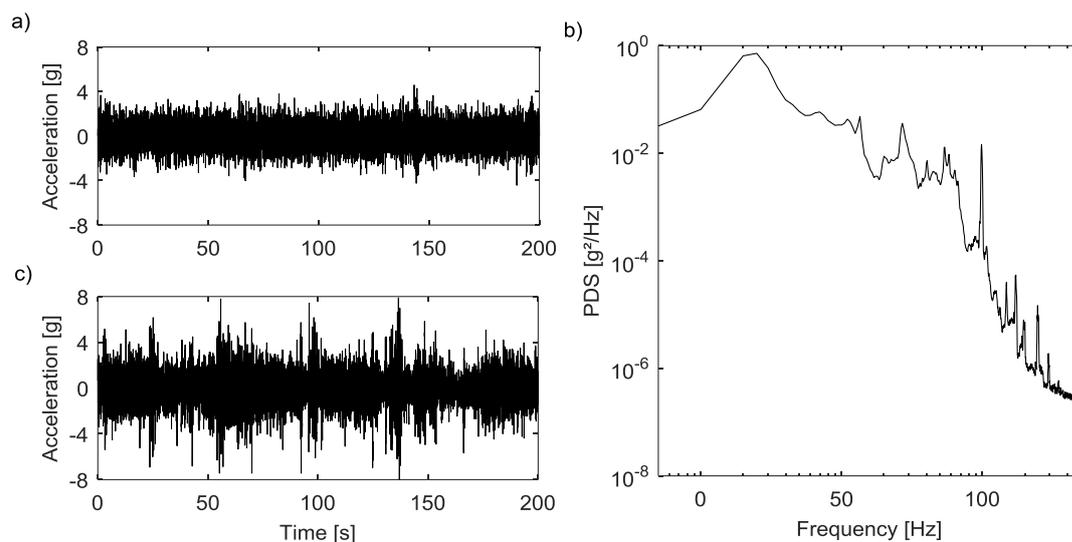


Figure 5: Gaussian signal synthesis of a typical RVV signal: a) Gaussian synthetic signal; b) PDS of the vehicle signal used for the synthesis; c) measured vehicle signal

The problem is that the steady-state nature (constant vibration intensity level *i.e.* RMS value) of the Gaussian signals produced by this method does not represent realistic RVV. Vibration fluctuations caused by variations in road roughness (profile characteristics) and vehicle speed cannot be represented by a constant RMS value signal (Figure 5 c). The high amplitude events are averaged-out in the PDS, so they are not present in the synthesised signal (Figure 5 a). Therefore, vibration tests based on PDS alone lack high amplitude events which can have a significant effect on product damage (Charles, 1993, Rouillard, 2007a, Kipp, 2008).

The target PDS provided in the standards to generate the random signal also concerns researchers because they are not necessarily representative of real RVV. These functions oversimplify vehicle dynamic behavior. They are approximate representations which include all vehicle dynamic behavior and the PDS shape is not accurately represented (Rouillard and Sek, 2013, Rouillard, 2008). This has a serious consequence in that the energy of the simulated signal is spread across a broader frequency band rather than focused around a narrow frequency region as is the case for real vehicles which exhibit mechanical resonances (Rouillard, 2008).

An important shortcoming also comes from the time compression. The artificial signal amplifications that reproduce high amplitude events of vehicle excitation are normally distributed (Gaussian) which is not the case with vehicle excitation. Therefore high amplitude events have much higher incidence in the time compressed signal which does not reflect reality.

2.2 Time History Replication

The more candid way to reproduce RVV without all the shortcomings of PDS-based Gaussian reproduction is to play back the time history vibration recorded from vehicle measurements on a vibration table. Simple as this solution may be, it presents the following fundamental issues.

One single time history replication has no statistical significance. In other words, using the time replication of one journey is only a sample of all possible journeys (population). This specific journey may have been more or less severe than usual (statistical extremum) and there is no way to verify this unless many journeys are recorded. The issue of dealing with several journeys, how to select a typical one in an objective manner, remains unresolved.

The solution could be to replicate different time histories of various journeys in a series to increase the test significance. This requires extremely long testing time and since the time history signal cannot be significantly compressed (Kipp, 2001). Therefore, it is unrealistic to use a series of time histories for packaging testing purposes.

2.3 Non-Gaussian Simulation

As previously explained, the PDS-based RVV replication method generates normally distributed (Gaussian) acceleration signals. This is problematic because RVV is well known to be non-Gaussian, more specifically leptokurtic (high kurtosis value, more details are presented in Chapter 4). Therefore for the same RMS value, a Gaussian signal has a lower maximum than the leptokurtic signal (Rouillard and Lamb, 2008, Rouillard, 2007a).

Some researchers addressed this issue by reproducing random vibration from a target PDS using different distribution (non-Gaussian) (Otari et al., 2011, Van Baren, 2005, Steinwolf and Cannon Iii, 2005, Hosoyama et al., 2012). One of the methods used to recreate non-Gaussian

random vibration based on a spectrum is to distort the waveforms using Zero-Memory NonLinear (ZMNL) monotonic function (Smallwood, 2005). However, this operation is made in the time domain, so the spectral properties of the synthesised signal are not necessarily conserved. To control the spectrum and the kurtosis of the synthetic signal, Van Baren proposes in his US patent (Van Baren, 2008) to use adaptive filters.

The main limitation of the non-Gaussian simulation, however, is its inability to reproduce the nonstationary nature of the vehicle excitation. Since the vehicle speed and the roughness of the road pavement vary throughout a journey, the vehicle excitation is therefore nonstationary (Richards, 1990). This nonstationarity can be evaluated with the variation in time of the statistical moments of the signal. For instance when the mean value and the RMS value (or standard deviation for zero mean) of a signal do not vary in time, the signal is called weakly stationary or stationary in the wide sense. However, because it is usually impractical to prove that a signal is strongly stationary (where all the possible statistical moments are time invariant), Bendat and Piersol (2011) propose that: “for many practical applications, verification of weak stationarity will justify an assumption of strong stationarity.”

According to that definition, the non-Gaussian simulations described above are necessarily stationary because they are based on a fixed distribution, so the statistical moments of the signal describing this distribution are time invariant. In that sense, this type of simulation does not faithfully represent RVV, which is nonstationary.

2.4 Nonstationary Gaussian Simulation

Simulating the nonstationary nature of RVV has been addressed by several researchers. They all use a similar approach of decomposing the signal into stationary segments. Several methods have been used to decompose a RVV signal into stationary segments. The following section presents a review of methods used to decompose the signal.

It is interesting to note that decomposing a RVV signal into Gaussian segments has shown that the non-Gaussianity of the signal is not inherent to the vehicle excitation but it is caused by the fluctuation of the RMS values of the signal, *i.e.* its nonstationary nature (Charles, 1993, Rouillard, 2007a, Rouillard, 2007b, Rouillard and Sek, 2000, Rouillard and Sek, 2010, Bruscella, 1997). In other words, it is not the vehicle dynamics that make the vibration non-Gaussian; it is actually caused by the combination of different road roughness levels and vehicle speed (constant for a certain period of time) which creates RVV segments of different RMS values. This further justifies the use of signal decomposition to accurately simulate RVV.

2.4.1 Split Spectra Decomposition

One of the first initiatives to simulate the nonstationary nature of RVV is to decompose the signal into low and high amplitude events. The idea is to calculate two PDS based on 70 % of the RVV signal with the lowest acceleration levels and on 30 % with the highest acceleration levels (Chonhenchob et al., 2012, Singh et al., 2006). This is based on the hypothesis that the low level spectrum represents the vehicle response to road profile random variations and the high level spectrum represents the response to transient events such as speed bump, cracks, potholes *etc.*

The actual selection of the signal amplitude to effectuate the separation in order to calculate the PDS is subjective. For instance Wallin (2007) used a 80 %:20 % ratio and Kipp (2008) used a three way split spectra for 70 %, 25 % and 5 % of the signal. Kipp (2008) even went further in the number of split spectra by proposing to define the probability split spectra as: “the probability that an encountered PDS level will be at or below the profile based on all data events recorded.” For instance, the profile proposed by him is 100 %, 99 %, 95 %, 90 %, 80 % and below. Unfortunately, there is not enough published information on the implementation and validation of this method to properly assess it.

Wolfsteiner and Breuer (2013) also proposed to decompose rail vehicle vibration into several PDS. This decomposition seems to be applied in the time domain, but once again, limited information exists about how it was achieved. Therefore, it is difficult to assess the potential of this technique, but according to the authors it is an effective method to use for finite element analysis simulation purposes.

The split spectra method has not been shown to be effective as very little research has been published on the assessment of the technique. At this stage, many of the parameters used to split the spectra are arbitrary and subjective. Furthermore, the technique produces an orderly sequence (incrementally increasing RMS level) of Gaussian random vibration conforming to a single PSD. It is not adequate to deal with realistic nonstationary RVV signals.

2.4.2 Random Gaussian Sequence Decomposition

Another approach to deal with nonstationary RVV is to assume that the PDS shape is constant and to decompose the signal into Gaussian segments of different amplitude. The Random Gaussian Sequence Decomposition (RGSD) was first proposed by Charles (1993) and then implemented by Rouillard (2007a). The general idea is to decompose a non-Gaussian distribution of the RVV signal into several Gaussian distributions. As shown in Figure 6, a leptokurtic normal distribution of RVV can be represented as a sum of Gaussian Probability Density Functions (PDFs).

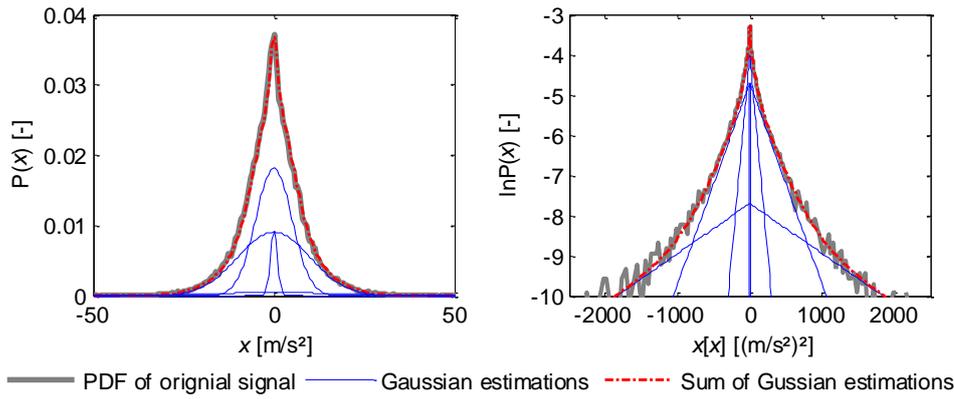


Figure 6: Gaussian decomposition in a linear scale on the left and a log scale on the right

The contribution of every sequence with its standard deviation σ_i (also called RMS value) is weighted with a term called Vibration Dose, D_i , to fit the signal distribution, $p(x)$:

$$p(x) = \sum_{i=1}^n \frac{D_i}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_i}\right)^2\right] \quad (\text{eq. 2})$$

This Vibration Dose also describes the time fraction for which each Gaussian sequence exists. However it does not give any information about the sequence length and position in the signal. To provide this information, the Gaussian segments are identified in the time signal using a cumulative sum - bootstrap algorithm. It is shown that this algorithm is a robust and reliable technique to detect significant changes in the statistics of the instantaneous acceleration magnitude signal (Rouillard, 2007a, Rouillard and Sek, 2010). The results however are dependent on the signal length (window) used to calculate the RMS value (Rouillard, 2014). A short window does not represent the true nonstationarities of the signal and a long window will not detect the short-duration nonstationarities.

The analysis of vibration measurements of different vehicles and roads shows that the statistical distribution of the Gaussian segments duration follows a hyperbolic curve (see section 4.1 for more detail). These measurements represent a variety of small utility trucks, vans, rigid trucks and semi-trailers with various suspension types and payloads riding on randomly selected poorly maintained local roads, country roads, urban roads and highways (motorways), located in Victoria, Australia. Further detail on measurement conditions are given by Rouillard (2007a). These measurement samples provide a good representation of roads in Australia and the conclusions drawn could be extended to other developed countries.

A nonstationary synthetic signal is recreated using the hyperbolic distribution of the segment duration combined with the RMS value distribution (based on the Vibration Dose) to first produce a modulation function (Figure 7 b). This modulation function is then multiplied by a

random Gaussian function (created with the measurement PDS, Figure 7 a) to generate the synthetic nonstationary signal used for simulation purpose (Figure 7 c). The level of nonstationarity of a signal can be estimated using the ratio of the number of Vibration Dose to the expected number of Vibration Dose for a stationary signal (Rouillard, 2014).

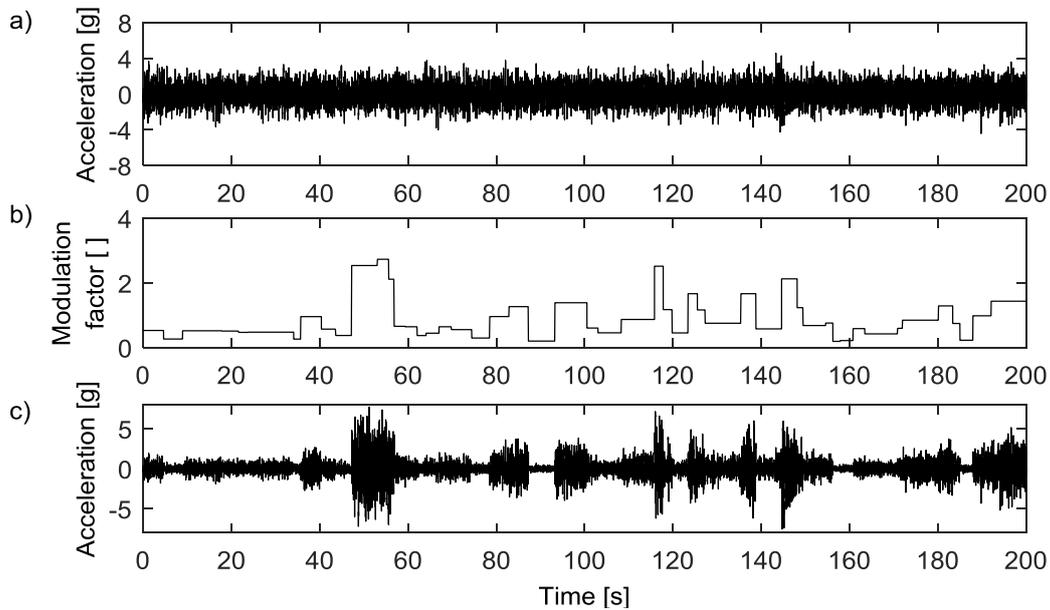


Figure 7: random Gaussian Sequence Decomposition: a) random Gaussian function; b) modulation function; c) synthetic nonstationary signal

The Random Gaussian Sequence Decomposition method, although being a significant improvement on the split spectra method, cannot properly account for randomly occurring shocks and varying harmonics occurring in real RVV signals. The method is limited to decomposing nonstationary RVV signals into shorter Gaussian segments. These segments are characterised statistically by an RMS distribution and a segment length distribution (explored in further detail in Section 4.1). Overall, the method is not appropriate for detecting and characterising shocks buried in RVV signals.

2.4.3 Vehicle and Road Characteristics-Based Simulation

Another simulation method similar to the RGSD is vehicle and road characteristics-based simulation. This method, proposed by Rouillard and Sek (2013), has the benefit of not requiring any road measurements to create a simulation. As with the RGSD, it uses a modulated random Gaussian function but is entirely based on established models.

For instance, the vehicle Frequency Response Function (FRF) is computed from parameters of different vehicle types which are readily available such as these published by Cebon (1999). The vehicle response PDS can be estimated by combining the FRF with the road elevation PDS

curve (*e.g.* ISO-8608) coupled with the vehicle speed (Lu et al., 2010, Garcia-Romeu-Martinez et al., 2008).

A transport journey vibration signal can be synthesised using known road roughness classification data and corresponding vehicle speed and segment durations. Vehicle speed dependence on the road roughness can be taken into account; higher speed on smooth roads or lower speed on rough roads.

This is a practical and affordable method to simulate RVV, but, according to the authors, it is only an approximation of the reality since the signal is based on generic PDS.

2.4.4 Wavelet Based Gaussian Decomposition

The Discrete Wavelet Transform (DWT) was used to decompose RVV signals into Gaussian components as presented by Griffiths (2012). This method exploits the good resolution in time and frequency domain afforded by the DWT to detect the Gaussian segments in the signal through an iterative process.

As shown at Figure 8, the first step of this process is to generate a random Gaussian signal based on the PDS of the vibration measured from a vehicle. The DWT is then calculated on both the Gaussian and measured signals. The envelope (maximal and minimal values) of the DWT of the Gaussian signal is used to sequence the vehicle signal; the segments of a signal are considered Gaussian when the DWT of the vehicle signal fits into the envelope. These segments are considered as the first Gaussian component and they are extracted from the signal. Once the extraction is done, a new Gaussian signal is generated from the PDS of the signal residual and the DWT functions comparison is undertaken again for a predetermined number of times. After all the iterations, the signal residual is considered as the non-Gaussian part of the signal and is fitted with the best Gaussian approximation.

The iterative loop can also be used a second time on each Gaussian component to improve their Gaussian fit. Once the Gaussianity of each component is adequate (based on a kurtosis criterion), a synthetic signal is created using the PDS and time duration of the segments of every component. Despite the fact that the duration distribution of each segment is not taken into account in the proposed algorithm, it is a good method to decompose a nonstationary signal with varying PDS between each segment.

A shortcoming of the Wavelet Based Gaussian Decomposition is that its rationale is not clearly formulated. It assumes that the spectra of the Gaussian segments are different but this hypothesis has not yet been clearly proven and it is subject to questioning. This approach has not been adopted by any other third party and therefore its effectiveness remains invalidated. As

demonstrated by Griffiths (2012), this technique does not offer any practical improvement on the Random Gaussian Sequence Decomposition technique.

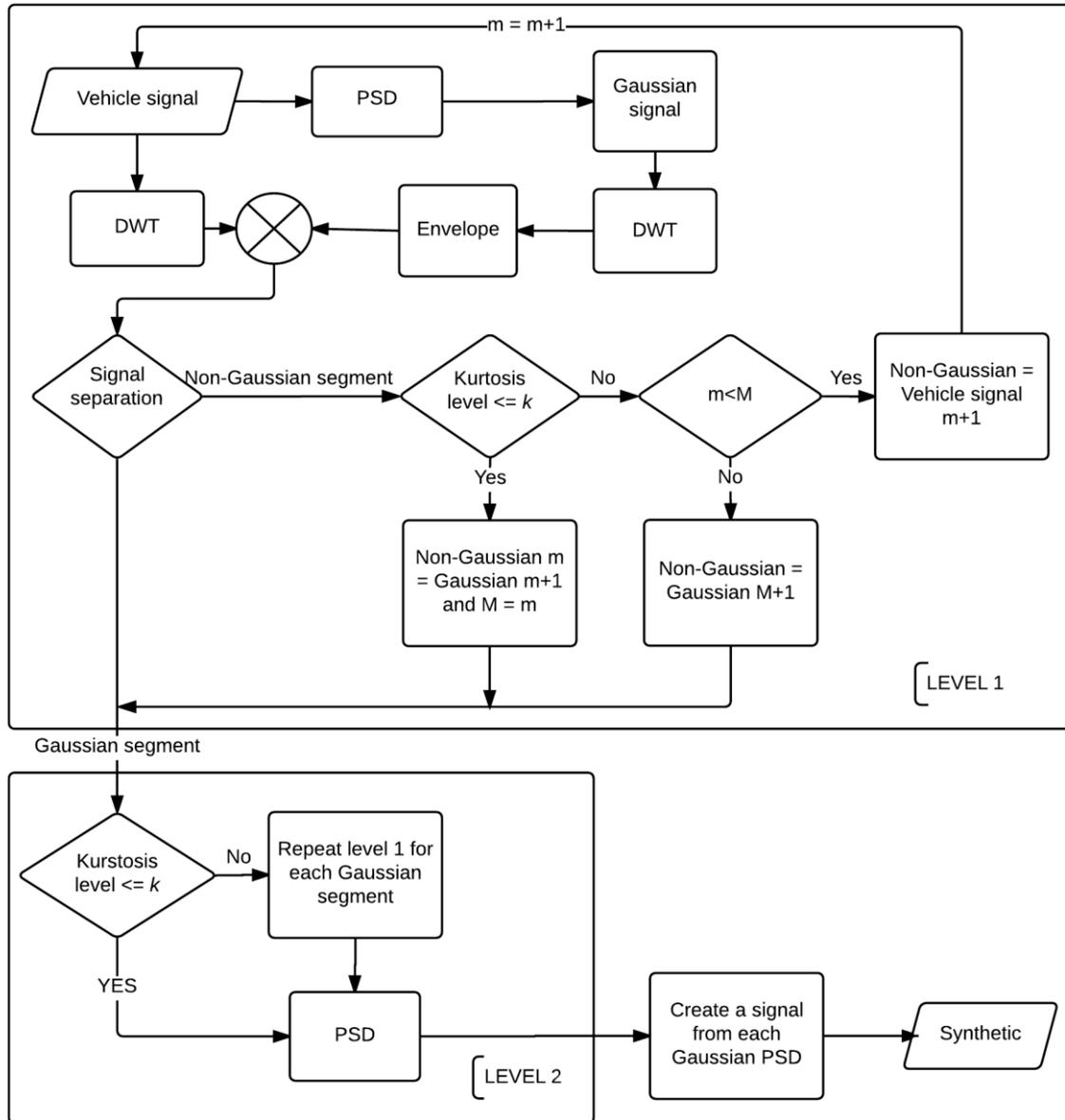


Figure 8: flow chart of the Wavelet Based Gaussian Decomposition, (reproduced from Griffiths, 2012) and PSD stands for Power Density Spectrum

2.4.5 Hilbert-Huang Transform

Another time-frequency analysis method used to analyse the nonstationary nature of RVV and road profile is the Hilbert-Huang Transform (HHT). Described in more detail in Appendix A, the HHT decomposes signals into the sum of narrow banded signals which can vary in frequency and amplitude over time. As with the DWT, the HHT has good resolution in time and frequency.

Rouillard and Sek (2005) applied the HHT to analyse railcar vibration. Using the HHT, they were able to analyse the variation in the frequency structure of the vibration which cannot be seen with the more conventional spectral averaging. Ayenu-Prah and Attoh-Okine (2009) analysed road profile using the DWT and HHT. They concluded that the HHT was the best method to analyse nonstationary road profile because as opposed to the DWT, its time resolution is independent of the frequency resolution.

The application of HHT to RVV is very recent. At about the same time as the author presented papers on the topic (Lepine et al., 2014, Lepine et al., 2015), Mao et al. (2015) developed a model of the RVV based on the HHT. As shown at Figure 9, the latter characterises a “Sampled” signal with the Hilbert spectrum (time-varying spectrum computed from the HHT). This “Target” Hilbert spectrum is then used to sequence stationary parts of the signal using changing point analysis. These segments are characterised with probability density functions which are then used to construct the “Simulated” Hilbert spectrum. Coupled with a random phase function, the Simulated Hilbert spectrum is transformed in the time domain to generate a synthetic signal with the same nonstationary proprieties than the sampled signal.

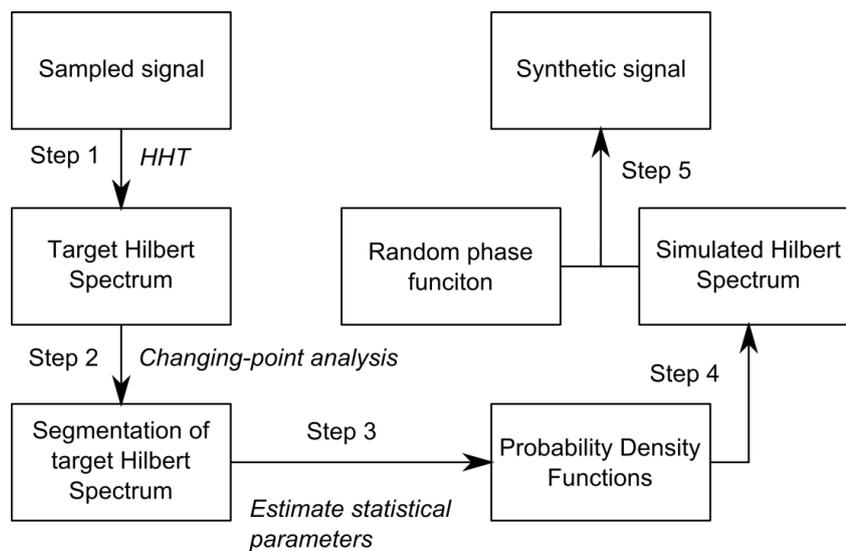


Figure 9: flow chart of the HHT RVV simulation, reproduced from Ayenu-Prah and Attoh-Okine (2009)

The application of the HHT to analyse RVV signals is promising. However, as this technique is relatively new, it is difficult to assess its performance based solely on a literature review. A more detailed critical evaluation based on its application on realistic RVV signals is presented in Chapter 6.

2.4.6 Bayesian Detector

A statistic-based approach to decompose the Gaussian sequence of RVV is proposed by Thomas (2005) who used a Bayesian detector to identify changes in the level and variance and/or

autocorrelation between successive measurement series. To ease the identification, a Box-Cox transformation (Box and Cox, 1964) is applied to the vehicle signal to change the measurement scale. An Automatic Merging for Optimal Clusters (AMOC) algorithm (Carlstein et al., 1994) is used on the transformed signal to divide the different random process of the signal. These algorithms may appear complex, but they are well known and widely used in many areas of applied statistics. One drawback of this iterative process is its risk of result instability. The algorithms also require several coefficients to be fixed. Further experimentation needs to be undertaken on this method to provide a significant assessment and guidelines to generalise its use.

2.5 Transient Events Simulation

Anyone who has ever used a road vehicle would agree that the RVV is not only defined by a series of random excitations but it also contains transient events. These events occur, for instance, when the vehicle travels over potholes, cracks, manholes, speed bumps, railway crossings and so on. Many attempts have been made by researchers to include the transient events in RVV simulations.

However, as Kipp (2001) advocates, this is far from trivial. For instance, simply adding (superimposing) impacts to a random signal has serious limitations. The user has to decide how many impacts and what amplitude to add and how to appropriately distribute them. The various shapes of the impacts also have to be defined taking into account the dynamic behaviour of the vehicle (structural modes excited by impact).

Rouillard and Richmond (2007) proposed a method to include the dynamic behaviour of vehicles when generating a synthetic impact using the IFFT of rail vehicles spectrum which was modified with a constant zero phase. This transform creates a wave packet similar to the impulse vibration response of the vehicle.

Prior to transient event simulation, their detection is also a challenge that needs to be addressed in order to characterise their occurrences in a real signal. The crest factor of the signal is used by many researchers to detect transient events (Garcia-Romeu-Martinez et al., 2007, Garcia-Romeu-Martinez et al., 2008, Steinwolf and Connon Iii, 2005, Bruscella et al., 1999, Rouillard et al., 2000, Rouillard and Sek, 2002, Rouillard, 2007a, Rouillard and Sek, 2013), which is the ratio between the absolute value of the signal and its RMS value:

$$CF(t) = \frac{|x(t)|}{x_{\text{RMS}}} \quad (\text{eq. 3})$$

This method is however highly heuristic because it is based on two subjective criteria: (1) the signal segment length used to calculate the RMS value is very important, especially for nonstationary signals and (2) the threshold defining a transient event. Bruscella (1997) proposed an enhanced method to extract transients based on the moving RMS drop-off distance and the crest factor. When these indicators reach a certain threshold a transient event is detected and the segment containing it is extracted from the signal. However the main shortcoming is that the method remains subjective because the detection relies on too many user-defined variables.

Lu and Ishikawa (Lu et al., 2008) also used a highly subjective method to detect transient events with video analysis to validate the transient detection made on the acceleration data of vehicle using *Oscope*® software (*Ono Sokki*®). Unfortunately, no further information about the software algorithm is provided. Analysis of truck vertical acceleration data and the same data with the transient events removed using the *Oscope*® software shows that shocks are more noticeable at certain vehicle speed and increase the frequency range of the vehicle response (Lu et al., 2010). Without giving a precise indication on how to detect transients, this paper highlights the importance of considering the transient and the random broad-band vibration components of a vehicle response separately.

Wavelet decomposition is proposed by Wei et al. (2005) to detect discontinuities (such as cracks and bumps) in road profile data which correspond to transient events on RVV. To do so, the road profile is decomposed into frequency sub-bands with the Wavelet. According to this study, when a discontinuity occurs in the road profile, some sub-bands are excited more than others which facilitates its detection. However, the method was only tested on discontinuities that were artificially added on road profile measurement. Therefore the frequencies contained within the profile discontinuities are determined by the user and do not necessarily represent reality. Nei et al. (2008) validated this technique on real transportation measurement and suggested that signal sub-bands decomposition could be an accurate method to characterised transient events on RVV measurements.

2.6 Harmonic Simulation Method

It is generally accepted that RVV is mainly composed of random processes (including the random occurrence of transient events) but that does not mean that harmonic components are not also present. This specific type of vibration is not induced by the road but by the vehicle itself such as engine or drivetrain-related vibration and unbalanced wheel(s) vibration (Gillespie, 1985). While harmonic vibration is well known in railway vehicle vibration (ISO-13749), to the best knowledge of the author, they are only briefly discussed in term of road vehicles by Charles (1993). It is suggested that waterfall plots could be used to study vehicle speed dependency to certain vibration, but no application is presented.

2.7 Conclusion on Simulation Methods

RVV simulation methods have been studied for several decades, and even though significant improvements on the earliest simulation methods have been made, the current methods contain some significant limitations. As described in this literature review, it is recognised that the Gaussian PDS-based (standardised), time replication and the non-Gaussian methods are not sufficiently accurate to realistically simulate RVV because of their statistical limitations and their discrepancies with respect to the real nature of the vibration.

Despite this major shortcoming, the Gaussian PDS-based method remains broadly used today to simulate RVV probably because it is recommended by several international standards. This is a proven practice to ensure that the packaging will be effective, but it is far from being the best method for optimisation. Its simplification of the nature of RVV generally increases the severity of the excitation. Therefore packages that survive the standardised simulation are more likely to be over-designed.

A good simulation method should reproduce at least the two dominant modes present in RVV, *i.e.* the nonstationary random and transient components. So far, a number of effective methods have been proposed to characterise the nonstationarities, such as: the Random Gaussian Sequence Decomposition (Rouillard, 2007a), Wavelet Based Gaussian Decomposition (Griffiths, 2012) and the Bayesian Detector (Thomas, 2005). But no definitive method has been developed and validated to identify and characterise the other modes of RVV.

Two elements should therefore be improved to provide more accurate RVV simulations. First, transient and harmonic signal components must be well characterised, and second, a generalised method including at least the nonstationary and transient modes must be developed.

Chapter 3: HYPOTHESIS AND OBJECTIVE

Many methods have been developed to characterise and simulate the RVV. However, the common shortcoming of these methods is that they consider only one mode of the RVV. For instance, the models that describe nonstationarity of the RVV signals neglect the presence of transient and harmonic components. None of the aforementioned methods consider RVV as mixed-mode signals composed of nonstationary, transient and harmonic components.

The main hypothesis of the thesis is that the distinct modes (or components) constituting the vibration produced by road vehicles can be reliably detected and separately identified.

The main objective of the thesis is to develop a technique to identify the two dominant components within a RVV signal. Once identified, each of the components can be individually analysed using the appropriate method. RVV could be separated into mono-mode signals; *i.e.* the nonstationary random and transient vibration each in a different signal. The characterisation of each component would then be improved which means RVV could be more accurately simulated.

3.1 Definition of the RVV Modes

Throughout the thesis, it is assumed that RVV signals can be decomposed into three modes. These modes are also called components to avoid any confusion with the usage of the word “modes” in modal analysis. There is not universal definition for these components and the following subsections define them in the context of the thesis.

3.1.1 Nonstationary Components

The definition of Rouillard (2007a) on RVV nonstationarities is adopted. It defines nonstationary components as a sequence of Gaussian random signals of different intensities (RMS values). These piecewise intensity variations increase the kurtosis value of the signal. All the Gaussian sequences have the same spectral shape.

The nonstationary components represent the variation of pavement profiles and vehicle speed. For instance, a vehicle will produce different levels of Gaussian random vibration at the same speed on two different roads and vice versa.

3.1.2 Transient Components

The transient components are referred as shocks in the RVV signal. Harris et al. (2002) define these transient components as “a temporarily sustained vibration of a mechanical system. It may consist of forced or free vibration or both”. However, this definition makes the difference between nonstationary and transient components ambiguous. The transient components are just defined in terms of time; so a very short Gaussian segment could be identified as shocks.

It is therefore important to introduce a notion of the signal’s profile or shape in the definition. For instance the more recent definition of transients by Bendat and Piersol (2011) is more appropriate:

“A special class of nonstationary data are those that physically exist only within a finite, measurable time interval, that is, where the input process $x(t)$ and the output process $y(t)$ have nonzero values only for $0 \leq t \leq T$. Such data are commonly referred to as *transients* and allow for a greatly simplified analysis because for a pair of sample records $x(t)$ and $y(t)$ with zero mean values,

$$\begin{aligned} X(f) &= \int_0^T x(t) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \\ Y(f) &= \int_0^T y(t) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} y(t) e^{-j2\pi ft} dt \end{aligned} \quad \text{” (eq. 4)}$$

This definition of transient components does not qualify a short Gaussian segment as a shock because by definition it is not finite; *i.e.* the Gaussian sequence does not decay by the end of the segment.

Ogata (1995) proposes a definition that clearly makes the distinction between the transient and nonstationary components: “The transient response refers to that portion of the response due to the closed-loop poles of the system, and the steady-state response [in this case the nonstationary components] refers to that portion of the response due to the poles of the input or forcing function.”

The latter definition is the more appropriate for the RVV modes identification. The transient components referred to as shocks are therefore defined as a vibration composed of the vehicle’s impulse response (transfer function poles) caused by a short duration and finite excitation.

3.1.3 Harmonic Components

The harmonic components constitute the overlooked mode of RVV. It refers to any periodic vibration. They principally come from the vehicle itself. They are potentially generated by wheels, drivetrain and engine. The periodicity of the vibration depends on the vehicle's speed and the engine's revolutions. They can be described by time variant sinusoidal function such as:

$$X(t) = a(t) \sin \left[2\pi \int_0^t f(t) dt \right] \quad (\text{eq. 5})$$

where $a(t)$ and $f(t)$ are respectively the time variant amplitude and frequency.

For well-maintained vehicles, the action of this mode is negligible compared to the other two modes. The predominance of harmonic components may even indicate a vehicle's malfunction. Because of this lower importance, the harmonic components are only considered in the first part of the thesis (Chapters 2 to 6). To respect the timeframe of the candidature, the development of the machine learning classifiers was simplified and only the nonstationary and transient components are considered in the second part of the thesis (Chapters 7 to 9). The development of machine learning classifiers capable of detecting harmonic components in RVV signal is left for future research.

3.2 Sub-Objectives

The thesis has the objective of developing a technique to identify these components within a RVV signal and more precisely the nonstationary and transient components. This main objective can be separated by these sub-objectives:

- RVV Analysis Methods Review - to review the different methods that could be used to identify the different RVV modes;
- Machine Learning Integration - to integrate the best nonstationary and transient analysis methods into a machine learning process to develop classifiers which can detect shocks;
- Evaluation and Optimisation of Machine Learning Classifiers - to evaluate and optimise the performance of different machine learning classifiers on the detection of shocks in synthetic RVV signals;
- Real Vehicle Implementation - to develop a machine learning process which trains classifiers on synthetic signals and can detect shocks buried in real RVV signals.

These sub-objectives constitute the main contribution of the thesis by their significance and originality. In order to carry-out these evaluations and develop the detection algorithm, calibration signals with the RVV properties are needed. The creation of such signals is described in the following chapter.

Chapter 4: SYNTHETIC ROAD VEHICLE VIBRATION SIGNALS

The Road Vehicle Vibration (RVV) is qualified as mixed-modes signals due to their multiple natures. In order to synthesise realistic RVV signals, the nature of these modes needs to be established. During a typical transportation journey the road roughness profile and the vehicle speed vary, which creates nonstationary random vibration; the road profile discontinuities and aberrations also create transient vibration (shocks) and the vehicle itself generates harmonic vibration due to its drivetrain system. These components are characterised by different vibration modes which have to be included in RVV simulation because they can cause different types of damage to freight. The nonstationary random vibration is more likely to cause fatigue failure because it induces a highly repetitive low level of stress. The transient vibration is more likely to cause sudden failures because its occurrence is less frequent than the random vibration but its stress level could be much higher. Failures caused by harmonic vibration depend on the dynamic behaviour of the freight. When the harmonic frequency is close enough to the natural frequency of freight, the stress level will increase in time because of the resonance phenomenon. Depending on the level of damping in the structure, a small harmonic excitation at natural frequency can lead to catastrophic damage.

To accurately replicate the RVV and simulate its effect on freight, these three vibration modes have to be considered. A simulation signal should have a similar degree of nonstationarity and level of transient and harmonic components to those of typical RVV signals. The main challenge here is to extract these modes in a signal in order to characterise them. This can be achieved using different signal analysis methods, which will be addressed in the following chapters of the thesis. Since these methods were developed for different applications and types of signals, it is important to assess their performance on the three specific modes of RVV.

Synthetic signals are more appropriate than real RVV signal to evaluate these methods because the RVV modes composing the signal can be precisely defined and known. To represent all three of them, the synthesis is made in three different sections (Figure 10). Each section is

independent and its characteristics can be modified to generate different types of road vehicle signal, in order to make it possible to assess the performance of different analysis methods on each mode separately or simultaneously.

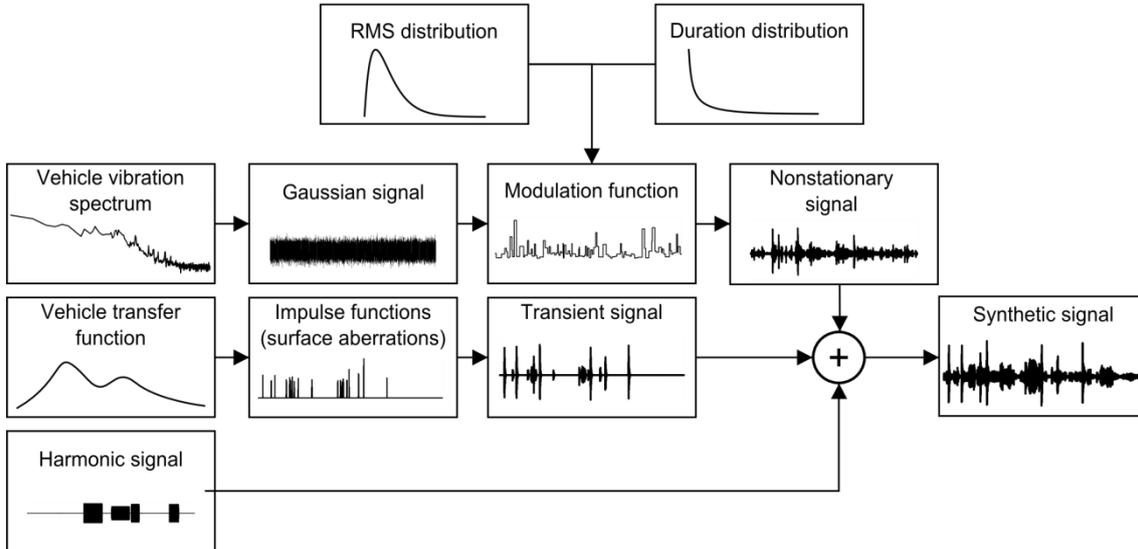


Figure 10: schema block of the signal synthesis

To give more detail on the mechanics of the mixed-modes signal synthesis, this chapter explains how each mode of the RVV is synthesised and how they individually affect the signals' overall characterisation.

4.1 Nonstationary Signal Synthesis

A simple way to synthesise RVV would be to generate a road profile with a nonstationary roughness profile superimposed with short negative and positive impulse functions to represent the transient components. This road profile combined with a generic vehicle transfer function creates a RVV signal. However the major drawback of this technique is that it is difficult to know the RMS value of the nonstationary segments once the signal has been synthesised. Therefore another technique is used based on the work of Rouillard (2007a) on the synthesis of non-Gaussian RVV. The basic idea is to modulate a Gaussian PDS-based signal to create a nonstationary signal. The Gaussian signal is synthesised from PDS of RVV. Only the shape of the spectrum is needed at the beginning because the signal will be rescaled subsequently. Therefore the PDS (P_{xx}) is normalised with the signal's power (*i.e.* RMS^2):

$$\hat{P}_{xx}(f) = \frac{P_{xx}(f)}{\sum P_{xx}(f)\Delta f}, \quad (\text{eq. 6})$$

where Δf is the frequency resolution of the PDS. This normalised PDS is transformed into an amplitude spectrum with a random phase, ϕ , spectrum uniformly distributed between $-\pi$ and π :

$$X(f) = \sqrt{(P_{xx}(f)\Delta f)}e^{j\phi}. \quad (\text{eq. 7})$$

This spectrum is transformed to the time domain using the Inverse Fast Fourier Transform (IFFT). The resulting time signal is a Gaussian random signal with an RMS value of 1 (from the normalisation). The signal becomes nonstationary when multiplied by a modulation function representing the variation in the RMS value of the signal, Figure 10.

As seen in Figure 11, the modulation function has a direct effect on the signal distribution. The more the modulation function has segments with different RMS values, the more leptokurtic (high kurtosis value) is the Probability Density Function (PDF) of the signal; *i.e.* the signal distribution shows that high amplitude accelerations occur more often than with a normal distribution. This explains why RVV signals are mostly leptokurtic.

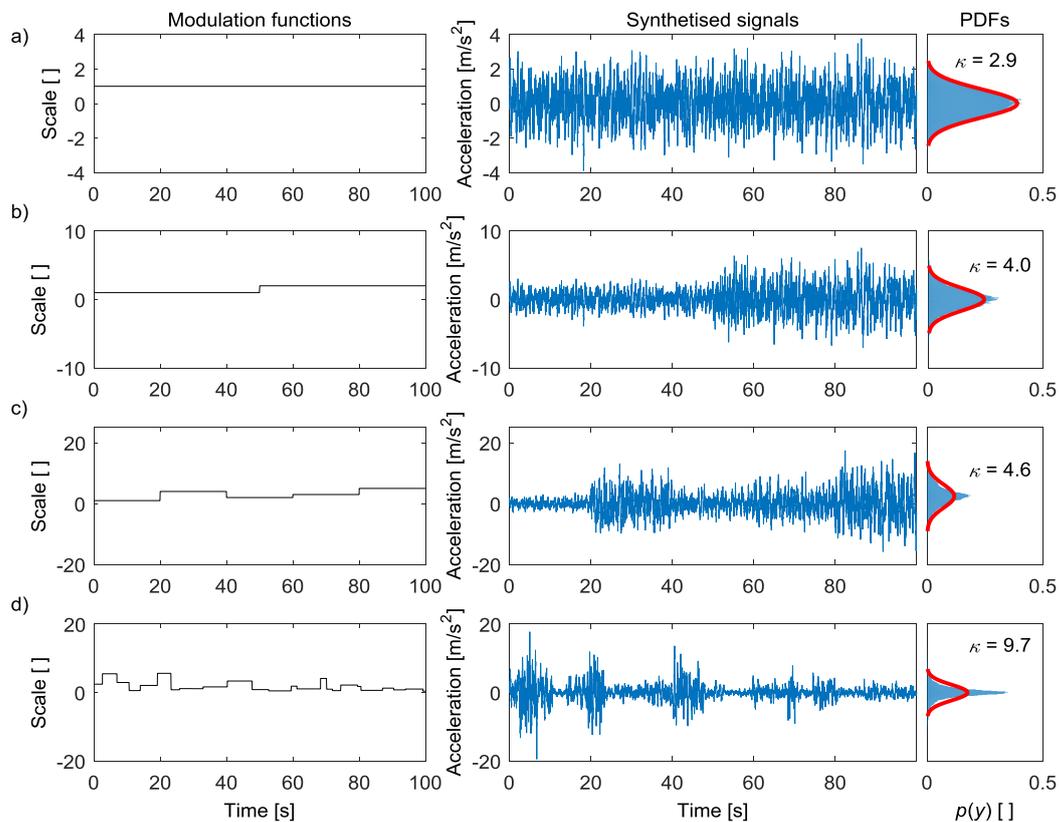


Figure 11: effect of nonstationarities on a Gaussian random signal with: a) constant RMS value (stationary); b) two segments with different RMS values and same duration; c) five segments with different RMS values and same duration; and d) 30 segments with different RMS values and durations; where the red line superimposed on PDFs is a Gaussian fit and κ is the kurtosis value of the distribution

Rouillard (2007a) studied the intensity and duration distribution of the Gaussian segments composed in typical RVV signals. The segments' intensity distribution, *i.e.* the distribution of

the RMS value of every segment, can be represented by a two-parameter Rayleigh function defined as:

$$P(m) = \frac{m}{(\beta\sigma)^2} e\left(-\frac{1}{2}\left(\frac{m}{\beta\sigma}\right)^\alpha\right) \quad (\text{eq. 8})$$

where m is the RMS value of the segment, σ the standard deviation, α the exponential parameter and β the scaling factor. Typical distribution is presented in Figure 12.

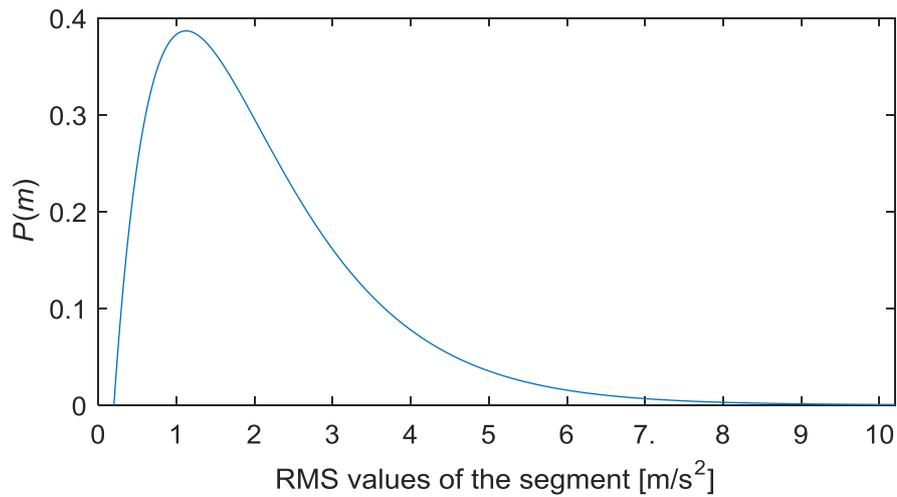


Figure 12: typical probability distribution of the RMS values of RVV Gaussian segments (based on Rouillard, 2007a)

The duration of the Gaussian segments was also studied by Rouillard. The distribution of the duration is modelled by a hyperbolic function as:

$$p(d) = \frac{C}{\sinh(kd)} \quad (\text{eq. 9})$$

where d is the segments' duration and C and k empirical constants. A typical distribution is presented in Figure 13.

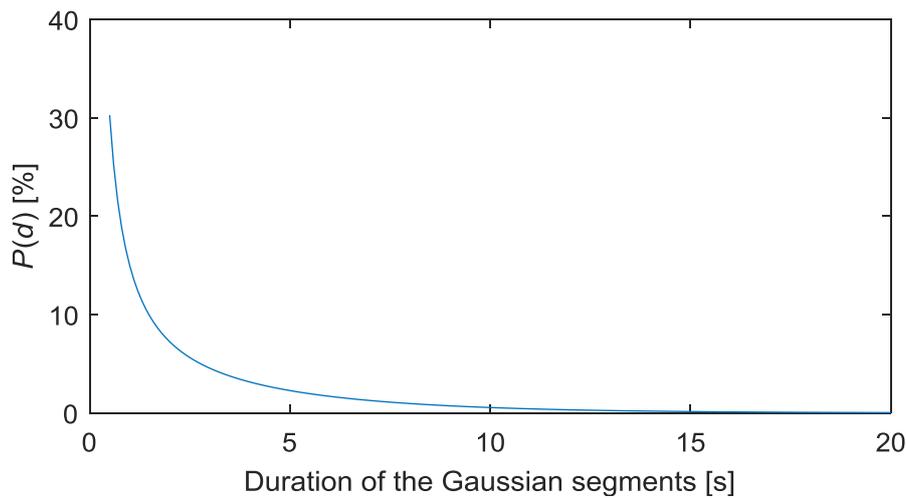


Figure 13: typical distribution of the duration of RVV Gaussian segment (based on Rouillard, 2007a)

Both distributions can be scaled to recreate different synthetic nonstationary signals which represent realistic RVV. These nonstationary signals are used throughout the thesis.

4.2 Transient Signal Synthesis

The transient signal represents the response of the vehicle over short time duration and high amplitude excitations such as potholes or bumps present on the road. This can be represented by the impulse response of the vehicle. The synthesis of the signal relies on the dynamic model of a typical vehicle.

A two degree-of-freedom model of the vehicle, called a quarter-car model, was used for the transients synthesis (Figure 14). The model is composed of a sprung mass M_s , an unsprung mass M_u , two springs (k_s and k_u) and two dampers (c_s and c_u). The numerical values of components used in the model represent typical values used by Cebon (1999) for “quarter-car” truck model, Table 1. The model input, $x(t)$, is the road profile and the output, $\ddot{y}(t)$, is the vehicle sprung mass acceleration. The FRF of the model (Figure 15) shows that the first peak is less damped than the second. The response decreases after the second peak and its magnitude is less than 10 % of the maximum value above 30 Hz.

In the thesis, what is called a shock is the quarter-car response to an impulse function such as a raised single period cosine function. This impulse function can have different amplitude and duration representing different road aberrations.

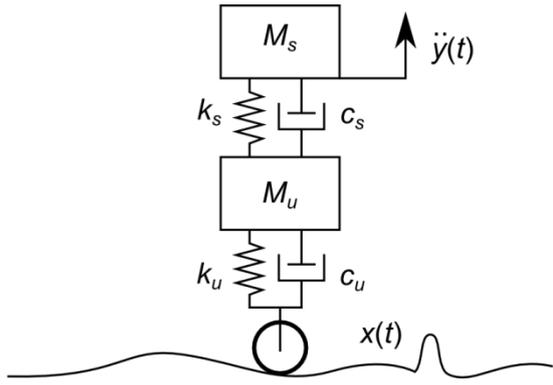


Table 1: components of the two degree-of-freedom model of a typical truck (Cebon, 1999)

Component	Value
M_s	8900 kg
M_u	1100 kg
k_s	2000 kN/m
k_u	3600 kN/m
c_s	40 kN s/m
c_u	4 kN s/m

Figure 14: two degree-of-freedom vehicle model

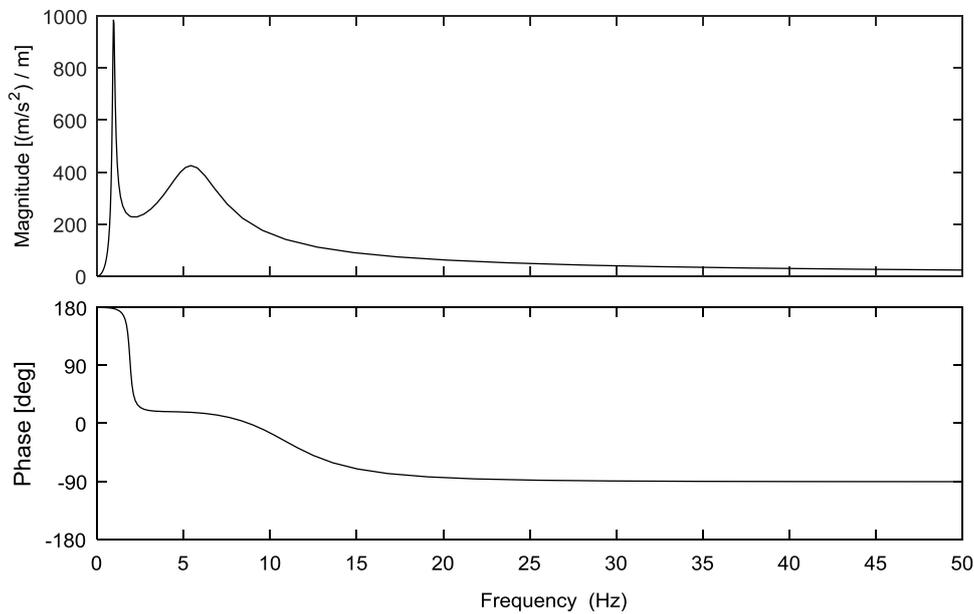


Figure 15: two degree-of-freedom model of a typical truck

Detecting shocks when they are the only component in a signal is relatively easy. A detector based on a single value threshold will locate the transient. However the task is more complex when the shocks are superimposed onto a Gaussian random signal of similar amplitude. The task becomes even more complex when the background random signal is nonstationary. A single value threshold cannot dissociate the transient from the signal. Therefore, it is pointless to generate transient signal of same amplitude and duration. To assess the time-frequency analysis tools, the transient signals are superimposed on Gaussian signals of different relative amplitude. The relative amplitude and the duration of the impulse used to synthesise the transient event superimposed on a Gaussian signal have a direct effect on the signal distribution. As seen in Figure 16, a short duration impulse (0.01 s) does not create shocks that modify the signal distribution; a medium duration impulse (0.1 s) transforms the normal distribution of the signal

into a leptokurtic distribution and an impulse of a longer duration (1 s) skews the distribution. The amplitude of the impulse has a direct effect on the kurtosis value of the distribution.

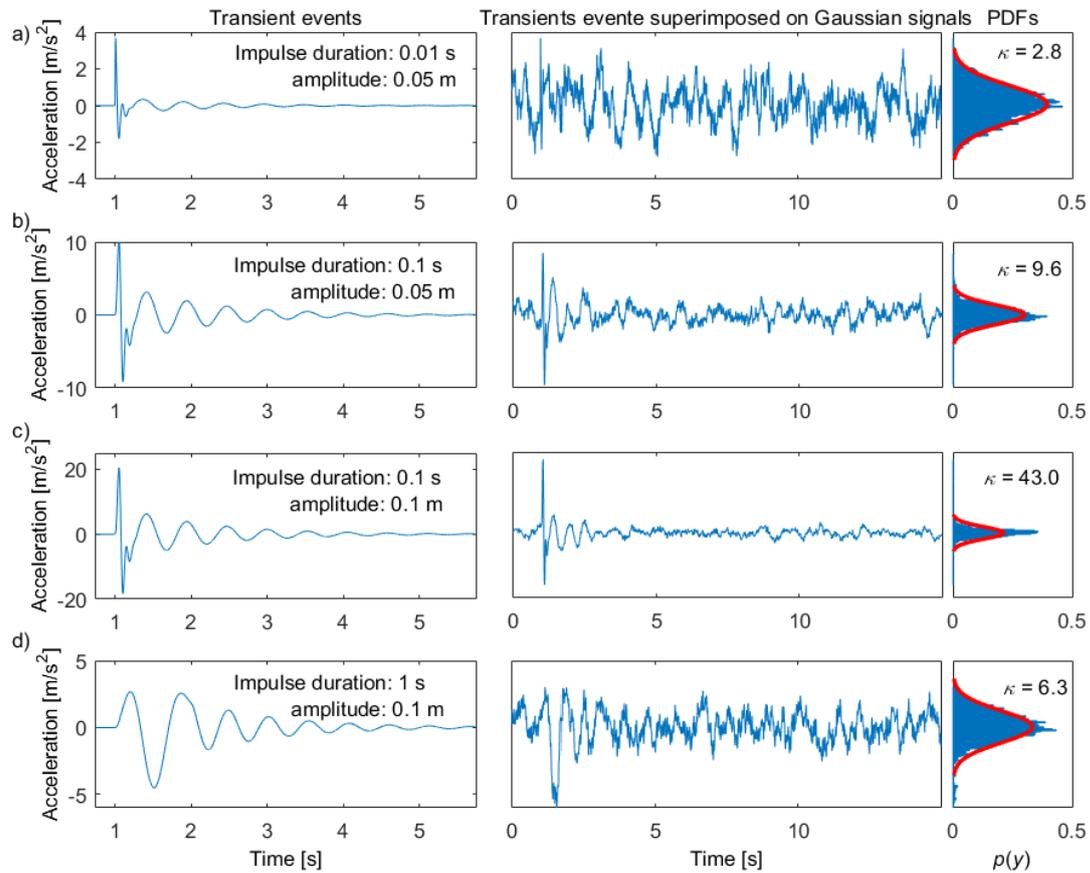


Figure 16: effect of the impulse duration and amplitude of a transient event and its superimposition on a Gaussian signal ($\text{RMS} = 1 \text{ m/s}^2$), where the red line superimposed on PDFs is a Gaussian fit and κ is the kurtosis value of the distribution

These shocks can be characterised with three distributions: impulse amplitude, impulse duration and time interval between shocks. Since there is no *a priori* information on the distributions of the shocks in RVV signals, they have been assumed Gaussian for the purpose of synthetic transient signal generation.

4.3 Harmonic Signal Synthesis

The harmonic components of the RVV come from the vehicle itself and more precisely the drivetrain system. The main frequency range of these harmonics can therefore be estimated from the vehicle speed and characteristics. For instance, an unbalanced wheel with a tyre diameter of 800 mm generates sinusoidal excitation from 6 Hz to 11 Hz between 50 km/h and 100 km/h and the first excitation mode of an engine at 500 to 2000 revolutions per minute is 8 Hz to 33 Hz. To represent the vehicle's intrinsic vibration, a synthetic harmonic signal

sinusoidal waveform was made to vary between 6 and 30 Hz. Because the speed of the vehicle and the engine vary, the amplitude A and frequency f of the waveform can also be time variant (eq. 5):

$$x(t) = A(t) \sin \left(2\pi \int_0^t f(t) dt \right).$$

It can be constituted of only one waveform or also of waveform harmonics to represent vibration emerging from varied sources along the drive train:

$$x(t) = A_1 \sin(f 2\pi t) + \sum_n A_n \sin(nf 2\pi t), \quad (\text{eq. 10})$$

where n is the order of the harmonics.

As with the transient events, the detection of a mono-component harmonic signal is not a challenge. The Fourier transform, or the Hilbert transforms in the case of time variant waveform, are well-designed to characterise this type of signal. To assess the time-frequency analysis tools, the harmonic signals are superimposed onto Gaussian random signals in order to evaluate their performance on more complex applications. As seen in Figure 17, a harmonic signal of the same amplitude as the Gaussian signal RMS value (1 m/s^2) does not greatly affect the signal distribution. When the harmonic signal amplitude is twice the Gaussian signal RMS the distribution becomes platykurtic (low kurtosis value). The kurtosis of a sine wave being 1.5, the kurtosis should never reach such low value as harmonics are not predominant in RVV.

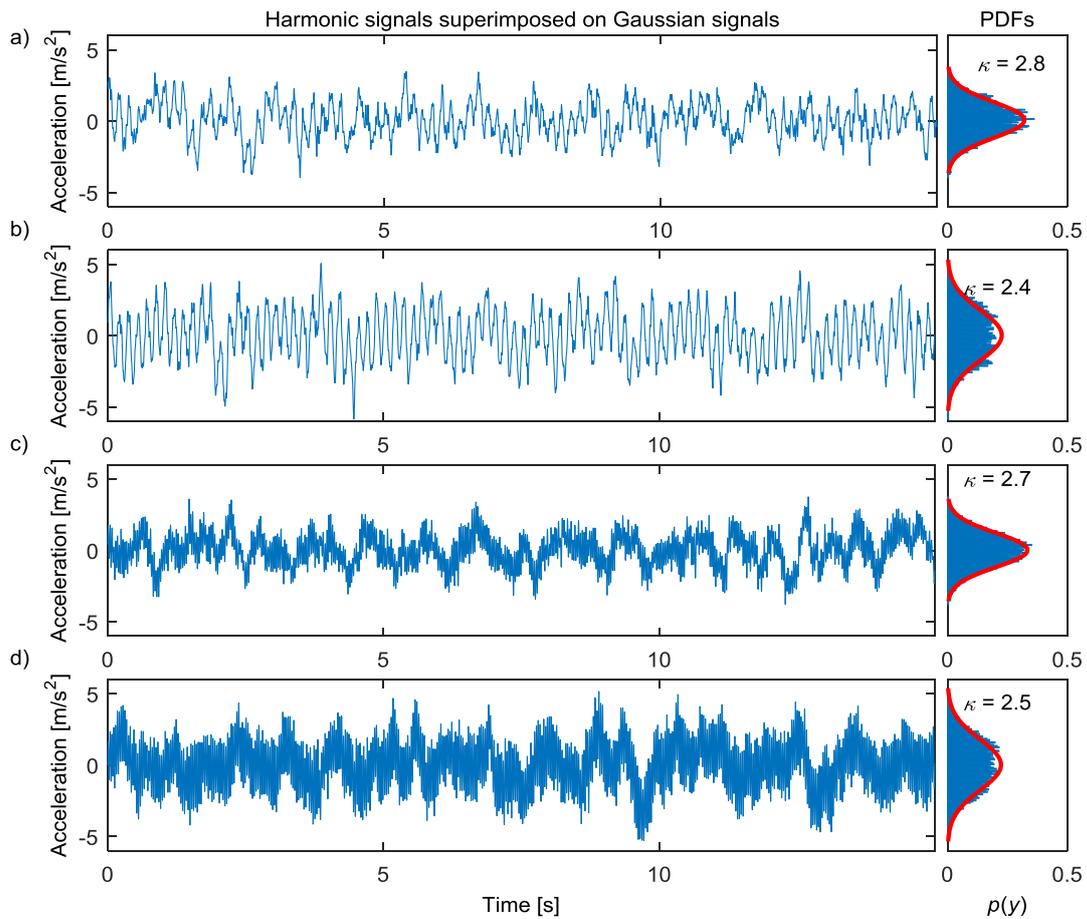


Figure 17: Gaussian random signal ($\text{RMS} = 1 \text{ m/s}^2$) superimposed by: a) a 1 m/s^2 6 Hz sine wave; b) a 2 m/s^2 6 Hz sine wave; c) a 1 m/s^2 30 Hz sine wave; d) a 2 m/s^2 30 Hz sine wave; where the red line superimposed on PDFs is a Gaussian fit and κ is the kurtosis value of the distribution

4.4 Conclusion on Synthetic RVV Signal

The proposed synthesiser is able to generate signals with the three modes present in RVV. The characteristics of each mode can be controlled which makes these signals a perfect evaluation tool to assess the signal analysis methods used to study these modes. Signals with only one or two modes at different levels can be generated to assess how accurately different methods can identify and characterise the components. The components have different effects on the signal distribution which is described with the kurtosis. The level of random nonstationarity increases the kurtosis value and the presence of harmonic components decreases it. The effect of transient events on the kurtosis is not as simple. It generally increases it but the magnitude depends on the impulse amplitude and duration.

These synthetic signals are a simplification of real RVV signals and they only consider the predominant vehicle dynamic behaviour. Therefore, the best performing methods are validated on real RVV signals at the end of the thesis.

Chapter 5: TIME DOMAIN ANALYSIS

The three modes contained within RVV signals cannot be analysed with the same statistical tools. The transient and harmonic components have to be separated from the underlying nonstationary random vibration. As was revealed in the literature review in Chapter 2, a myriad of existing signal analysis techniques in the time domain can be used to identify each of the modes. The objective of this chapter is to review the merits and limitations of the most promising techniques, such as the crest factor and moving statistics. It must be noted that the results presented in this chapter are typical representation of numerous analyses performed on the topic.

5.1 Crest Factor

The crest factor is the ratio between the absolute value and the intensity of a signal. The intensity is represented by the Root Mean Square (RMS) value of the signal $x(t)$ of duration D ,

$$x_{RMS} = \sqrt{\frac{1}{D} \int_0^D x(t)^2 dt} . \quad (\text{eq. 11})$$

The crest factor (eq. 3) is calculated from the absolute value of the signal's extremum divided by the RMS value,

$$CF(t) = \frac{|x(t)|}{x_{RMS}} .$$

By definition, the crest factor is sensitive to values standing out from the other values of a signal. For instance, it is used in machine monitoring to quantify impulsive loads and detect bearing wear. In RVV application, the crest factor is used to detect shocks (transient components). However, the nonstationary nature of RVV signals reduces the effectiveness of this technique. The presence of shocks increases the crest factor values but the variations in the signal intensity also increases it making it difficult to differentiate between the two.

Figure 18 shows a Gaussian random sequence of 20 s duration with an average RMS value of: 20 m/s² from 0 to 8 s, 10 m/s² from 8 s to 15 s and 25 m/s² from 15 s to 20 s, superimposed with five transient events (vehicle response to 0.05 m of amplitude and 0.05 s duration impulse functions located at 1, 5, 6, 10 and 12 s). The crest factor value of this signal increases at the shocks but also when the local RMS value is above the RMS of the full signal (eq. 11). The shocks occurring during the 10 m/s² RMS segment (8 s to 15 s) could be easily detected. However, the crest factor is not different between the 20 m/s² RMS segment superimposed with shocks and the 25 m/s² segment without shocks. The crest factor in its current form cannot effectively be used to analyse nonstationary signals because it is a statistic that is computed on the whole signal. To make it more effective in this situation, the crest factor must be computed from a shorter segment of the signal. This is where the moving statistics could be useful.

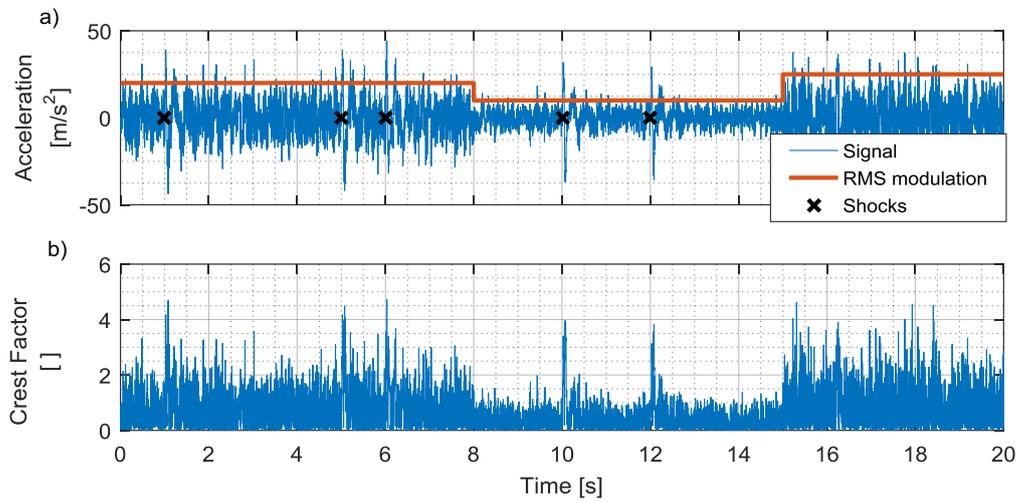


Figure 18: a) nonstationary signal superimposed with 5 shocks b) the crest factor

5.2 Moving Statistics

Most of the common statistical tools are designed to work with a stationary signal. A signal is considered stationary in the wide sense when its first and second moments are time invariant; *i.e.* when its mean and variance do not vary in time (Bendat and Piersol, 2011). The RVV signals have constant zero mean, but their variance is not constant. If the signal mean μ is zero, the RMS value (eq. 11) equals the square root of the variance, σ^2 ,

$$\sigma^2 = \frac{1}{D} \int_0^D [x(t) - \mu]^2 dt . \quad (\text{eq. 12})$$

When the synthetic RVV signals are generated with segments of different RMS values they are therefore nonstationary. However these segments can be considered stationary. Measuring the statistics of a single moving segment (window) of the signals could therefore be useful.

5.2.1 Moving RMS

Measuring the fluctuations in the signals' RMS values gives an overview of their nonstationarities. This can be performed with the computation of the moving RMS of a window length, T ,

$$m_{\text{RMS}}(t) = \sqrt{\frac{1}{T} \int_0^T x(t+\tau)^2 d\tau} . \quad (\text{eq. 13})$$

Note how the moving RMS is calculated “forward” ($t+\tau$) to consider the vehicle causal reaction from the road excitation.

The selection of this window length is very important. A shorter window performs better with a short nonstationary segment than a longer window, but is ineffective for a long nonstationary segment where the longer window performs best. For instance Figure 19 shows that a 0.1 s window length is too short to analyse this nonstationary RVV signal because the RMS value varies with the signal's peaks and troughs. On the other hand, the 8 s window is too long to reveal the signal's real RMS variation. Figure 19 shows that the more appropriate window length for RVV application is within 0.5 s and 4 s. The same conclusion has been made by Rouillard (2014) when quantifying the nonstationarity of vehicle vibration.

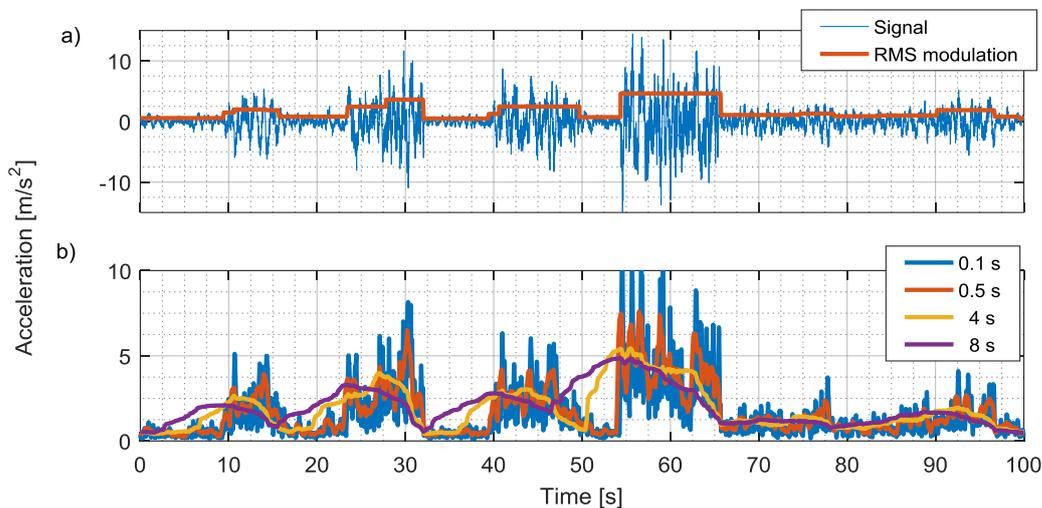


Figure 19: effect of moving RMS window length; a) 100 s sample of a nonstationary signal; b) moving RMS functions

Transient components are short and have high amplitude so they have more effect on the moving RMS of a shorter window period. For example, the 0.5 s moving RMS function has bigger responses to shocks buried in a nonstationary signal than the 4 s moving RMS function (Figure 20). However, when the window length is too short (*e.g.* 0.1 s) the moving RMS

becomes too sensitive to any signal amplitude variation and is inadequate to detect the presence of shocks. For the synthetic RVV signal used in the analysis, the ideal moving RMS window length for shock detection is approximately 0.5 s. Whereas, the 4 s window is better to quantify the nonstationarity because it has good RMS variation sensitivity without being too affected by transient components.

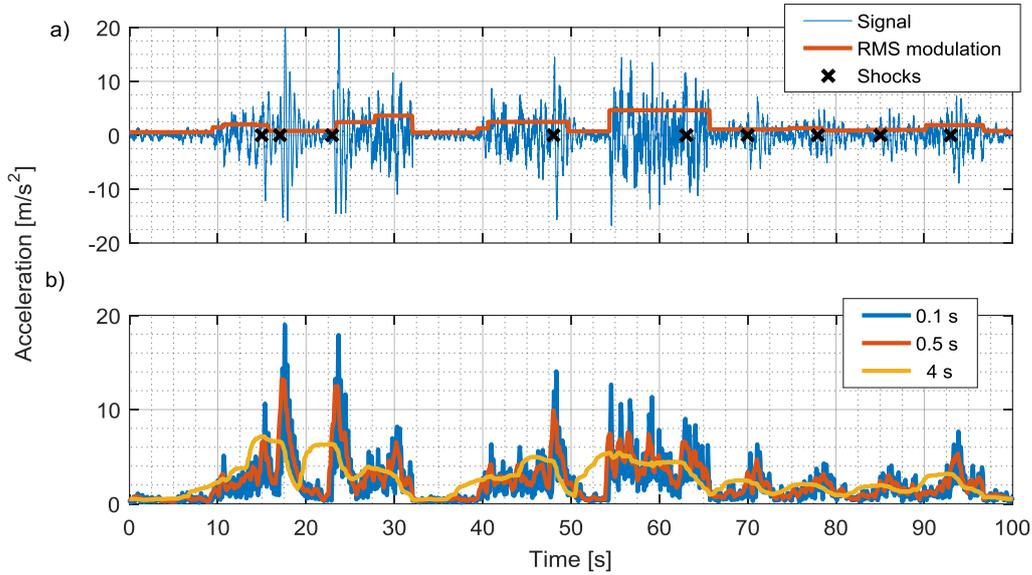


Figure 20: effect of moving RMS window length; a) 100 s sample of a nonstationary signal superimposed with shocks; b) moving RMS functions

5.2.2 Moving Kurtosis

The second moment, the variance or RMS^2 value in the case of a zero mean signal, quantifies the intensity (dispersion) of a signal. When increasing in order, the third moment, the skewness, quantifies how evenly the data is distributed around the mean and the fourth moment, the kurtosis, quantifies how uniformly distributed the data are. Two signals with the same RMS value can have different kurtosis if one has a few very high peak values. The kurtosis could also be interpreted as how close a signal amplitude distribution is to a Gaussian distribution.

The kurtosis of signal duration, D , and mean, μ , is defined as:

$$\kappa = \frac{\frac{1}{D} \int_0^D [x(t) - \mu]^4 dt}{\left(\frac{1}{D} \int_0^D [x(t) - \mu]^2 dt \right)^2} \quad (\text{eq. 14})$$

The kurtosis of stationary Gaussian signals is equal to 3. Signals with peak values of high amplitude have a kurtosis greater than 3 and are called leptokurtic. This is the case for signals

with transient events. As shown in Chapter 4, nonstationarity can also explain a high kurtosis value. Signals with kurtosis below 3 are called platykurtic. Uniformly distributed data have a kurtosis of 1.8 and the cosine and sine functions (harmonic components) have a kurtosis of 1.5.

The three modes of RVV signals affect the kurtosis of a signal. The nonstationary and transient components increase its value and the harmonic components reduce it. Therefore the kurtosis in its moving form could be useful to identify each component individually:

$$\kappa(t) = \frac{\frac{1}{T} \int_0^T [x(t+\tau)]^4 dt}{\left(\frac{1}{T} \int_0^T [x(t+\tau)]^2 dt \right)^2} \quad (\text{eq. 15})$$

Note how the moving kurtosis is calculated “forward” ($t + \tau$) to consider the vehicle causal reaction from the road excitation.

As for the moving RMS, the window length of the moving kurtosis has an important effect on the mode detection sensitivity. Window lengths of 4 s and 8 s show similar sensitivity to nonstationarities (Figure 21). As seen more in detail in Figure 21 c, their kurtosis value increases at the end of important Gaussian random segments. The kurtosis value of the 16 s window does not vary much with the change in Gaussian segments.

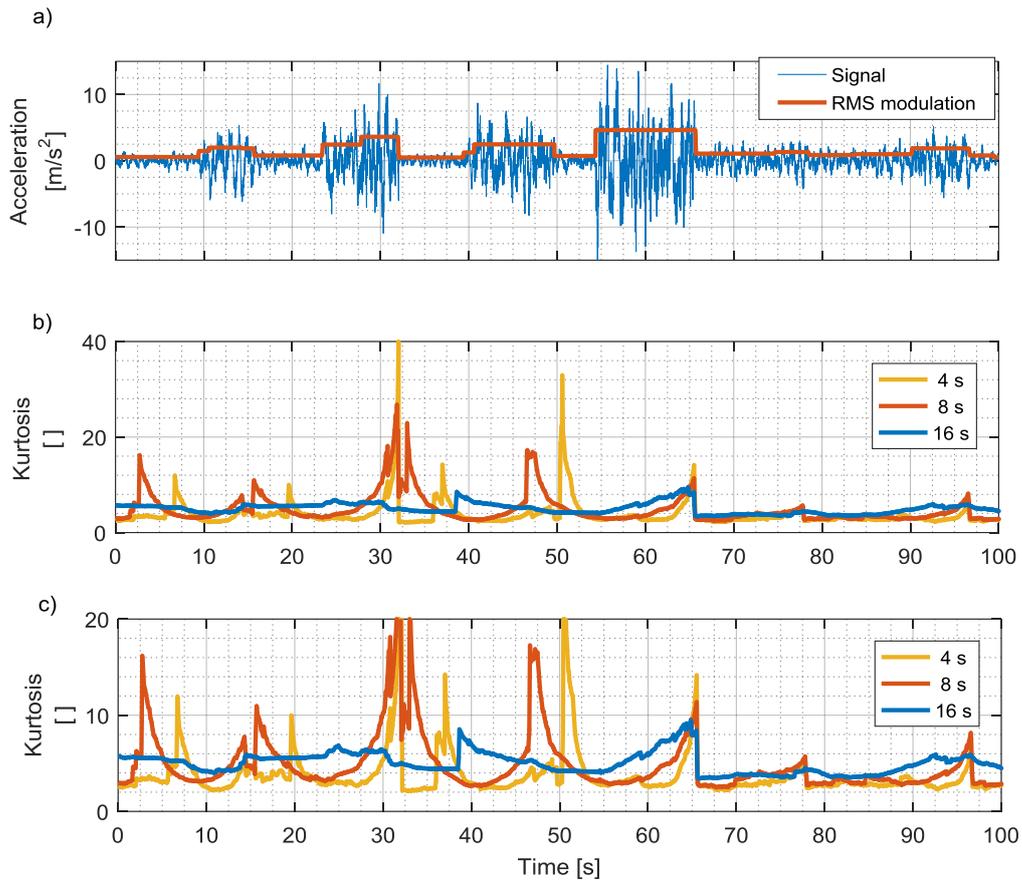


Figure 21: effect of moving kurtosis window length; a) 100 s sample of a nonstationary signal; b) full moving kurtosis functions (0-40 y-axis scale); c) moving kurtosis functions zoomed on 0-20 y-axis scale

Discrepancies appear between the 4 s and 8 s windows when shocks are superimposed onto the same nonstationary signal (Figure 22). With the shocks, the 8 s window has a more constant kurtosis value than the 4 s window. For the 16 s window, the moving kurtosis is too broad and does not have much sensitivity to either nonstationarity or shocks.

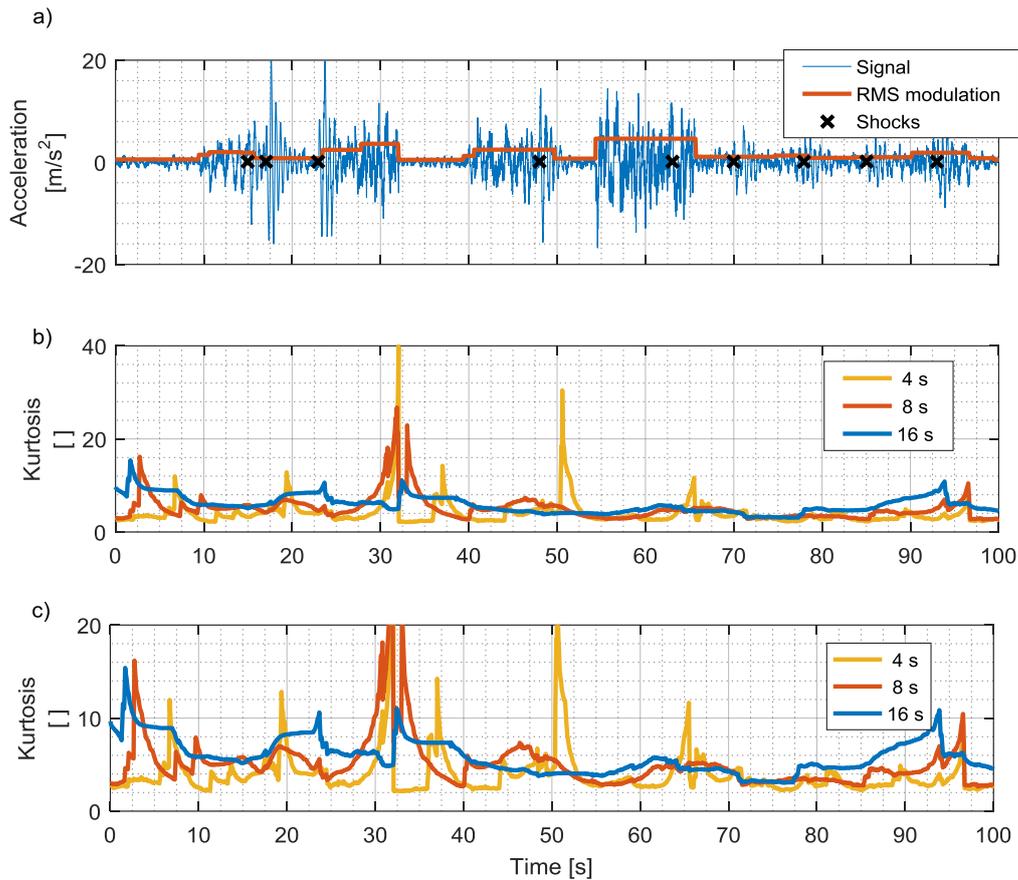


Figure 22: effect of moving kurtosis window length; a) 100 s sample of a nonstationary signal superimposed with shocks; b) full moving kurtosis functions (0-40 y-axis scale); c) moving kurtosis functions zoomed on 0-20 y-axis scale

The comparison of the moving kurtosis functions calculated on the same nonstationary random signal with and without superimposed shocks shows that the shocks have an insignificant effect on the 4 s window functions (Figure 23 a). Interestingly, there are significant discrepancies between the 8 s window functions (Figure 23 b). In this case, the shocks remove the sudden kurtosis change appearing at 47 s and 65 s on the signal without shocks.

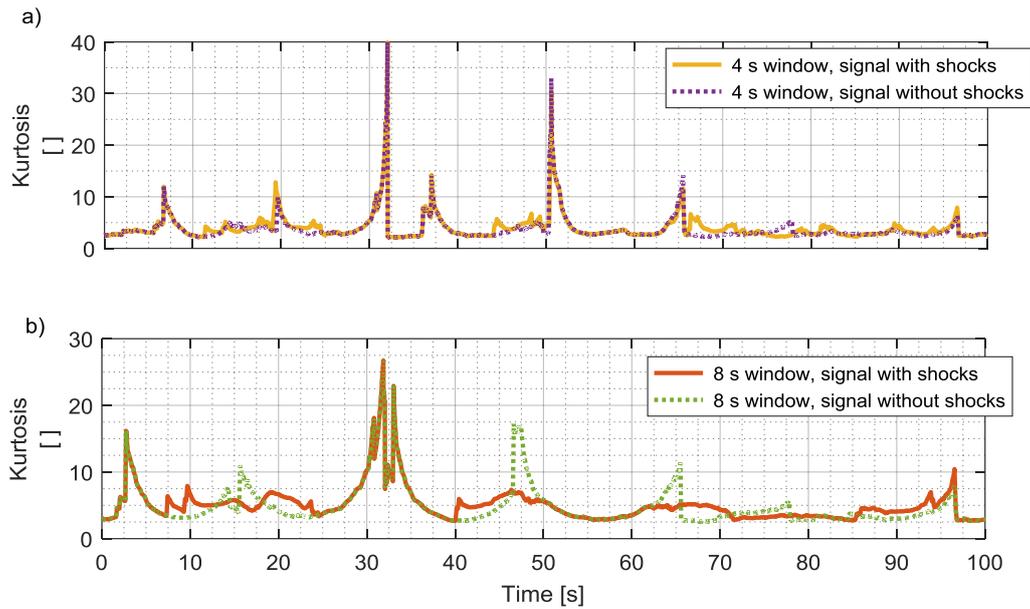


Figure 23: comparison of the moving functions calculated on the same nonstationary random signal with and without superimposed shocks using: a) 4 s window and b) 8 s window

Because the sine functions have inherently a low kurtosis value, their presence in nonstationary signals brings the moving kurtosis function down. This is shown in Figure 24 where from 45 s to 60 s a 5 m/s^2 sine wave is superimposed onto the nonstationary signal shown in Figure 21. In the presence of a sine wave the moving kurtosis barely goes above 3 for both tested window lengths. This completely removes the kurtosis peaks around 50 s which makes the moving kurtosis very sensitive to harmonic components.

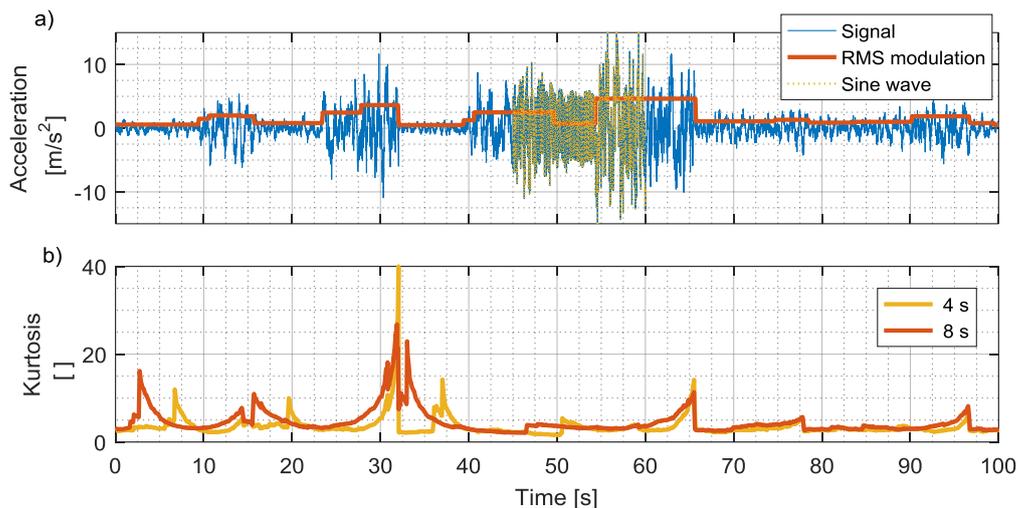


Figure 24: effect of moving kurtosis window length; a) 100 s sample of a nonstationary signal superimposed with a 10 Hz sine wave with an amplitude of 5 m/s^2 ; b) moving kurtosis functions

5.2.3 Moving Crest Factor

As mentioned previously, the crest factor, as a single-value statistic, is designed to be used on a stationary signal. Therefore it does not perform well on RVV signals to identify shocks. However, it is possible to enhance the crest factor performance on a nonstationary signal by using its moving form; *i.e.* using the moving RMS instead of the overall RMS value in the crest factor computation,

$$m_{CF}(t) = \frac{|x(t)|}{\sqrt{\frac{1}{T} \int_0^T x(t+\tau)^2 d\tau}} \quad . \quad (\text{eq. 16})$$

Note how the moving crest factor is calculated “forward” ($t + \tau$) to consider the vehicle causal reaction from the road excitation.

The effect of the window time length, T , was studied by varying the moving crest factor window length from 8 s to 64 s of a nonstationary random signal superimposed with shocks. Depending on the window length, the moving crest factor is more or less sensitive to the signal’s nonstationary segments. As shown in Figure 25, a relatively long window length (64 s) is more sensitive to shocks in general (*e.g.* at 15, 23 and 48 s) but is also sensitive to high RMS segments present in the signal (*e.g.* at 55 s to 65 s segment). On the other hand, the shorter window (8 s) is less sensitive to high RMS segments but indicates falsely high values when the RMS level drops (*e.g.* at 30 s).

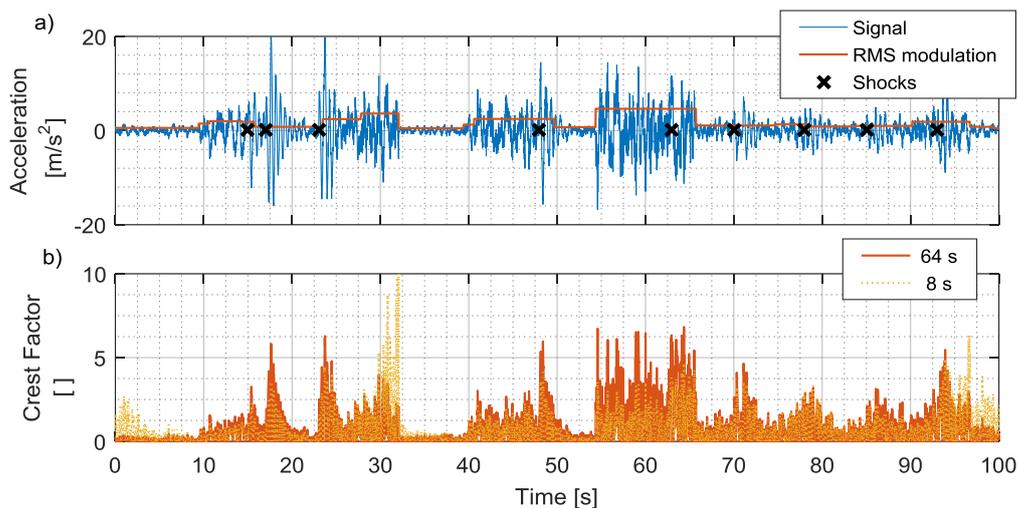


Figure 25: effect of moving crest factor window length; a) 100 s sample of a nonstationary signal superimposed by the shocks; b) moving crest factor functions

5.3 Conclusion on Time Domain Analysis

The mixed modes nature of RVV signals complicates their analysis in the time domain. For instance, the nonstationary components make the crest factor analysis inadequate to detect the transient components. The moving statistics perform better with nonstationarity and are therefore more suited for RVV analysis. Three moving statistics were evaluated, *i.e.* moving RMS, kurtosis and crest factor, and each has different merits. The moving RMS with a window length of 4 s performs best to characterise level of nonstationarity while the 0.5 s is better for shock detection. With window length combination of a 4 s and 8 s, the moving kurtosis is sensitive to nonstationarities and shocks. The same window lengths can also detect the presence of harmonic components. For the moving crest factor, an amalgam of short and long windows (*e.g.* 8 s and 64 s) gives the best compromise between nonstationary and shocks sensitivities.

Unfortunately, none of the time domain analysis techniques presented offers a clear solution. They have some merits but they cannot clearly distinguish between the three different modes present in RVV. Another analysis domain is required to enhance the modes detection. Including the spectral characteristics of each mode with the time-frequency domain brings new solutions to analyse RVV signals.

Chapter 6: TIME-FREQUENCY ANALYSIS

The Road Vehicle Vibration (RVV) is nonstationary and non-Gaussian because of the nature and the sporadic presence of its different vibration modes (nonstationary, transient and harmonic components). These components can be characterised using time-frequency analysis. The spectral variations of the signal over time can be correlated with different components which have distinctive frequency spectra. By looking in the time and frequency domains simultaneously instead of only in the frequency domain, the signal's modes can be more easily analysed even if they are superimposed. For example, a transport vehicle with a faulty transmission system could generate harmonic vibration on a certain gear, so its duration is sporadic and short compared to the other RVV modes occurring during the vehicle normal operation. The harmonic has therefore virtually no effect on an average spectrum calculated on a typical vehicle journey, but has a significant effect in the time-frequency domain at specific times of the journey.

Several methods are used to express a signal in the time-frequency domain. The first methods developed were the short-time Fourier transform and Wigner-Ville distribution, also based on the Fourier transform. However these methods have an important limitation due to the Gabor limit, also known as the signal processing equivalent of the Heisenberg's Uncertainty Principle. More specifically the issue is that an exact frequency cannot be known at an exact time. Fine frequency resolution requires long time segment analysis (*i.e.* coarse time resolution) and fine time resolution results in coarse frequency resolution. The short-time Fourier transform and the Wigner-Ville distribution are inherently limited by the Gabor limit because of the continuous nature of the cosine functions composing the Fourier transform (Newland, 2012).

This limitation can be minimised using more advanced time-frequency analysis methods such as the Hilbert-Huang Transform (HHT) and the Discrete Wavelet Transform (DWT). The fundamentals of these methods are presented in Appendix A for the HHT and Appendix B for the DWT. These appendices provide enough information to understand and even apply these advanced time-frequency analysis methods.

In this chapter, these methods are assessed on their capabilities to provide enough information to detect and simulate the nonstationary, transient and harmonic components contained in real RVV signals. In order to maintain the same statistical distribution, the starting point and duration of the different components have to be identified in real RVV signals. Key elements of each component also have to be characterised in order to reproduce them properly. These elements are: (1) for the nonstationary components, the spectrum and RMS value; (2) for the transient components, the shock duration and amplitude and, ideally, the transient waveform; and (3) for the harmonic components, their frequency, amplitude and duration. Synthetic RVV signals are used to perform this assessment as these elements are *a priori* known. It must be noted that the results presented in this chapter are typical representation of numerous analyses performed on the topic.

6.1 Hilbert-Huang Transform Assessment

The HHT is an adaptive time-frequency analysis method providing different types of predictors from RVV signals. The HHT uses a sifting process to decompose a signal into different narrow-banded components, called Intrinsic Mode Functions (IMFs) which have the following characteristics as defined by Huang et al. (1998):

1. in the whole dataset, the number of extrema and number of zero-crossings must either equal or differ by at most one;
2. at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The sifting process is made by fitting a spline function over signal local maxima and a second over local minima. The difference of these functions is then subtracted from the signal which has the effect of removing the low frequency components of the signal. This process is repeated until the number of extrema and zero-crossings remain the same and are equal or differ, at most, by one for 8 consecutive iterations (Huang et al., 2003a). The remaining signal becomes the first IMF. The next IMFs are then calculated using what remains from the signal using the same process. This iterative process stops when no more IMF can be fitted and the residual is called the signal trend. As the IMFs are a narrow-banded, their instantaneous frequency and amplitude can be computed using Hilbert transform. The instantaneous frequencies and amplitudes can be expressed as individual functions or regrouped together to form the Hilbert spectrum which presents the amplitude of the instantaneous frequency as a function of the time.

A more detail description of the HHT is presented at Appendix A. It is important to notice that the HHT's mathematical framework still has to be developed. Therefore, it is good practice to assess the capabilities of the HHT on synthetic RVV signals. This assessment is also useful to

compare the HHT with the DWT to ultimately select the time-frequency analysis tool that suits best the RVV.

Three synthetic calibration signals were created to assess the HHT performance as described in Chapter 4. The first signal used for the assessment is a Gaussian random signal with two different RMS values within the same vehicle vibration spectrum. The duration of the signal is 15 s. From 5 s to 10 s, the average RMS value is 12 m/s² and then 6 m/s² for the remainder of the signal (Figure 26 a). The sampling frequency (1024 Hz) is sufficient to provide a smooth representation of the signal waveform. The second signal is a stationary Gaussian random signal based on the same vehicle vibration spectrum (6 m/s² RMS value) superimposed with five shocks (Table 2 and Figure 26 b). The shocks consist of the impulse response function of a typical truck (quarter car, such as presented by Cebon, 1999) to a Hanning function excitation with varying amplitude and duration. The last signal once again contains the same stationary Gaussian random signal, with a 10 Hz sinusoidal waveform of 10 m/s² amplitude superimposed between 5 s and 10 s (Figure 26 c). These signals were designed especially to test and evaluate the effectiveness of the HHT to identify the three RVV modes.

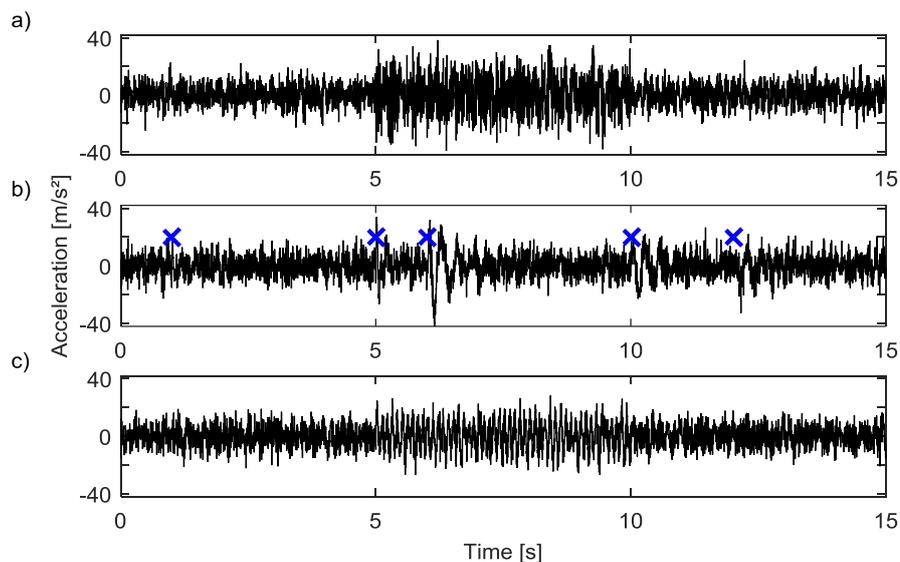


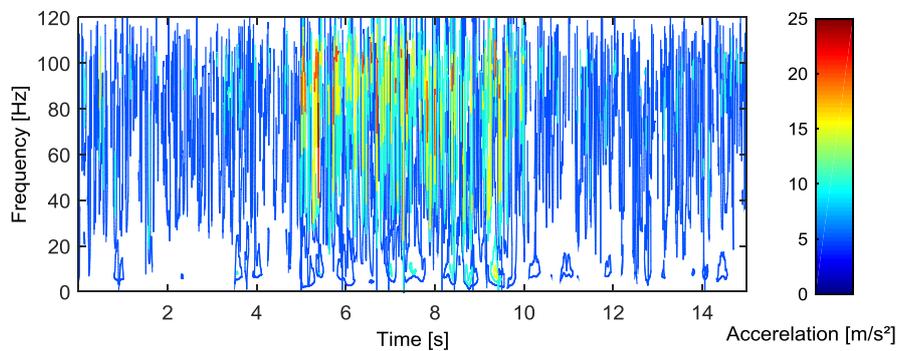
Figure 26: a) signal comprising three Gaussian segments with two different RMS values; b) stationary Gaussian random signal with shocks (asterisk marks); and c) stationary Gaussian random signal with a 10 Hz sinusoidal component between 10 s and 15 s

Table 2: properties of the road surface aberrations used to produce the synthetic shock

Impact [#]	Position [s]	Duration [s]	Amplitude [m]
1	1	0.015	0.05
2	5	0.03	0.05
3	6	0.15	0.05
4	10	0.15	0.03
5	12	0.15	0.02

6.1.1 Nonstationary Component

From Figure 27, it appears to be possible to visually identify the different Gaussian components using the Hilbert spectrum where the global spectrum amplitude increases between 5 s and 10 s. However, this Hilbert spectrum presents some issues concerning the Gaussian characterisation. Its general shape contains discrepancies with the Power Density Spectrum (PDS) function of the Gaussian signal itself (Figure 28). The dominance of the frequencies around 5 Hz and 10 Hz is clearly not represented in the Hilbert spectrum.

**Figure 27:** Hilbert spectra on Gaussian signal with two different RMS values

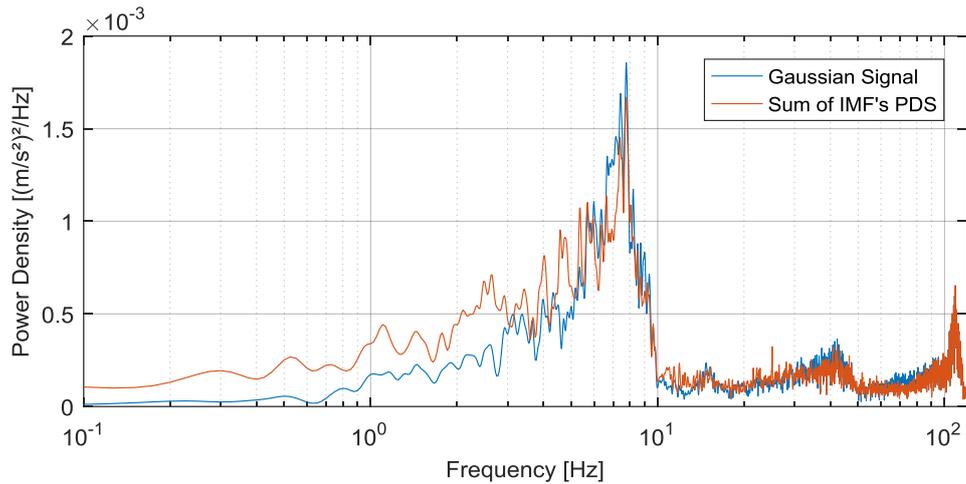


Figure 28: PDS of the Gaussian signal and the sum of the PDS of the signal's IMF

The discrepancies between the spectral shapes can also be explained by the IMFs' instantaneous frequencies principle. The high order IMFs have narrower and lower frequency content which varies slowly through time, while the low order IMFs have broader and higher frequency content which varies quickly. In order to represent the same power density, the low order IMFs need greater amplitude which creates a bias in the spectral shape obtained in the Hilbert spectrum. To illustrate this, the PDS function of each IMF is presented in Figure 29 where IMFs 4 and 6 dominate the between 6 Hz to 10 Hz.

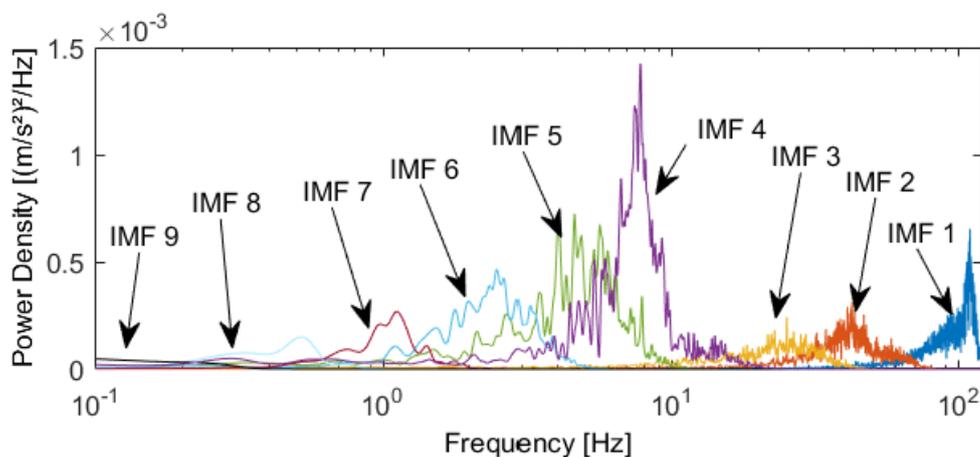


Figure 29: PDS function of each IMF

A second issue appears when these PDS are summed; the frequency content below 6 Hz is over-represented (Figure 28). This is mainly caused by the IMFs decomposition and their instantaneous frequencies. IMFs 6 to 9 show instantaneous frequencies that are artefacts of the limit criteria of the decomposition method (Figure 29). The algorithm keeps decomposing the signal until the IMF or residual amplitude reach a certain threshold, but since this process acts in a similar way to a low pass filter, the remaining signal is modelled only with very low frequencies, which creates this bias. Advice on how to deal with this shortcoming is given by

Huang and Shen (2005). The nonphysical IMF should be left out of the analysis, but there are no guidelines on how to define a nonphysical IMF. Therefore, it is impossible to know that IMFs 6 to 9 do have physical sense if the original spectrum of the signal is unknown, as it is the case for nonstationary signal. Another technique proposed by Peng et al. (2005) using Wavelet Packet Transform (WPT) and correlation test on IMF with the signal can also remove the artefact. However, because this technique uses the DWT the HHT loses its fully adaptive property.

6.1.2 Transient Component

The Hilbert spectrum of the signal with shocks (Figure 30) shows high amplitude values at low frequency at 6 s and 10 s which are long duration shocks, but there is no clear evidence of the other events. The IMFs themselves could also be used to detect shocks. For instance, IMF 3's amplitude increases at 1 s and 5 s when short duration shocks occur and IMF 4's amplitude increases at 6, 10 and 12 s when the long duration shocks occur (Figure 31). This could give information on the shocks' duration, but once again there are artefacts that could lead to false detections based on the IMF amplitude at 4 s and 11 s for instance. One could also use IMF 5 to falsely detect shocks. This indicates that the HHT does not enhance shock detection compared to time domain analysis where these events can be visually appraised (Figure 26 b). There is also no indication on the waveform of the transient with the HHT.

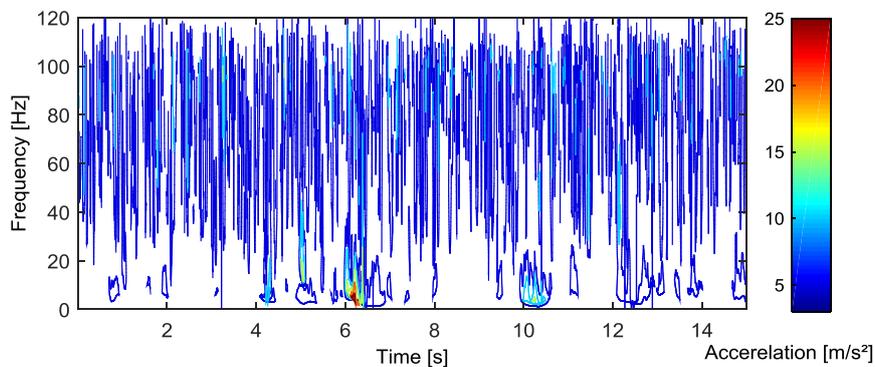


Figure 30: Hilbert spectrum on Gaussian signal superimposed with shocks

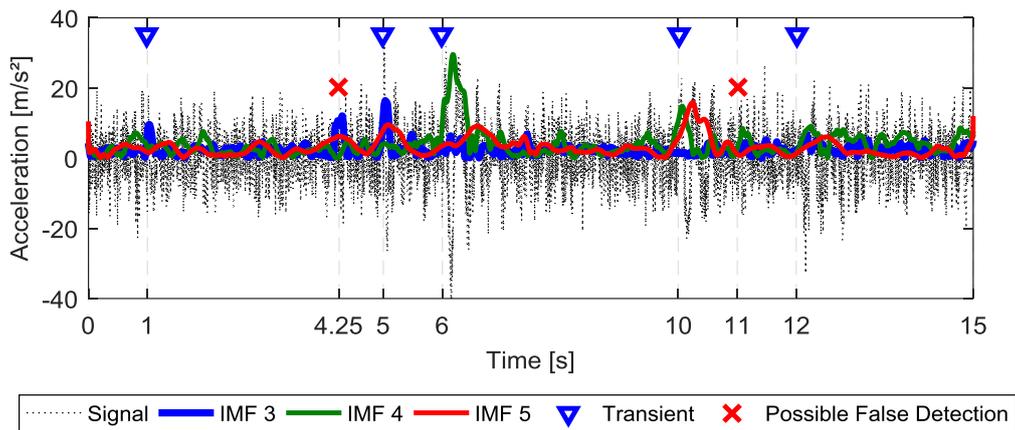


Figure 31: IMFs 3 to 5 of the Gaussian signal superimposed with shocks

6.1.3 Harmonic Component

The Hilbert spectrum shown in Figure 32 reveals concentrated low frequency activity between 5 s and 10 s. However, this frequency bandwidth is relatively broad and does not represent the single frequency sine component. The continuous 10 Hz sine wave is instead represented by scattered broad-band components.

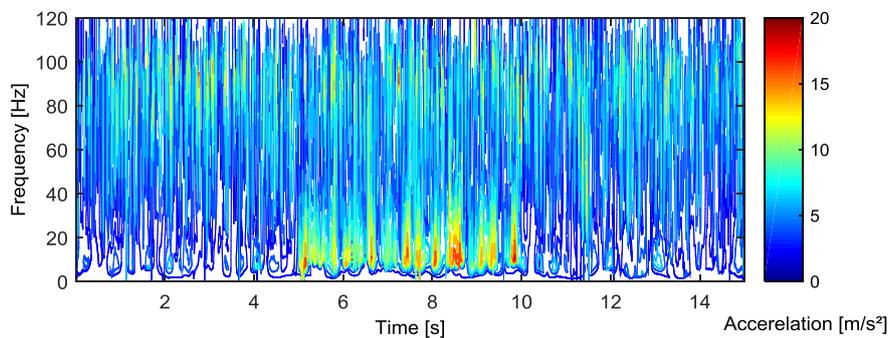


Figure 32: Hilbert spectrum of Gaussian signal with a 10 Hz sinusoidal component

This discontinuity in frequency and amplitude is explained by the shift (or leak) of the sine wave from one IMF to another adjacent IMF. The EMD performed on the signal with lower relative amplitude to the Gaussian component (RMS value is 10 % of that of the sine amplitude) shows that even with a relatively low noise level, the sine wave appears in two IMFs with an amplitude fluctuation (Figure 33). This mode-mixing gives a false representation, in the Hilbert spectrum, of the stationarity of the sine wave, resulting in misleading information on its amplitude, frequency and duration.

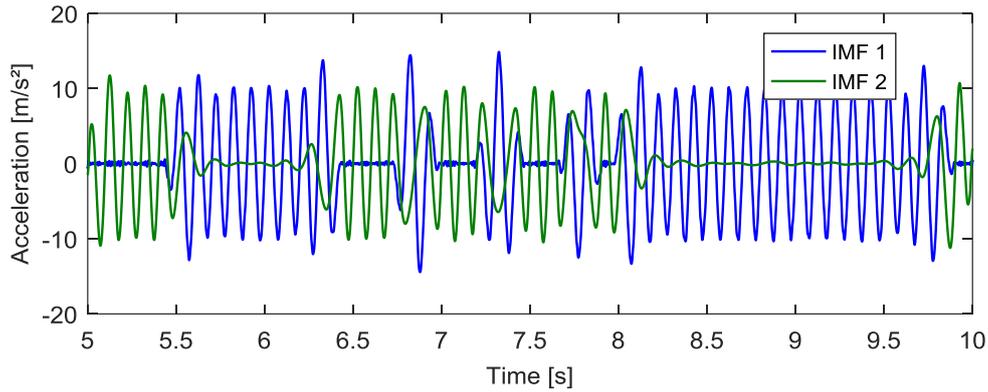


Figure 33: FIRST TWO IMF of the 10 Hz sinusoidal component with 10 % Gaussian noise

Huang et al. (2003a) propose a bandwidth limitation as an EMD stoppage criterion to ensure that a single frequency component remains in the same IMF. However, to use such a criterion, the instantaneous frequency of the component has to be known, which is not necessarily the case for RVV. This subjective intervention also limits the fully adaptive nature of the EMD.

In order to maintain the adaptive nature of the EMD, Wu and Huang (2009) propose the Ensemble EMD (EEMD) to avoid mode-mixing between IMFs. This method uses a characteristic of the EMD, that of being a dyadic filter bank on Fractional Gaussian noise such as a white noise (Flandrin et al., 2004). By adding white noise on a signal, the EEMD ensures that the IMFs are the result of the filter bank, such that a mode will more likely remain in the same IMF. It is then possible to remove the effect of the white noise by repeating the EMD on the signal with different white noise superposition. Since these noises are uncorrelated, the effect can be minimised by computing the average of the replicated IMFs. The effect of the additive noise, ε_n , decreases with the number of EMD repetitions, N , as follow:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{N}}, \quad (\text{eq. 17})$$

where ε is the amplitude of the added noise.

EEMD is not broadly used because the mode-mixing problem is not always important enough to justify the extra computing time required. The method also has some drawbacks such as not totally correcting the mode-mixing issue when the mode frequency is located in the overlapping region of the filter bank. Also, nothing guarantees that the mean IMFs respect the definition in regards to the number of extrema and zero-crossings. Before using the EEMD, these considerations must be assessed in terms of the significance of the mode-mixing problem.

6.1.4 Energy Density/Average Period Significance Test

EEMD uses noise to improve the HHT analysis but in most cases the noise present in a signal is undesirable. As presented in Appendix A.4.1, the HHT can be used to distinguish the “noise” from a “signal”. For RVV, the nonstationary component, which is a Gaussian random signal, can be qualified as “noise” and the transient and harmonic components, which are deterministic, can be qualified as the “signal”. So in theory, the noise filtering capacity of HHT can be used to separate the different RVV components.

The first HHT filtering method assessed on RVV is the energy density/average period significant test. As explained more in detail in Appendix A.4.1, EMD of white noise (Gaussian) has a known logarithmic energy density/average period relationship for every IMF. So EMD showing any discrepancy with this relation means it is a signal in the specific IMF which does not follow the relationship. The same principle could be applied to other random noises where the energy density/average period relationship between IMFs will follow a specific pattern.

To define the energy density/average period relationship of the nonstationary component of the RVV, the distribution of the IMF energy density E in function of the average period T was calculated using a Monte Carlo simulation of a Gaussian signal made from the same spectrum and 1000 different randomly distributed phase spectral functions. The RMS value of all these signals is approximately 8 m/s^2 . The result shows that the spread function has an irregular pattern which does not follow the logarithmic decrement of the white noise case (Figure 34).

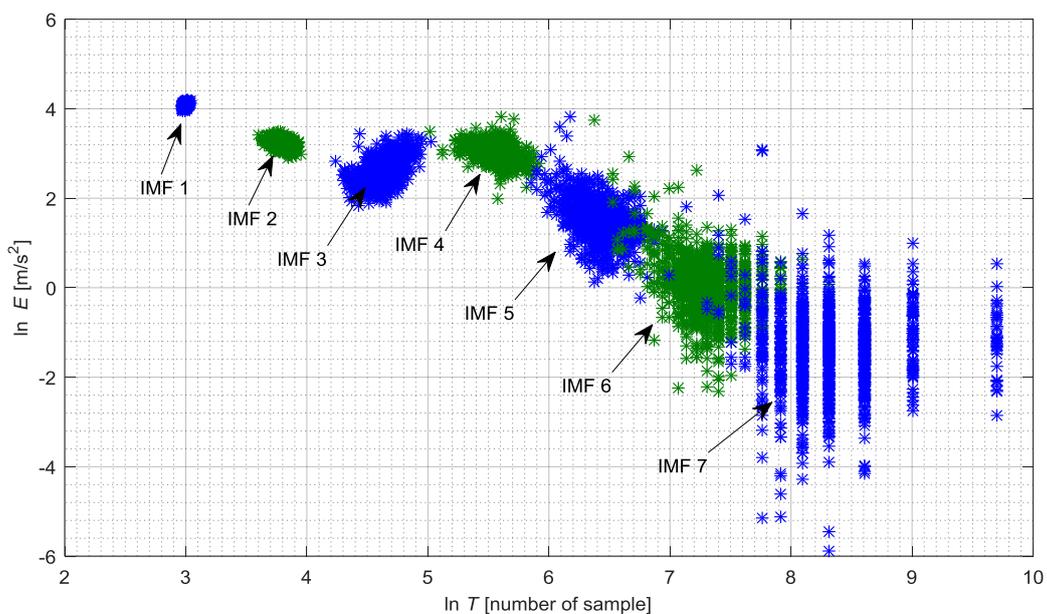


Figure 34: spread function of the energy/average period relationship of IMFs of 1000 vehicle Gaussian signals; the odd IMFs are in blue and the even in green

Before using this method to identify the IMFs containing the transient and harmonic components of a RVV signal, the effect of the nonstationary Gaussian signal on the distribution is assessed. The nonstationary signal is created using the same PDS based Gaussian signal as the one used during the Monte Carlo simulation but its RMS value varies (Figure 26 a), from 5 s to 10 s, the average RMS value is 12 m/s² and then it is 6 m/s² for the remainder of the signal. The overall RMS value of the signal is 8 m/s², the same as the stationary Gaussian signal used in the Monte Carlo simulation. As seen in Figure 35, the nonstationarities affect the number of IMFs generated by the EMD. For the same sampling frequency, there are seven IMFs for the stationary Gaussian signal and twelve IMFs for the nonstationary. The energy density values are situated in the cluster generated by the stationary Monte Carlo simulation which means that the energy density/average periods relationship is not affected by nonstationarities.

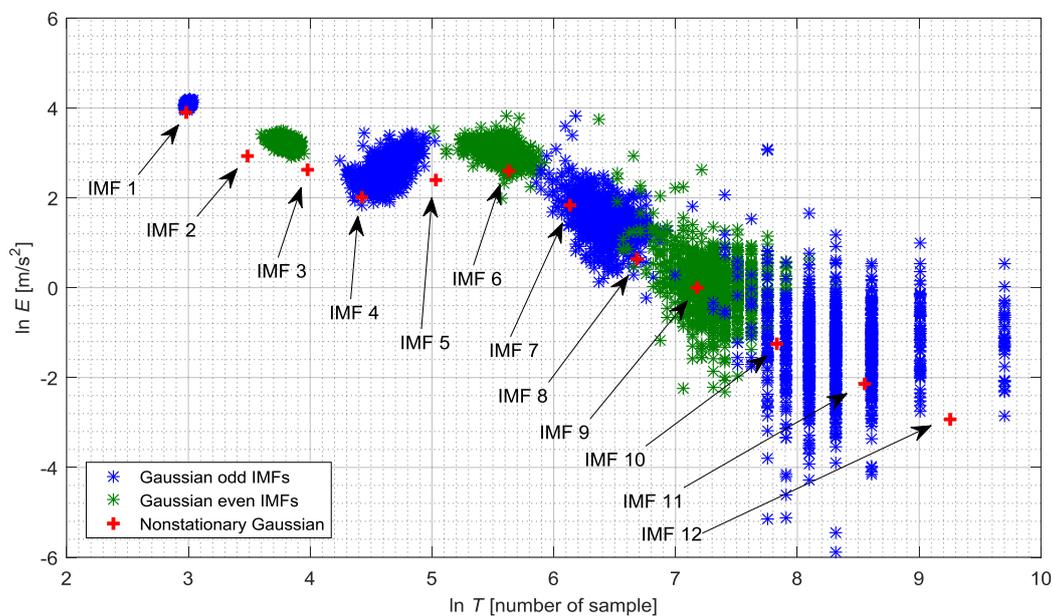


Figure 35: energy/average period functions of the nonstationary Gaussian signal and Gaussian Monte Carlo simulation; the IMFs are in ascending order starting from the left, the odd Monte Carlo IMFs are in blue and the even in green

The energy density/average period significance test shown in Figure 36 is not able to detect a harmonic component superimposed on a stationary Gaussian signal (Figure 26 b). As for the nonstationary signal, the number of IMFs is superior to those of the stationary Monte Carlo simulation. The relationship between the energy density and average period of the harmonic signal is within the trend of the Monte Carlo distributions. The IMFs are also positioned near the nonstationary Gaussian IMFs. The IMFs containing the harmonic component do not clearly stand out of the Monte Carlo distribution or the nonstationary Gaussian IMFs which can be caused by the mode-mixing problem which spreads the energy of the harmonic component into

several IMFs making the energy density/average period significance test unreliable to detect harmonic components.

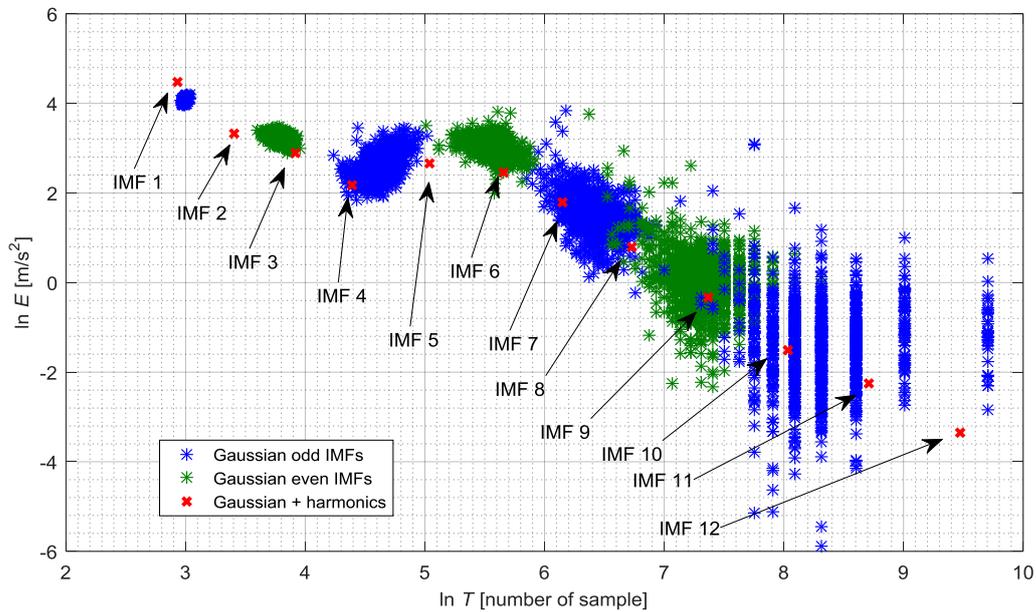


Figure 36: energy/average period functions of the Gaussian signal superimposed with harmonics and Gaussian Monte Carlo simulation; the IMFs are in ascending order starting from the left, the odd Monte Carlo IMFs are in blue and the even in green

As seen in Figure 37, the energy density/average period significance test can detect the IMF containing shocks superimposed on stationary Gaussian signal (Figure 26 c). As seen in Figure 35, the fifth IMF has more energy density than the Monte Carlo spread distribution. The average period of the fifth IMF (about 150 samples or 0.15 s) coincides with the average frequency of the impulse response used to synthesise the impact in the transient signal (6 Hz). However, the problem is that other IMFs are present outside the confidence intervals such as the second, third sixth and seventh IMFs, which have no physical relation with the shocks superimposed on the Gaussian signal which affects the reliability of the technique.

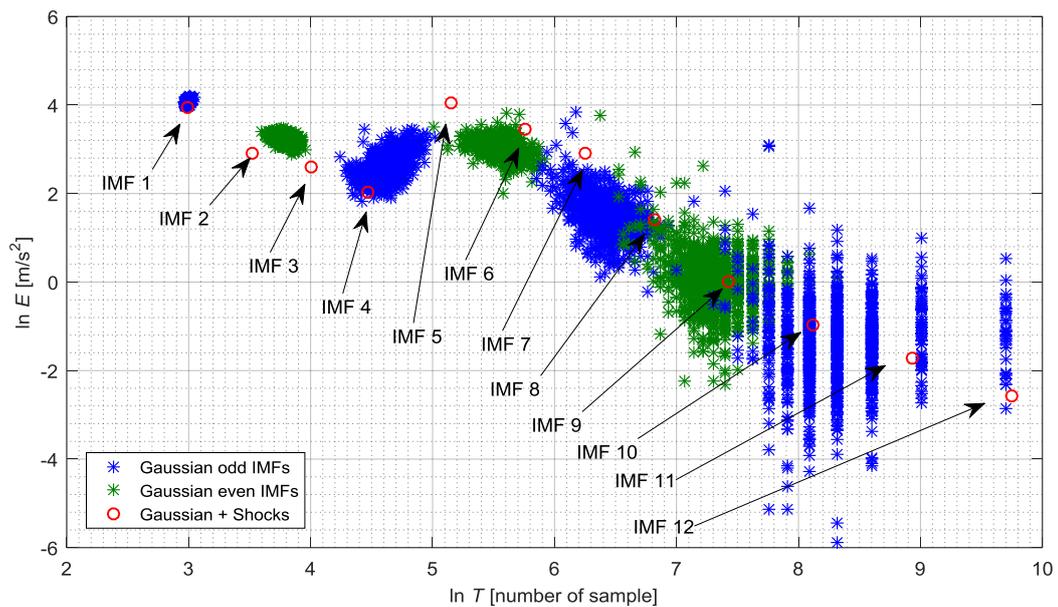


Figure 37: energy/average period functions of the Gaussian signal superimposed with shocks and Gaussian Monte Carlo simulation; the IMFs are in ascending order starting from the left, the odd Monte Carlo IMFs are in blue and the even in green

6.1.5 Fractional Resampling Technique

The energy density/average period significance test does not totally suit the RVV analysis which could have been anticipated because it was specifically created for white noise (Wu and Huang, 2004). Facing the same issue, Chen et al. (2013) developed the Fractional Resampling Technique (FRT), also based on HHT, to detect a signal in fractional Gaussian noise such as red (Brownian) noise, see Appendix A.4.2 for more detail. Since the road profile which creates the random noise in RVV could be modelled as a red noise (Cebon, 1999), FRT could also be applied in RVV signal analysis. Based again on the principle that the nonstationary component is the “noise” and the transient and harmonic components are the “signals”, the FRT should determine which IMFs contain the shocks and harmonics.

The first parameter to set up is the initial sampling frequency of the FRT. RVV signals are usually over sampled to give a more detailed curve. In other words, the Nyquist frequency is well above the highest frequency of the signal which gives several points around signal’s peaks and troughs. However oversampled signals spoil the FRT because the average frequencies of the first IMFs do not correspond to the filter bank. Following the principle of a dyadic filter bank, the first filter or IMF should have an average frequency of 3/4 of the Nyquist frequency but an oversampled signal does not have any content at these frequencies so its first IMF has an average frequency below this value. The signal resampling does not have any effect if the average frequency of the first IMF remains below 3/4 of the Nyquist frequency of the down-

sampled signals. This leads to a stable normalised average frequency as seen on a stationary Gaussian signal created from a vehicle vibration spectrum sampled at 1024 Hz, Figure 38.

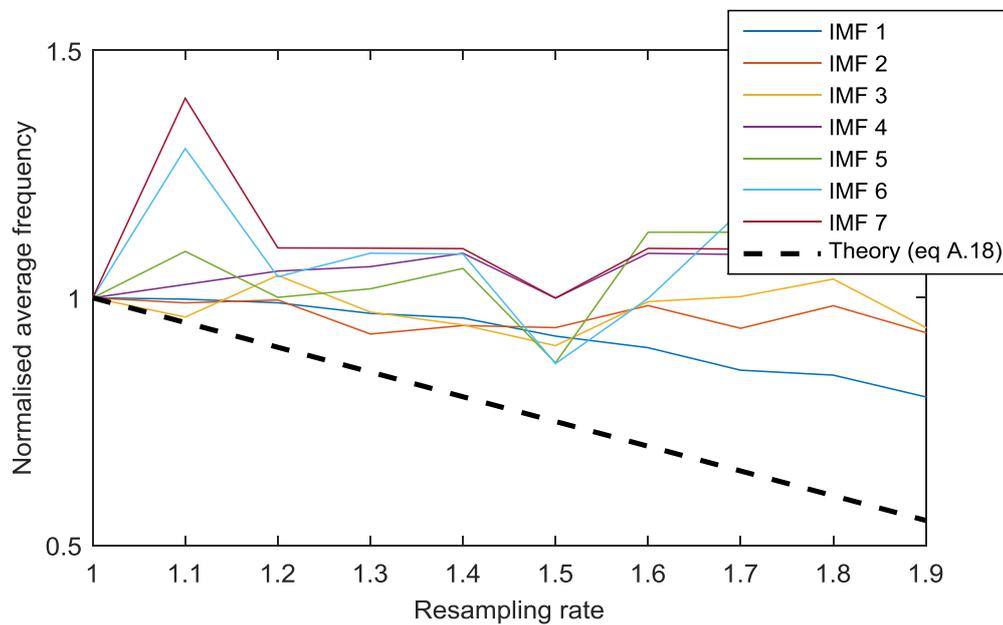


Figure 38: normalised average frequencies of IMFs as a function of the resampling rate for a Gaussian signal created from a vehicle vibration spectrum sampled at 1024 Hz

The decrement of the normalised average frequency is seen on the Gaussian signal at 256 Hz sampling frequency. However, shocks superimposed on a stationary Gaussian signal as presented in Figure 26 b are not detected with the FRT. The normalised average frequencies of the IMFs mostly follow the theoretical linear relationship of the red noise, Figure 39.

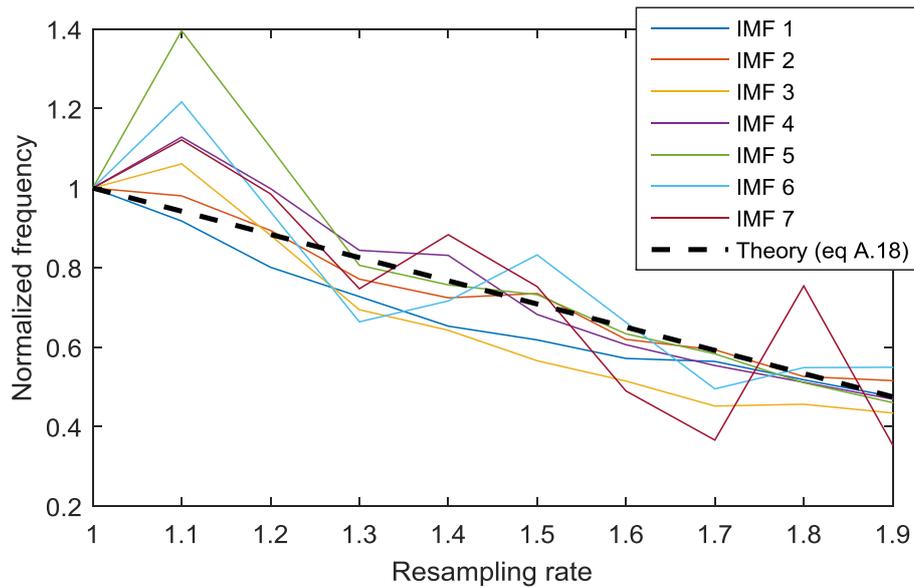


Figure 39: normalised average frequencies of IMFs as a function of the resampling rate for shocks superimposed on Gaussian signal sampled at 256 Hz, with the dashed line representing the theoretical relationship

The harmonic component seems to be detected with the FRT, but this is not totally correct (Figure 40). Applied to a signal composed of a harmonic component superimposed on a Gaussian signal, the seventh IMF indicates the normalised average frequency and resampling rate theoretical function, but it does not represent the harmonic component. The sinusoidal frequency of the harmonic component is at 10 Hz but the seventh IMF has an initial average frequency of 1 Hz. So the outlining IMF is only an artefact.

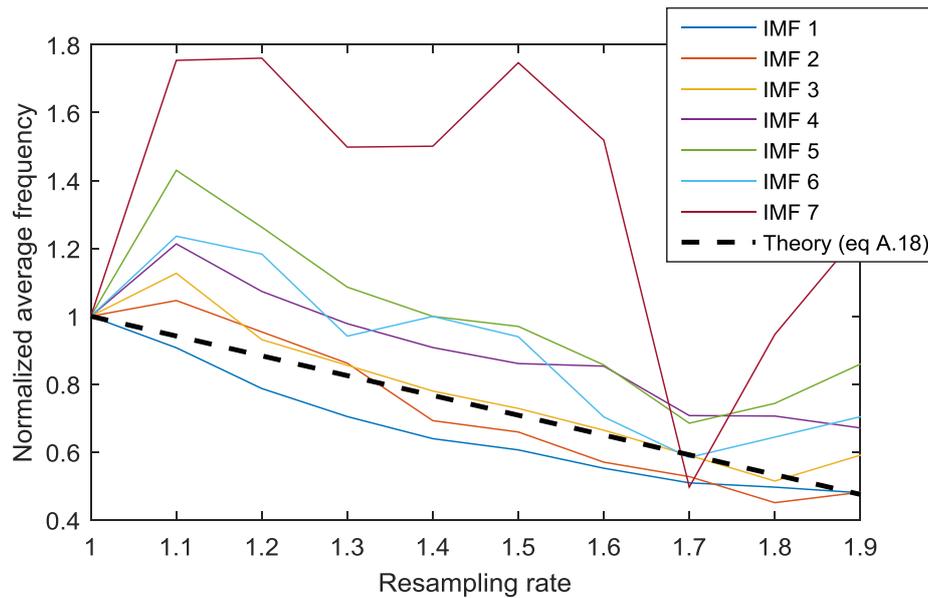


Figure 40: normalised average frequencies of IMFs as a function of the resampling rate for harmonic components superimposed on Gaussian signal sampled at 256 Hz, with the dashed line representing the theoretical relationship

6.1.6 Conclusion on the Hilbert-Huang Transform

The HHT is an adaptive time-frequency analysis tool with interesting potential but has no practical application to RVV in its current state of development. Instabilities due to stopping criteria selection and modes-mixing within IMFs are two of the issues that complicate the application of the HHT on RVV. This is the drawback of an adaptive signal processing technique; it adapts too much to a signal, so the outcome might end-up as complex as the signal itself.

The signal detection and noise filtering capabilities of HHT have interesting features but they were found not to be sensitive or reliable enough to be used with RVV signal. It appears that the techniques presented are not versatile enough to analyse the multimodal nature of the RVV.

Considering these limitations, other time-frequency analysis tools such as the DWT, which has a complete mathematical framework, might be more appropriate and give a more reliable method to analyse RVV signal.

6.2 Discrete Wavelet Transform Assessment

As opposed to the HHT, the wavelet transform has a fully proven mathematical framework. Therefore the wavelet transform of a RVV signal has an analytic solution and does not need to be extensively validated for this application. The wavelet transform consists of series of filters applied in sequence on a signal. The outcome of each filter represents a different scale of the signal which is related to its frequency content.

The DWT is used over the Continuous Wavelet Transform (CWT) to avoid redundancy between the scales and to make sure the different components of the signal are present only in one scale (see Appendix B for more detail on wavelet transform). The drawback of the DWT is that its scalogram is a coarser scale resolution than the CWT. Note here that the DWT results are presented in the time-scale domain rather than the time-frequency domain because the analysed signal is divided in terms of scales and not frequencies. The scale domain may be transformed into the frequency domain depending of the mother wavelets, *i.e.* the shape of the wavelet.

There are several types of mother wavelets and it is important to define which one is suited to RVV application. Different mother wavelets have been used in the literature to analyse road and vehicle vibration signals. Ayenu-Prah and Attoh-Okine (2009), Lee and White (2000), Lee and Son (2001), Nei et al. (2008) used the Morlet wavelet to analyse RVV signals but they do not justify their choice. Staszewski and Giacomini (1997) justified the usage of the Morlet wavelet because its scale can be more easily expressed in the frequency domain since the Morlet wavelet is a complex exponential function multiplied by a Gaussian window. Griffiths (2012) employed a more extensive selection process. She compared Haar, Mexican hat, Daubechies 2, 6, and 10ⁱ and Morlet wavelets. She suggested that Daubechies 6 and 10 and Morlet wavelets should be used to analyse RVV. Wei et al. (2005) also tried several Daubechies wavelets (1 to 10) and found that at least three vanishing points are needed to analyse the RVV signal. Nei et al. (2008) also used Daubechies 8 wavelet but there is no mention in the paper why this order is used and why Morlet wavelets are used for some applications and Daubechies for others.

Based on a review of the literature, both Morlet and Daubechies wavelets seem to be appropriate for with RVV signals. However for indexing the different modes in RVV, the Daubechies family is preferred because it has an orthogonal basis which ensures there is no redundancy between the scales. This means that there is no correlation between the scales which leads to a number of statistical simplifications.

It is now a question of selecting the order of the wavelet (or its number of vanishing points) which corresponds to the k^{th} moments where a function, $f(x)$, equals zero,

$$m_k = \int_{-\infty}^{\infty} f(x)x^k dx . \quad (\text{eq. 18})$$

The number of vanishing points p of the wavelet function gives an indication of the precision of every scale. In other words, any degree- p polynomial can be exactly reproduced in every scale. The more vanishing points there are, the more concentrated is the signal information in the coarser scales. Therefore fewer coefficients are needed to analyse singularities and

ⁱ These numbers represent the number of vanishing points of the wavelet.

discontinuities. More vanishing points means that the wavelet filters need longer support (more filter coefficients); for instance, a Daubechies wavelet with p vanishing moments has $2p - 1$ filter coefficients.

Because RVV signal analysis does not require real time processing, the filter length is not an issue. A high number of vanishing points can be used to further separate the information between the scales. In the context of the thesis, 10 vanishing points are used with Daubechies wavelets (Daubechies 10). The number was selected using a real typical RVV signal. This signal recorded on a truck travelling at 100 km/h on Victorian roads (Australia), was analysed with Daubechies wavelets 1 to 15. The RMS value of the first six wavelets was compared for every number of vanishing points. As seen in Figure 41, the RMS values reach an asymptote around seven or eight vanishing points. That means that the wavelet decomposition of a real RVV signal does not change when there are more than eight vanishing points. As the filter length of the wavelet analysis is not an issue in this case, the number of vanishing points was fixed at 10 to have smoother wavelets with a reasonable length, *i.e.* 19 coefficients.

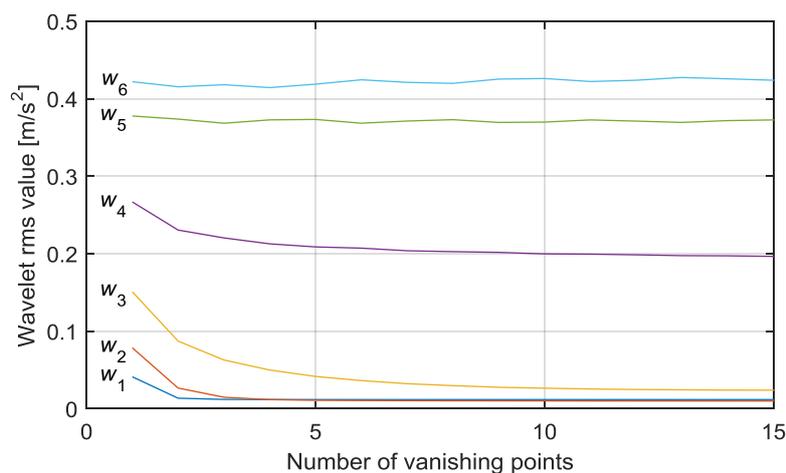


Figure 41: RMS values of the first six Daubechies wavelets of a real RVV signal in function of the number of vanishing points

The Daubechies 10 wavelet has all characteristics required for analysing a RVV signal and its performance in detecting nonstationary, transient and harmonic components needs to be assessed. The same synthetic signals used for the HHT assessment are used to facilitate the comparison between the two time-frequency analysis methods. These components are individually tested using scalograms which present the coefficients in detail for every scale as a function of time. The finer scale represents the high frequencies with the finer time resolution and the coarser scale represents the lower frequencies with the coarser time resolution.

6.2.1 Nonstationary Component

As for the HHT, the performance of the DWT for identifying nonstationary components is assessed using the same Gaussian random signal. This signal was synthesised from a vehicle vibration spectrum and has two different RMS values sections, 6 m/s^2 from 0 to 5 s and from 10 s to 15 s and 12 m/s^2 from 5 s to 10 s (Figure 42 a). The DWT does not clearly identify the high RMS section of the signal (Figure 42 b). The coefficients are higher between 5 s and 10 s at scales 2 to 8, but the difference between the RMS sections is not as clear as it was for the HHT (Figure 27). This can be explained by the DWT being more sensitive to signal variation than to the stationary segment (Hubbard, 1998). In other words, the DWT is better suited to detect changing points than to quantify steady-state intensities.

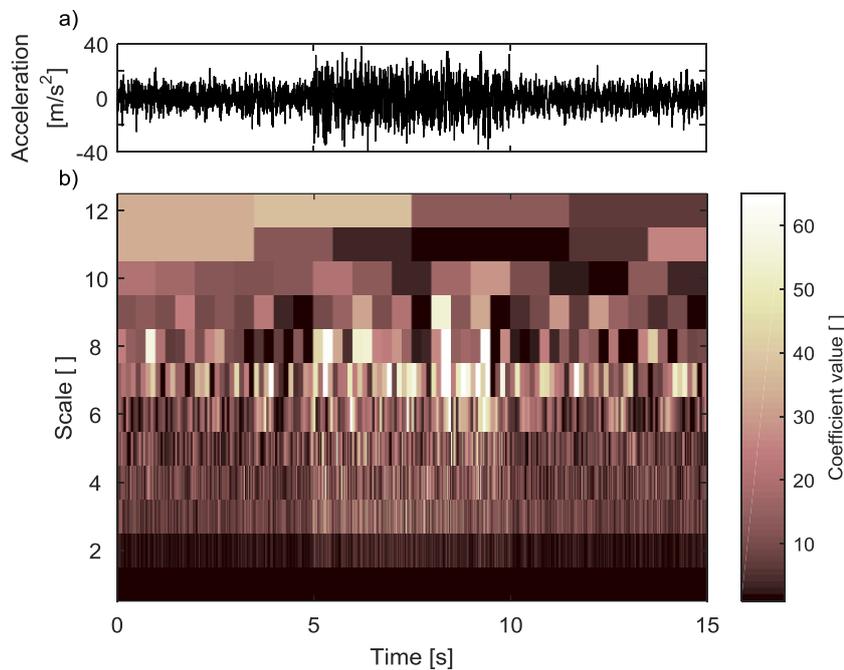


Figure 42: nonstationary signal (a) and its scalogram (b)

6.2.2 Transient Component

Shocks are short time high amplitude events in a signal, which can be seen as important signal variation. Therefore, the DWT is expected to have an inherent capacity to detect such events. This is the case for the stationary Gaussian signal (based on the same vehicle vibration spectrum with a constant 6 m/s^2 RMS level) superimposed with five shocks (Table 2 and Figure 43 a). As seen in the scalogram (Figure 43 b), scales seven and eight contain high amplitude coefficients when the shocks occur, excluding the first event which has a very short duration, hence not much energy (Table 2). This means that the shocks can be mainly expressed with coefficients of scale seven and eight (Figure 44). The envelope p , the magnitude of the analytic signal, of this reconstruction signal s_{7-8} also gives a good indication of the shocks' position;

$$p = \left| s_{7-8}(t) + iH \left[s_{7-8}(t) \right] \right|, \quad (\text{eq. 19})$$

where $H \left[s_{7-8}(t) \right]$ is the Hilbert transform of the reconstruction signal (see Appendix A.1 for more detail on Hilbert transform).

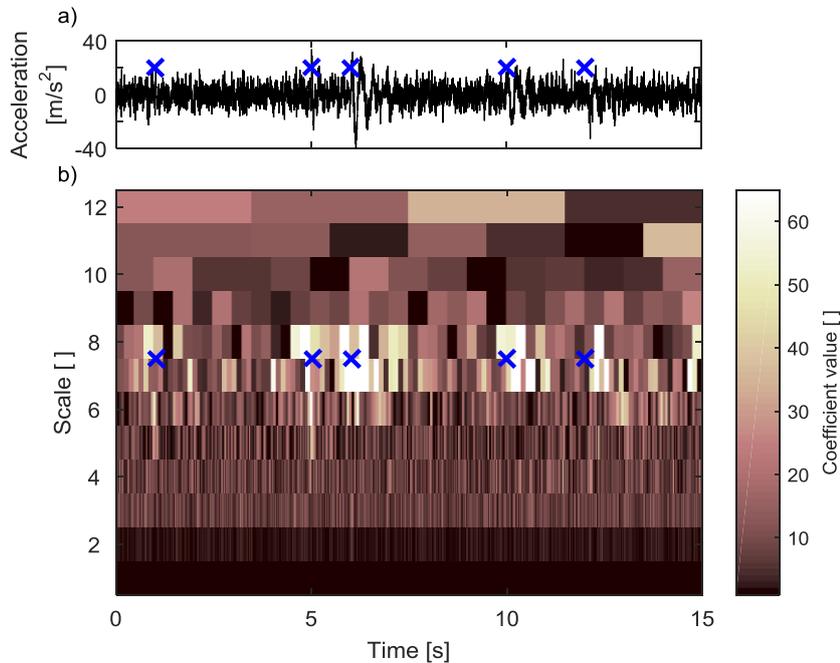


Figure 43: a) signal superimposed with shocks (marked with blue X); b) its scalogram

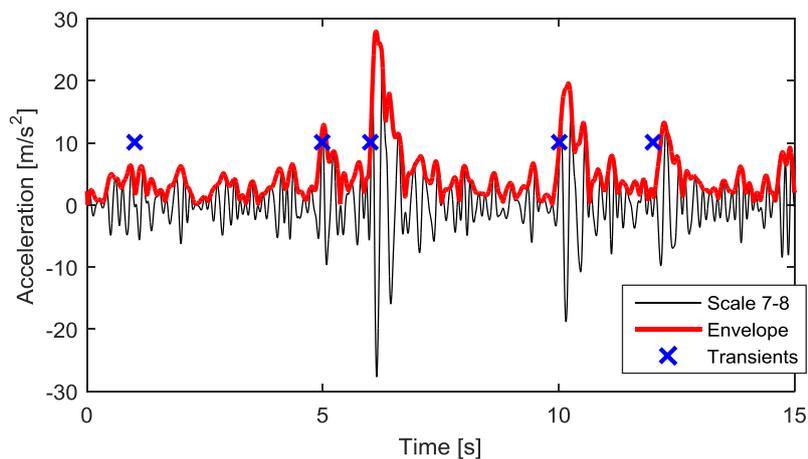


Figure 44: transients signal reconstruction using scales 7 and 8 of DWT coefficients

These results show that DWT is relatively well suited to detected shocks buried in a stationary random signal. As the shocks have very short duration, detecting them is equivalent to finding changing points in the random signal which, as stated in the previous subsection, is a known application of the DWT.

6.2.3 Harmonic Component

The DWT performs very well in finding the 10 Hz sinusoidal waveform component (amplitude 10 m/s^2) superimposed from 5 s to 10 s on a stationary Gaussian signal (same signal used for the HHT assessment, Figure 45 a). The scalogram shows a high amplitude coefficient at scale 6 when harmonic component is present. The reconstruction of the signal using only this scale also clearly shows the waveform (Figure 46) suggesting that the DWT is suited to detect harmonic component buried in random signal.

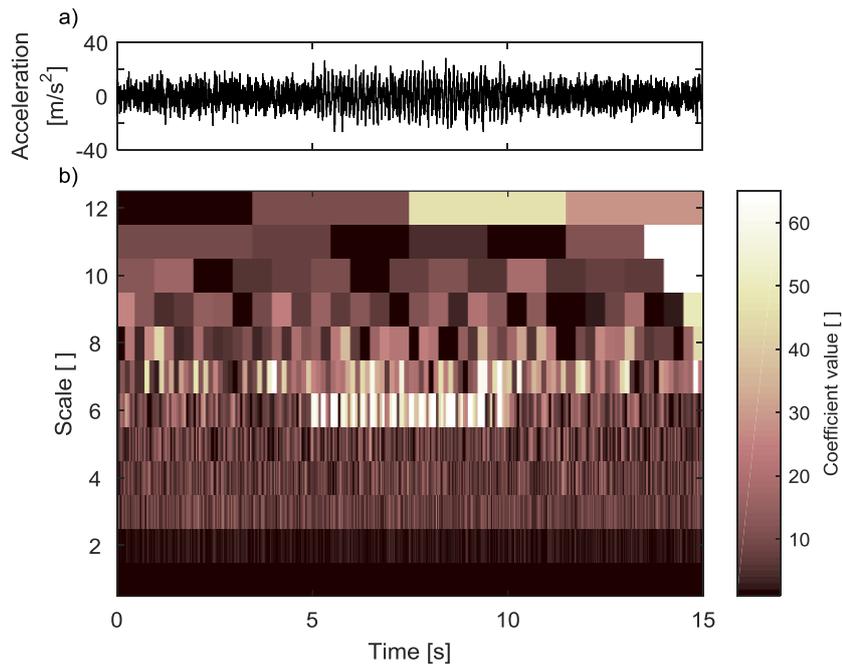


Figure 45: a) signal with harmonic component, b) its scalogram

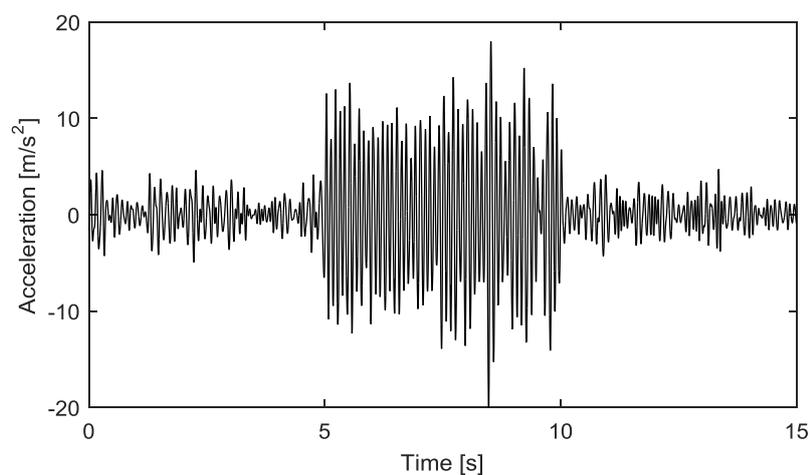


Figure 46: harmonic signal reconstruction using scale 6 of DWT coefficients

6.2.4 Conclusion on Wavelet Transform

The DWT is a well-established time-frequency (more specifically time-scale) analysis tool. Its application to RVV signals presents interesting features. The DWT using Daubechies 10 wavelet has good detection sensitivity for transient and harmonic components. However, the performance at detecting the nonstationary segment is poor due to the sensitivity of the DWT to signal state changes which creates misperception between shocks and nonstationary segment changes.

6.3 Conclusion on Time-Frequency Analysis Tools

The performance of the HHT and the DWT to analyse a synthetic RVV signal was assessed in this chapter. Both transforms have specific benefits and limitations. The DWT has the advantage over the HHT of being a more established technique with a fully proven mathematical framework. Many researchers have worked on its development and applications. The HHT does not have yet the theoretical background of the DWT but has important potential due to its complete adaptability and its simplicity (few user defined parameters).

On the test bench, the performance of the HHT was average. It detected the nonstationary components but lacked sensitivity for the shocks. The harmonic detection was good, but the frequency leakage between the IMFs (mode-mixing) did not allow frequency identification. The signal detection methods based on HHT (*i.e.* energy density/average period significance test and FRT) did not present sufficiently acceptable and reliable results due to the type of random signal created by the vehicle vibration spectrum.

On the other hand, the DWT presented promising results in detecting transient and harmonic components, but the HHT was better at detecting the nonstationarity segments. This shows that the HHT and the DWT complement each other well for a complex mixed-mode signal analysis such as RVV signals.

Chapter 7: MACHINE LEARNING

As seen in the previous chapters, both time and time-frequency analysis tools studied have their specific benefits and limitations when applied to RVV signals. The moving statistics such as the moving RMS, kurtosis and crest factor cannot clearly distinguish the three modes present in RVV signals. The HHT can detect the nonstationary segments of the synthetic RVV signal, but it does not have enough sensitivity to detect the transient and harmonic components. On the other hand, the DWT performs well at the transient and harmonic detection, but it shows poor performance at detecting nonstationary segments. An approach to get the most out of the time and time-frequency tools is to combine their outcome to perform the predictions. The combination can be done using machine learning processing. With the correct training, some machine learning algorithms should be able to detect and index shocks in nonstationary RVV signal using all analysis tools together. To simplify this first attempt to use machine learning to index the different RVV modes, only the shocks and nonstationary components are assessed, as they are the preponderant modes in terms of product damage during road transport. The same process could be applied to detect the harmonic components, as the analysis tools used in the machine learning algorithms can also detect this type of component. However this is beyond the scope of the thesis.

7.1 Machine Learning Overview

Machine learning is a branch of Artificial Intelligence that involves teaching (or training) a computer program to solve a problem. Once the training process is completed, the program can solve similar problems to those used for learning, using the relationship that was learnt during training. For instance, a computer can learn to recognise a person's face by analysing several pictures of that person. Once the training is done, the computer can recognise that person's face in other pictures containing other people. This type of machine learning where the analysed data are divided into discrete classes is called "classification problem".

There is also a second type of machine learning practice known as "regression problem". Instead of dividing a signal into discrete clusters or classes, regression algorithms find

relationships between the data variables. This is often used with time series, *i.e.* time dependant data. For instance, the electricity supply-demand of a city can be predicted using the time of the day, the date and the temperature forecast.

The same workflow is used for both classification and regression (Figure 47). It starts with a learning dataset which is a set of data where the classes are known. For vehicle vibration, this is an acceleration signal where the locations of the shocks are known. This dataset is processed to reveal different data behaviours and characteristics in a format compatible with machine learning algorithms. This is where machine learning is different to other classification approaches because it can base its prediction on several different signal processing methods. For instance, it can combine shocks and nonstationary analysis techniques to distinguish transient events (shocks) from signal intensity variations (nonstationarity). Once the processing is completed, the data is randomly partitioned into two sets: the training set and the validation set. Both sets have the same proportion of classes. The training set is used to train the algorithm and develop the classifier. The trained classifier is then validated using the validation set. Other approaches can be used to validate the model if there is not enough data to divide into two sets, such as cross validation where all the data are used for both training and validation, however this is not necessary to analyse RVV signals when a synthetic dataset of any size can be used. The validation phase is useful to compare the performance of different algorithms. The best algorithm is then used to generalise the classification or regression model to a new set of dataset.

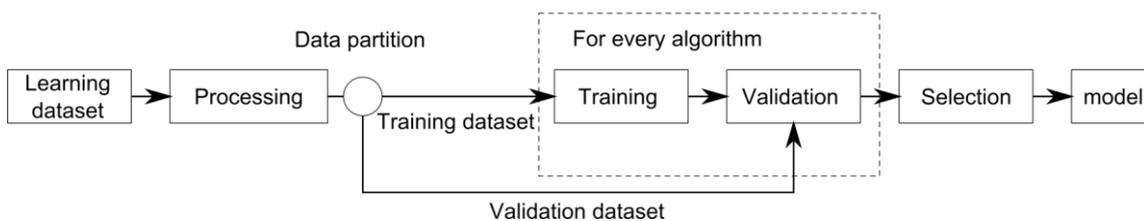


Figure 47: machine learning workflow

7.2 Predictors

Machine learning prediction performance depends on the data processing undertaken before the training phase. This processing reveals different signal characteristics and statistics called predictors. In order to detect the components of RVV signal, the predictors come from the relevant analysis techniques discussed in the previous chapter such as the moving RMS, moving crest factor, moving kurtosis, Hilbert-Huang Transform (HHT) and Discrete Wavelet Transform (DWT). The integration of the predictors in the classification learning process is shown at Figure 48.

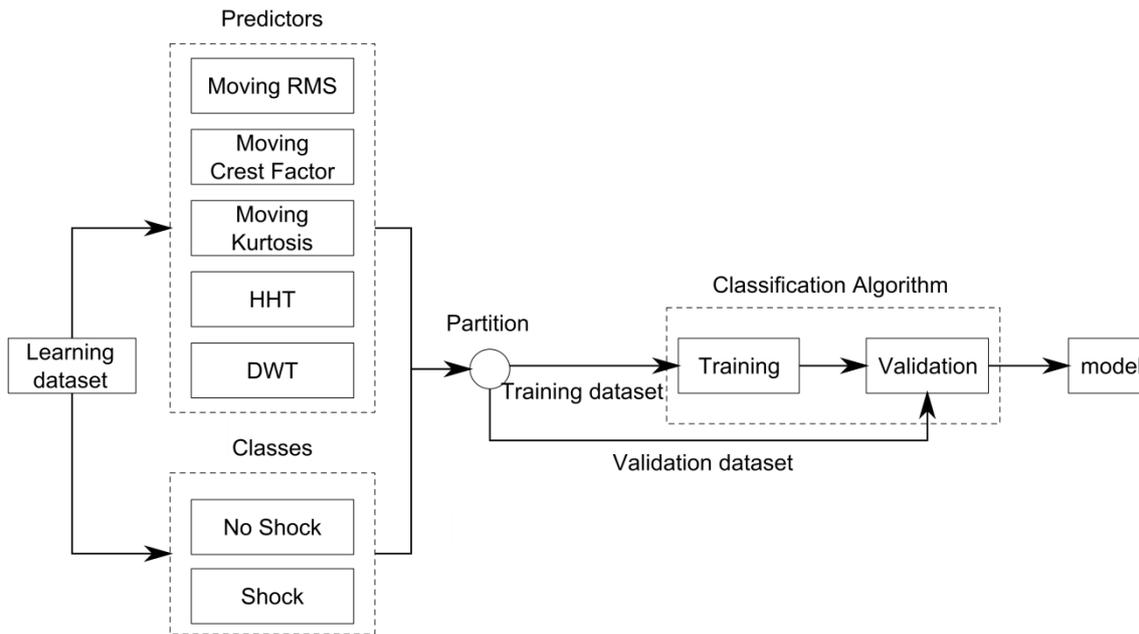


Figure 48: classification learning workflow for shocks detection

7.2.1 Moving RMS

The moving RMS can be used to characterise the nonstationary nature of RVV signals. The major shortcoming of this predictor is its dependency on its window duration, *i.e.* the segment length used to compute the RMS values. A shorter window is better to detect the short transient events, but is ineffective for long Gaussian changes and vice versa for a longer window. Therefore there is no ideal window size. Fortunately, machine learning classification has the capability to use multiple predictors. The moving RMS predictors are not limited to one window length, so two different window lengths (0.5 s and 4 s) are used. The shorter window is expected to be more responsive to shocks and the longer one to signal RMS variation (nonstationarities).

As for the following moving statistics, the moving RMS functions are computed forward in post processing to take into account the response of the system (eq. 13). In other words, at one moment, the RMS values represent the signal intensity of the next window. The classifiers base their prediction using both RMS values (short and long windows) for given moment in the signal. The discrete values of the moving RMS functions are the predictors (sufficient statistics) for every sampled time.

7.2.2 Moving Crest Factor

In general, the moving crest factor of a signal increases with the presence of a shock. Therefore shocks can be detected when the crest factor is above a certain threshold. As opposed to the moving RMS predictor, the moving crest factor predictor is more accurate when using a longer

moving window. This is because a longer window averages out the effect of the shock at the crest factor's denominator without affecting its numerator, which results in a greater sensitivity to shocks. However, windows of too long duration have been shown to misclassify short RMS variation as shocks. Based on Chapter 5, two crest factor predictors with window lengths of 8 s and 64 s are used to develop the machine learning classifiers.

As for the moving RMS, the discrete values of the moving crest factor functions (eq. 16) are the predictors for every sampled time.

7.2.3 Moving Kurtosis

The moving kurtosis is the fourth moment of a signal. It gives a measure of the "peakedness" of the Probability Density Function (PDF) of a signal segment. Transient and nonstationarity components create leptokurtic distribution (kurtosis > 3). Depending on the window length, the moving kurtosis does not have the same sensitivity to signal RMS variations and shocks. By combining two window lengths (4 s and 8 s), the moving kurtosis functions could identify shocks within a nonstationary signal. The discrete values of both functions (eq. 14) are used machine learning predictors for every sampled time.

7.2.4 Hilbert-Huang Transform Predictors

The HHT is an adaptive time-frequency analysis method providing different types of predictors from RVV signals. The HHT divides the signal into different narrow-banded components called IMFs which provide information that can be useful for the classification. The IMFs are not directly used as predictors because their properties are not explicit and vary too much in time to give accurate statistics. Instead, the IMFs instantaneous amplitude and frequency functions are used as predictors. These functions vary slowly and reveal trend changes in the signal which makes them more efficient predictors. Figure 49 shows for example IMFs 1, 5 and 9 of a signal along with IMF 5's instantaneous amplitude envelope and instantaneous frequency functions.

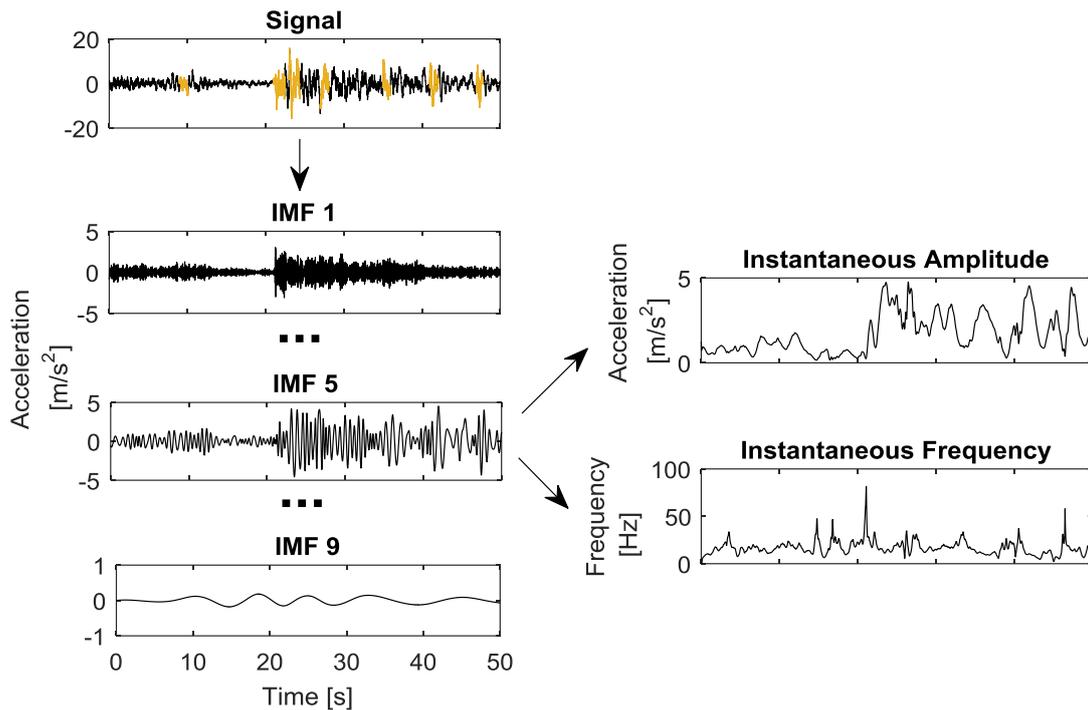


Figure 49: example of the synthetic signal (shocks are in yellow) and the HHT predictors: IMF number 1, 5 and 9; instantaneous amplitude and frequency functions of IMF number 5

As presented more in detail in Appendix C, oversampling signals have many advantages for shock analysis but Empirical Mode Decomposition (EMD) does not cope very well with it. Because there are very few high frequency components in RVV signals, the first IMFs (which are composed of the high frequency components) are difficult to compute using EMD. In this case, the EMD sifting process is not stable and cannot clearly isolate the signal's modes which results in a substantial amount of leakage and mode drifting between the IMFs. Reducing the sampling frequency has been shown to minimise this effect by ensuring constant level of frequency content in the first IMF. Therefore, the EMD naturally acts as a filter bank and the frequency range of every IMF is steadier. Fewer sifting iterations are required to obtain narrow banded IMFs, which greatly improves the computing speed, especially when applied to large and mixed-mode signals such as RVV signals. The synthetic RVV signal is sampled at 1024 Hz and its spectral density is negligible after 50 Hz. Resampling the signal by a factor 10 creates a Nyquist frequency that is above 50 Hz which is sufficient to cover the signal's spectrum. A low-pass filter is applied prior to the decimation to avoid any aliasing (Chebyshev Type I IIR filter of order 8 with a cut-off at 50 Hz).

The sampling frequency of the predictors issued from HHT is brought back to its original value to match the other predictors. A linear interpolation is used to resample the predictors.

7.2.5 Discrete Wavelet Transform

The DWT is another time-frequency (or more specifically, time-scale) analysis method that provides predictors which are more sensitive to signal changes and can be useful for shocks detection. The DWT coefficients issued from the Daubechies 10 wavelet analysis are directly used as predictors. Figure 50 shows as an example the DWT predictors of four different scales of signal.

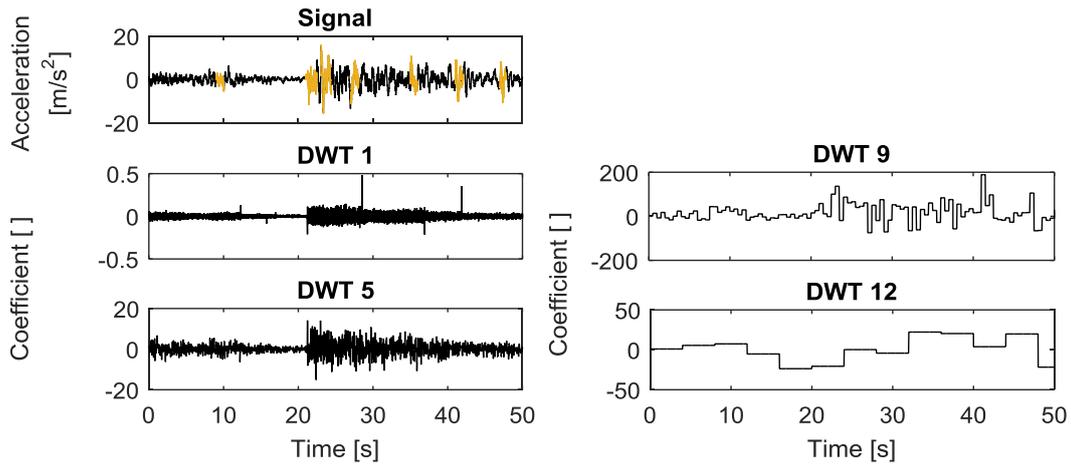


Figure 50: example of the synthetic signal (shocks are in yellow) and DWT number 1, 5, 9 and 12

During the DWT analysis, the signal sampling rate is halved for every scale. The number of scales is limited to 12, so the largest scale has a frequency range up to 0.25 Hz (for a sampling rate of 1024 Hz) which can be considered to be refined enough for RVV analysis purposes. This resampling also causes the number of coefficients to decrease at every scale. To create predictors with sampling rates that match the original signal and the other predictors, the coefficients are replicated to match the sampling rate of the signal as shown in Figure 51.

Signal:	$x(t) =$	$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$	$x(6)$	$x(7)$	$x(8)$
Coefficients scale 1:	$a_1(n) =$	$a_1(1)$		$a_1(2)$		$a_1(3)$		$a_1(4)$	
Predictor 1:	$p_1(t) =$	$a_1(1)$	$a_1(1)$	$a_1(2)$	$a_1(2)$	$a_1(3)$	$a_1(3)$	$a_1(4)$	$a_1(4)$
Coefficients scale 2:	$a_2(n) =$	$a_2(1)$				$a_2(2)$			
Predictor 2:	$p_2(t) =$	$a_2(1)$	$a_2(1)$	$a_2(1)$	$a_2(1)$	$a_2(2)$	$a_2(2)$	$a_2(2)$	$a_2(2)$

Figure 51: example of the first two DWT predictors' samples

7.3 Classifiers Training

The classifiers used in this chapter were trained using synthetic signals containing nonstationary and transient components only which are considered to be the two most important components for distribution packaging optimisation purposes. The same process can be applied to the

harmonic components which could be part of further studies on the topic once a good shock classifier model is fully developed.

In machine learning, data is king. In general, the classifier prediction accuracy depends on the quantity of data used in its training. In the case of the detection of shocks in a RVV signal, it is unrealistic to gather significant amounts of data on real vehicles to train a classifier because that requires a survey of many kilometres of road profile to know exactly where the aberrations are and to drive a vehicle exactly on the surveyed path. It is more appropriate to use synthetic acceleration signals that mimic typical RVV as there are no length limitations and the signal components are accurately known *a priori*.

To define the optimal training dataset size, the classifier training is performed using various signals of different duration (100 s to 10,000 s). These are created using the RVV signal synthesiser of Chapter 4 with equivalent nonstationary and transient properties (Figure 52). Every 100 s segment is composed of at least 15 stationary Gaussian segments of different intensity (distribution in Figure 12) and length randomly distributed (distribution in Figure 13).

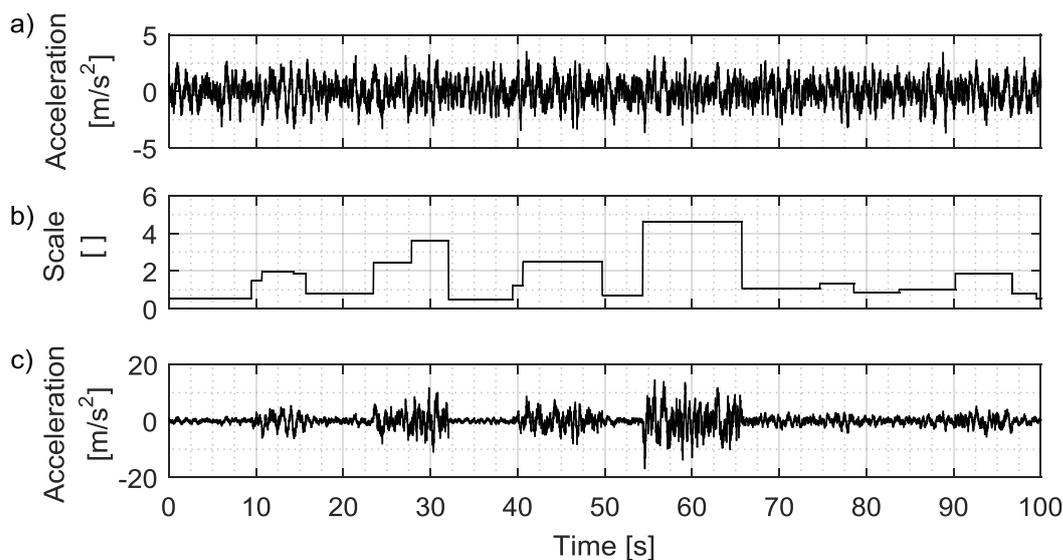


Figure 52: nonstationary signal synthesis: a) stationary Gaussian signal based on RVV spectrum; b) modulation function; c) nonstationary signal with Gaussian sections

These 100 s nonstationary segments are superimposed on 10 randomly distributed vehicle shocks responses to different impulses of amplitude and duration ranging between 5 mm and 40 mm and 0.5 s and 1.4 s, respectively (Figure 53 a). These impulses generate vehicle transient responses (*i.e.* shocks) with maximum absolute amplitudes ranging between 1.6 m/s^2 and 22 m/s^2 (Figure 53 b).

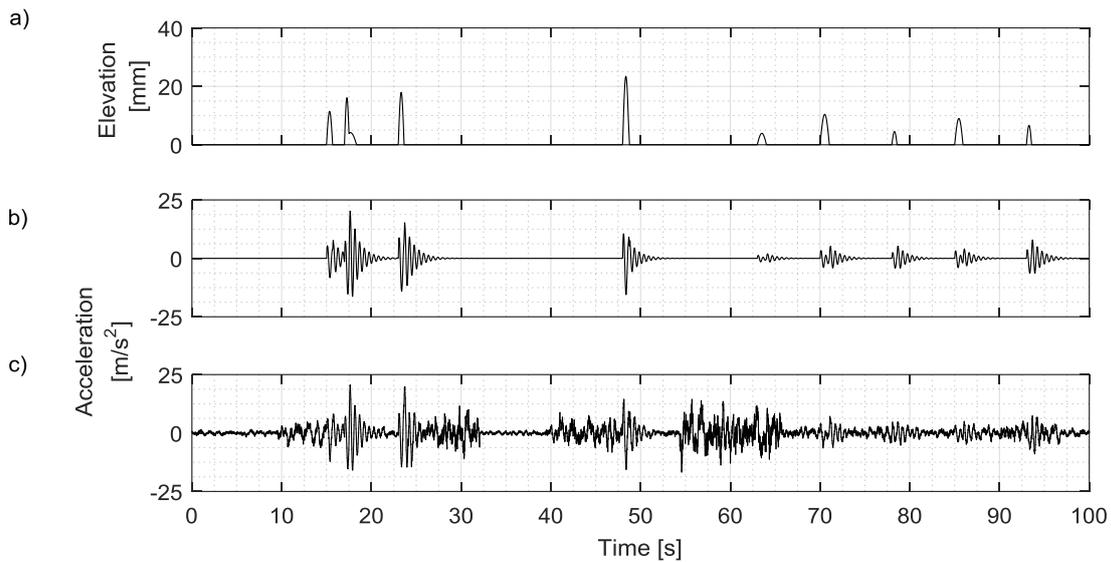


Figure 53: transient signal synthesis: a) impulse function; b) vehicle shock responses; c) vehicle shock responses superimposed on a nonstationary signal

The signal format is set to a single-precision floating point which uses 32 bits instead of 64 bits for the usual double-precision format. This increases the learning and fitting speed by removing the unnecessary precision of the signal. The process is also sped-up by reducing the sampling rate of the predictors by the factor of 64. The signals used for the training have a sampling frequency of 1024 Hz to ensure the peaks have a good resolution. The predictors are computed on these 1024 Hz signals to take advantage of this extra resolution. However there is no need to keep this resolution because the impulse functions are longer than 0.5 s. So the predictor sampling frequency is reduced to 16 Hz which according to Nyquist theorem is enough to analyse 0.5 s period components. To avoid aliasing, a low pass filter is applied before the predictor decimations (Chebyshev Type I IIR filter of order 5 with a cut-off at 8 Hz).

The class identification needed for the classification (Figure 48) is defined using the input function shown in Figure 53 a). The class is “no-shock” or “-1” when the input function is zero and “shock” or “1” when it is not zero.

7.4 Classifiers Validation

The classifier validation is an essential step of machine learning; it assesses the detection accuracy. The validation dataset used to assess the classifiers lasts 500 s and has the same characteristics as the training dataset (*e.g.* nonlinear random vibration and shocks distribution, 10 shocks per 100 s of signal). The position of the shocks is known and is used to assess the detection accuracy.

7.4.1 Detection Enhancement Algorithm

The detections made by the selected classifiers are discrete and independent of the predictors' order. In other words, the classification at one time is not affected by and does not affect the previous and future classifications. Therefore the causality of the shock responses is only considered in the predictor computations and not in the classification algorithms. This means that the transient detections can be scattered rather than continuous. For instance, the shock buried in the signal in Figure 54 between 48 s and 48.5 s is detected in two distinct segments. These incomplete detections decrease the classifier accuracy.

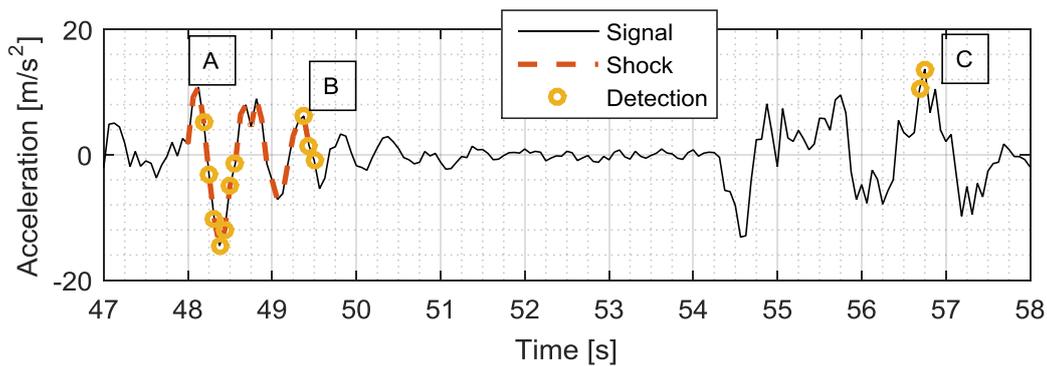


Figure 54: typical discontinuous shock detection of example signal grouped in three continuous sequences (A, B and C)

An algorithm was specially developed to enhance the continuity of the detections for RVV analysis and shock detection. The detection enhancement algorithm extends the detection sequences to ensure they have at least the same duration as the longest impulse function (*i.e.* 1.4 s) which is the period of the first natural frequency of the vehicle shock response. This is more than the period of the natural frequency of the two degree of freedom vehicle model used to synthesise the signal.

The algorithm is an iterative process described in Figure 55. It starts at the detection point with the maximum absolute value of the signal. It then creates a 1.4 s window starting 0.28 s before the absolute maximum. This places the absolute maximum at 20 % of the window length which approximately places the window at the beginning of the vehicle shock response. The algorithm considers all the data points in this window as a shock if at least 10 % of these data points were classified as shock by the classifier as shown in Figure 55 a and b. In this case the window points and the continuous detections adjacent to it are considered as a shock segment. This 10 % overlap criterion is based on an arbitrary value which only affects the classifier's operating point threshold (see section 7.4.2). If the 10 % detection overlap criterion is negative, the window is classified as no-shock and the detection points made by the classifier are voided as shown at Figure 55 b and c. This concludes the first iteration. The algorithm continues this iterative

process by finding the next maximum absolute detection acceleration excluding the points from the segment created in the previous iterations. The windowing process is then performed and the 10 % detection overlap criterion is applied using all the data points, including those from the previous iterations.

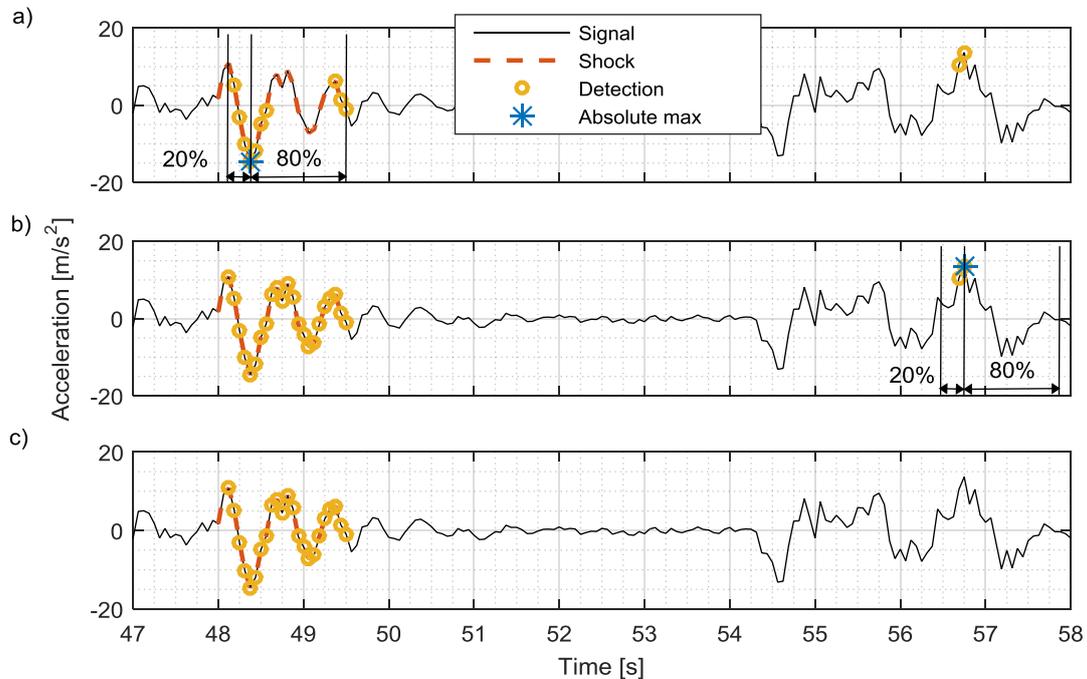


Figure 55: detection enhancement of example signal; first iteration (a) and second iteration (b) of the extension algorithm and (c) final result

7.4.2 Detection Accuracy

Classifiers base their prediction on a sufficient statistic which is calculated from predictors. The sufficient statistic is compared to a threshold value to attribute a class to a data point. This can be seen as the probability that a data point belongs to one class over the other. For various reasons explained in section 8.1, the classification is not necessarily made when there is more than a 50 % chance that a data point is from a specific class; other probabilities or threshold values are often used. Changing this threshold changes the number of true-detections, misdetections and false-detections which can be assessed in terms of sensitivity and specificity.

In this context, the sensitivity is defined as the proportion of shock data points in the signal that are correctly classified as such. Using this definition, the example signal's sensitivity (Figure 54) went from 45 % to 90 % through the enhancement process (Figure 55 c). For RVV signal analysis purposes however, this detection could be considered perfect (sensitivity equals 100 %) because the objective is to characterise the transient amplitude and occurrence. Therefore, the sensitivity is assessed in the thesis using a slightly different definition than usual. The transients

are considered as a sequence of data points instead of individual data points. The sensitivity definition becomes the number of true detection sequences over the total number of shocks,

$$\text{Sensitivity} = \frac{\text{Number of True Detections}}{\text{Number of shocks}}, \quad (\text{eq. 20})$$

where a detection has to overlap at least 75 % of the shock duration to be considered true. As for the 10 % overlap criterion, this is an arbitrary value which only affects the classifier's threshold value.

As opposed to sensitivity, the specificity is the proportion of the signal without shock that is correctly classified as such. This represents the true negative rate of the classifier. The definition of the specificity is not modified and it is considered as the number of true nondetections data points over the number of data points without transients,

$$\text{Specificity} = \frac{\text{Number of Nondetection Data Points}}{\text{Number of Data Points without shocks}}. \quad (\text{eq. 21})$$

The sensitivity and specificity are specific to an operating point, *i.e.* the threshold value used for the detection. Using a low detection threshold increases the sensitivity to the detriment of the specificity and vice versa. This relationship is represented by the Receiver Operating Characteristic (ROC) curve which displays the sensitivity as a function of the fall-out (*i.e.* probability of a false detection, or $1 - \text{specificity}$). The ROC curve always starts at the coordinates (0, 0) because when the detection threshold is maximal (*i.e.* very low detection sensitivity), there are no detections, thus there cannot be any true or false detections. It also always ends at the coordinates (1, 1) because when the detection threshold is reduced to a minimum (*i.e.* very high detection sensitivity), the whole signal is considered as a detection. Therefore all the shocks are necessarily detected which means the sensitivity equals one. This also implies that all the signal segments without shocks are falsely classified as shock which means the fall-out equals one. As the detection threshold is proportional to the amount of detection, the ROC curve is necessarily a monotonic function. However because of the enhancement algorithm used on shock detection, there could be some slight divergences to this property in some ROC curves presented in this thesis.

To illustrate how the ROC curve works, three typical curves are presented in Figure 56. The ROC curve of the ideal classifier would be a vertical line that goes from 0 sensitivity to 1 and the ROC curve of a classifier only based on chance (guessing) would be the diagonal going from (0, 0) to (1, 1). For the typical classifier, the ROC curve sensitivity rises rapidly and then reaches a plateau. The curve is above the diagonal line which means the typical classifier does better than prediction based on chance.

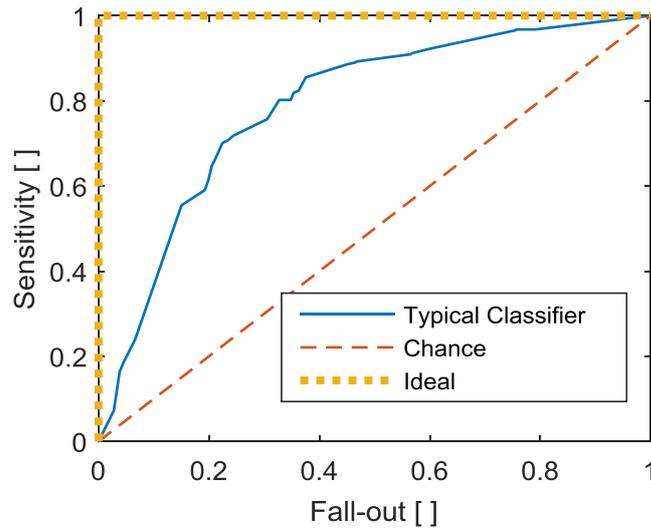


Figure 56: typical ROC curves

A simple method to assess and compare classifiers' performance is to calculate the Area Under the ROC Curve (AUC) (Bradley, 1997, Ling et al., 2003). The ideal classifier has an AUC of 1 and a classifier based on chance has an area of 0.5. This means that a classifier with an AUC above 0.5 has better prediction performance than chance. As pointed out by Fawcett (2004), the ROC curves are stochastic rather than deterministic. So without a measure of the variance, classifiers cannot be compared. Fawcett calculates the variance of every points of a ROC curve using several validation datasets. However to compare AUC values, it is more practical to calculate the variance directly from the AUC values calculated from different validation datasets.

All the ROC curves have the same abscissa and ordinate ranges (both 0 to 1) which mean the AUC is equivalent to the average value. According to the central limit theorem, the distribution of the AUC values can therefore be considered Gaussian. Hence, the confidence interval of the AUC values is calculated as:

$$\text{c.i.} = \mu \pm t_{\alpha, (n-1)} \frac{\sigma}{\sqrt{n}} \quad (\text{eq. 22})$$

where μ and σ are, respectively, the approximate mean and standard deviation of the AUC values calculated on validation datasets of the same size and composition, $t_{\alpha, (n-1)}$ is the Student's t -distribution for a confidence level, α , and the number of validation datasets, n . Classifiers with a relative large standard deviation can indicate that a portion of their predictions are based on chance. In this chapter, every AUC value is presented with its 95 % confidence interval calculated on 11 validation datasets composed of 50 shocks (500 s of duration).

7.5 Classifiers Evaluation and Selection

An important number of machine learning classifiers have been developed and selecting the best classifier for a new application can be cumbersome. There is no standardised nomenclature in machine learning. Therefore, the same or very similar classifier algorithms have completely different names. Some algorithms are also empirically modified for specific application and are not necessarily optimised for others. Since 2006, Mathworks® regrouped and implemented the most common classifier algorithms in Matlab® 2015a Statistics and Machine Learning Toolbox.

These classifiers are divided into seven groups: Discriminant Analysis, Naïve Bayes Classification, Ensemble, Decision Trees, Nearest Neighbours, Support Machine Support and Neural Network. These groups are briefly introduced here and for more information refer to Hastie et al. (2005), Rogers and Girolami (2011), Cherkassky and Mulier (2007), Shalev-Shwartz and Ben-David (2014).

The performance of the classifiers is assessed with four parameters: the AUC value which provide the prediction accuracy; the prediction speed which is the time required to analyse the validation dataset, memory requirement of the classifier and the minimum training dataset length needed to have maximal prediction accuracy. A basic optimisation is performed on the presented classifiers in order to identify those that are suitable for RVV analysis. In the next chapter, the parameters of the identified classifiers are refined to provide optimal shocks detection performance.

7.5.1 Discriminant Analysis

Discriminant Analysis is the most common classifier because of its simplicity. It uses a Gaussian distribution to predict the class of dataset. The learning phase consists of fitting a Gaussian function to every class for every predictor. To classify a new dataset, the algorithm minimises the misclassification error (or cost) using these functions. However, this type of algorithm is not compatible with RVV analysis because of its complexity, its number of predictors and, more importantly, its non-Gaussian nature of RVV, represented by high kurtosis values (see Chapter 4 and 5) causes the different vibration components.

7.5.2 Naïve Bayes Classification

Similar to the Discriminant Analysis, the Naïve Bayes Classification fits distribution functions on the learning dataset. The difference is that these functions do not need to be Gaussian. An important aspect is that every predictor must be independent. Handling non-Gaussian jointed distribution becomes rapidly overwhelming and the Naïve Bayes Classification algorithm cannot cope with this. This type of algorithm is therefore not suitable for RVV signal analysis.

This is because there are many redundancies between the predictors (*e.g.* two time-frequency methods are used, moving statistics are repeated with different window length) which makes the predictors highly dependent on each other.

7.5.3 Decision Trees

The Decision Tree is also a relatively simple algorithm to classify data. The classification is conducted with a cascade of statistical tests as shown in Figure 57. These tests consist of value comparisons. For instance, a test is true if the predictor value is greater than a certain threshold, otherwise it is false. Tests can sometimes have more than two outcomes and different paths can lead to the same final class.

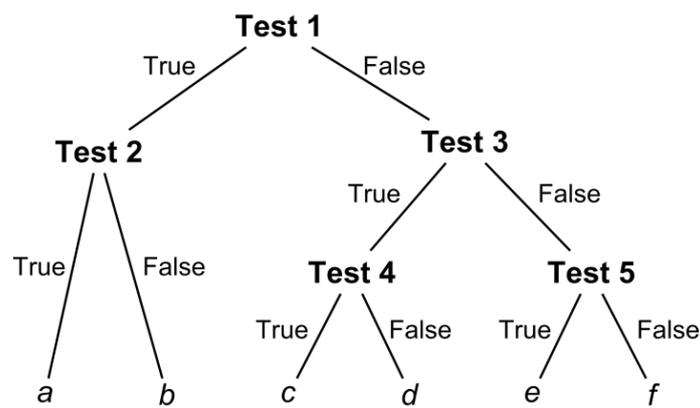


Figure 57: example of Decision Tree made up of 5 binary tests (splits) and 6 final outcomes (a-f)

The first tests use the predictors that have the most discriminant power, *i.e.* that have the biggest impact on the classification. These tests are defined through an iterative process during the algorithm learning. The tree's complexity is defined by its number of branch splits (or tests). For instance, a Decision Tree can be qualified as having a medium complexity if it uses the most relevant predictors as branch splits. More complex trees can be created by using the same predictors more than once. For instance in Figure 57, the predictor used for Test 2 can also be used on the right side of the tree at Test 4 and the thresholds used for both tests are not necessarily the same.

A three splits Decision Tree shows the three predominant predictors of the training dataset which are the 9th DWT coefficient, the instantaneous frequency of IMF 4 and the 8th DWT coefficient (Figure 58). As the several predictors are computed on the training dataset, the complexity of the decision tree rapidly increases making its analysis difficult.

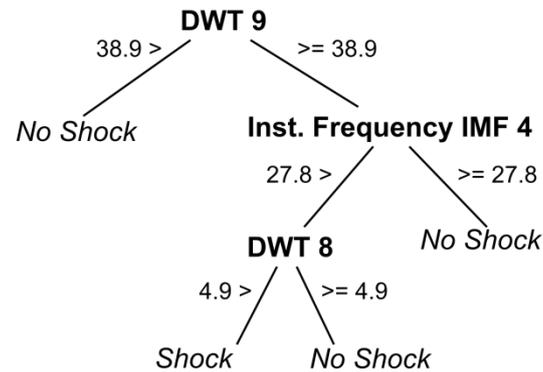


Figure 58: first three splits of Decision Tree detecting RVV shocks

Depending on their complexity, Decision Trees generally have a medium predictive accuracy, a fast fitting and prediction speed, low memory usage and are easier to interpret. The optimum learning dataset size also depends on their complexity. As shown in Figure 59, a Decision Tree with 20 splits quickly reaches an AUC plateau between 0.79 and 0.81 for training datasets above 50 shocks. The more complex 82 splits Decision Tree does not clearly reach a plateau because its AUC value drops with a certain training dataset size. Its AUC values are also generally inferior to those of the 20 splits Decision Tree. This suggests that its extra splits only add more uncertainty to the classification and that a more compact Decision Tree has a more reliable prediction. Both trees perform fast classification and can analyse a 500 s signal in less than 0.05 sⁱⁱ and require between 20 KB to 60 KB of memory. According to this initial optimisation, the more appropriate Decision Tree classifier would have 20 splits and be trained from a 200 shocks training dataset (AUC of 0.80).

ⁱⁱ Using Windows 7 64 bit with Intel® Xeon® E5-1650, 6 cores, 12 threads, 3.80 GHz, processor and 32 GB of RAM

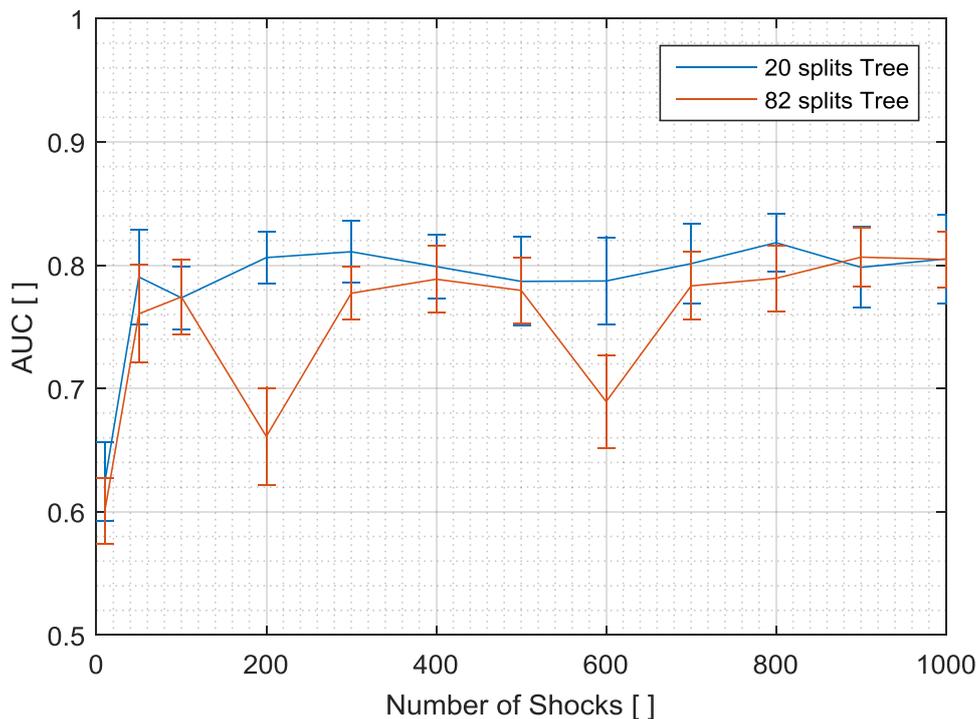


Figure 59: decision Trees' performance as a function of the number of shocks in training dataset size (10 shocks per 100 s of signal; 95 % confidence interval)

7.5.4 k -Nearest Neighbours

The k -Nearest Neighbours algorithm (k NN) groups the training dataset by class in as many spaces as there are predictors. It classifies new data points by grouping them to the most common class of their k nearest neighbours. In other words, the classification is made by association. If the majority of a data point's nearest neighbours belong to one class, it is more likely that the point belongs to this class too. The accuracy of the classification depends on the number of nearest neighbours, k , used in the algorithm. A k value that is too small could lead to an over fitted model and an excessively large value could lead to an insensitive model which in both cases would provide inaccurate classification.

For instance, the Figure 60 presents data points where 15 are squares, 5 are circles and the lozenge class is unknown. The lozenge's first nearest neighbour is a square which means it would be classified as such by a 1NN classifier. However this particular square seems to be an outlier because it is at the bottom right corner of the figure when all the other squares are gathered at the upper left corner. This means the lozenge would be more likely misclassified by a 1NN classifier. When more than three nearest neighbours are included in the classifier algorithm, the lozenge's predicted class changes to circle, which seems a more appropriate classification. However, because the number of circles in this example is limited to five occurrences, if the classifier uses more than 11 nearest neighbours any predictions would

necessary be squares and create systematic misclassification. It is therefore important to select an appropriate number of nearest neighbours that will give the more accurate classifications.

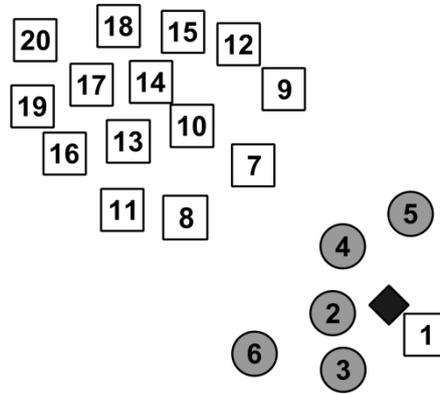


Figure 60: *k*NN process, the lozenge is a new data point and the two classes of the training points are the square and circle numbered from the nearest neighbours to the farthest

Another parameter that can change the *k*NN classifier's accuracy is the types of distance used to find the nearest neighbours. The most common distance is the Euclidean distance or the straight-line between two points. It is mathematically defined as the square root of the squared difference of all n dimensions (predictors) between points x and y ,

$$d_{\text{euclidean}} = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}. \quad (\text{eq. 23})$$

In the context of shock detection, 36 detectors are used. Therefore, the distance calculation is made in 36 dimensions.

Another way to define the distance is to consider the data points as vectors and compute the collinearity between them. This is called the cosine distance, which is the angle between the inner product of two vectors (x, y) expressed as:

$$d_{\text{cosine}} = 1 - \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}}. \quad (\text{eq. 24})$$

The *k*NN algorithms lose accuracy when the classes are not regrouped in well separated clusters, because there are no clear boundaries between them. One way to overcome this is to weight the contribution of the neighbours depending on their distance from the data points to be classified, such as,

$$\text{Weight} = \frac{1}{\text{Distance}} . \quad (\text{eq. 25})$$

This weighting algorithm ensures that the closer neighbours have more influence on the classification than the farthest ones.

The computation speed of k NN algorithms is considered medium to slow. It decreases with the number of predictors and the size of the training dataset because they are the reference used to classify new data points. For instance, the tested k NN algorithms can classify a 500 s RVV signal in around 16 s per 100 s of dataset used in the algorithm trainingⁱⁱⁱ. Therefore a k NN classifier training with a 1500 s signal (300 shocks) will take 240 s to analyse a 500 s signal. The memory required for the classifier can also be substantial because all of the training dataset has to be included in the algorithm. Fortunately, the memory requirement is not too important for this study in which only about 2.7 MB are required per 100 s of training dataset.

There are three elements that define the k NN classifier accuracy: the number of nearest neighbours, the distance and the training dataset length. Figure 61 summarises this information by presenting the AUC of different k NN algorithms as a function of the number of shocks used in the training dataset. To keep the figure clear, the confidence intervals are not shown because they are similar at every point (ranging between 0.022 and 0.042). The 100NN classifiers provide slightly more accurate results than 10NN classifiers. This suggests that using only 10 nearest neighbours may overfit the data. There is not significant accuracy difference between Euclidean and cosine classifiers at the same number of nearest neighbours. The weighting process does not significantly enhance the classification.

As it is generally good practice to use the simplest model or algorithm, the 100NN Euclidean classifier would be the more appropriate k NN algorithm to detect shocks in RVV signals. This classifier reaches an AUC plateau at 0.84 with a training dataset of more than 300 shocks.

ⁱⁱⁱ *Ibidem* ii

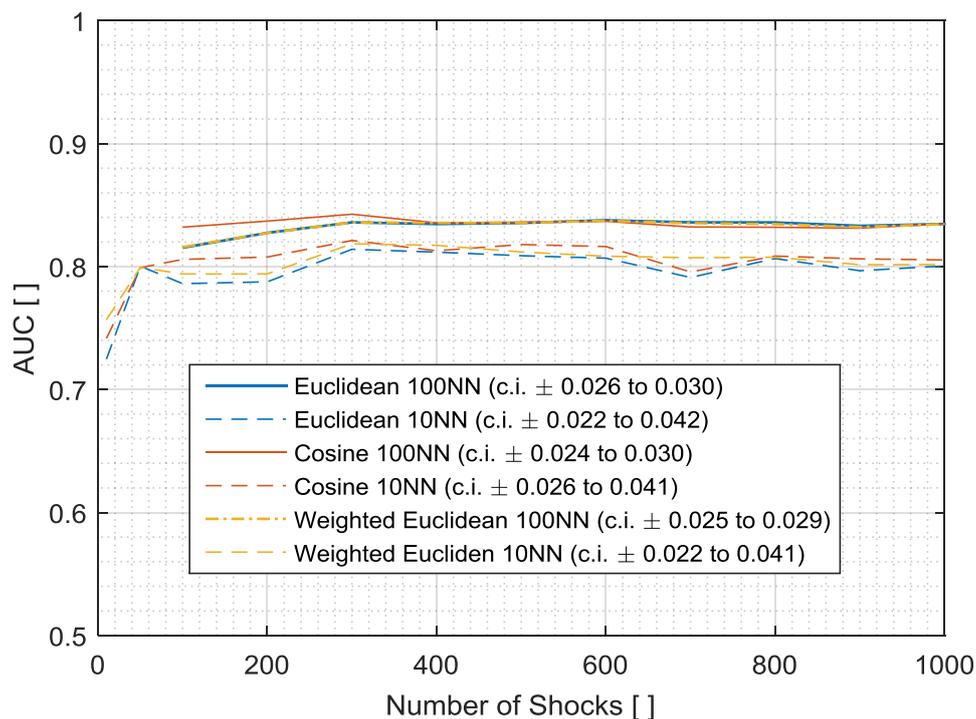


Figure 61: *k*NN classifiers' performance as a function of the number of shock in training dataset size (10 shocks per 100 s of signal; 95 % confidence interval)

7.5.5 Ensemble

Ensemble learning uses an ensemble of simpler classifiers to increase the classification performance. These simpler classifiers are for instance Classification Trees, *k*NN or Discriminant Analysis. Ensemble algorithms merge the outcome of an ensemble of classifiers of the same type to enhance their classification. The size of this ensemble goes from a few tens to many hundreds of classifiers. The ensemble size increases both the classification accuracy and processing time. When too large, an Ensemble algorithm can be overtrained and leads to an inaccurate prediction. As for the *k*NN, Ensemble models can be interpreted using the importance of every predictor on their outcomes.

Different categories of algorithms are commonly used such as Subspace, Boosted and Bagged (Cherkassky and Mulier, 2007, Rogers and Girolami, 2011). Subspace algorithms are designed to work best with *k*NN algorithm. They basically divide the training dataset into different subspaces and applies a *k*NN classifier on each of them. The algorithm then classifies new data points with every *k*NN classifier and use the average prediction to define their class. The signal subspaces are randomly selected and may overlap each other.

Bagged and Boosted algorithms use a combination of Trees on which they base their prediction. Bagged algorithm stands for “bootstrap aggregation”. It generates Decision Trees on different

random data samples. The classification is made by using the average response of each of these Decision Trees. Boosted algorithm is based on the same principle as the Bagged algorithm but it adds a weight to every Decision Tree prediction that minimises the mean-square error of the prediction. This algorithm also tends to generate shallower Decision Trees than the Bagged one and generally needs a larger ensemble. The Boosted algorithm works best with evenly distributed classes. A modification of the algorithm called RUSBoosted (Random UnderSampling) can be used for highly skewed data, *i.e.* where there are fewer occurrences of one class as is the case for the amount of shocks presented in the synthetic signal used for the learning. This modified version of the Boosted classifier ensures there are the same number of class occurrences in every data sample.

The Subspace algorithms can be slower than the RUSBoosted and Bagged algorithms because they use an ensemble of slower classifiers (*i.e.* k NN). As for the k NN, the computing time increases with the length of the training dataset. For instance, a Subspace composed of an ensemble of 200 k NN classifiers can classify a 500 s signal in around 50 s per 100 s of dataset used to train the classifier^{iv}. This is about only three times slower than a standard k NN classifier. The Bagged and RUSBoosted algorithms do not depend on the training dataset and can analyse the same 500 s signal in 20 s and 7.5 s, respectively, using an ensemble of 200 Decision Trees.

The Ensemble classifiers have a more extensive memory requirement as their algorithm is composed of many simpler classifiers. The Subspace has the biggest requirement with 15 MB per 100 s of training dataset. This means that the 1000 shocks training dataset set (10,000 s) creates a 1.5 GB classifier. The Bagged classifier requires about half this amount, *i.e.* 87 MB per 100 s and the RUSBoosted requires much less, *i.e.* 1.5 MB for any training dataset sizes.

The Subspace algorithm did not improve the k NN classification performance. Figure 62 shows that, for any training dataset the AUC values of the Subspace classifier are inferior to the k NN classifiers' values presented in Figure 61. The Bagged and RUSBoosted algorithms slightly improve the Decision Trees' performances. Their AUC values plateau respectively at 0.85 and 0.83 compared to 0.81 for the best performing Decision Tree. The AUC plateau is reached around 50 shocks for the RUSBoosted classifier and around 300 shocks for the Bagged classifier.

^{iv} *Ibidem* ii

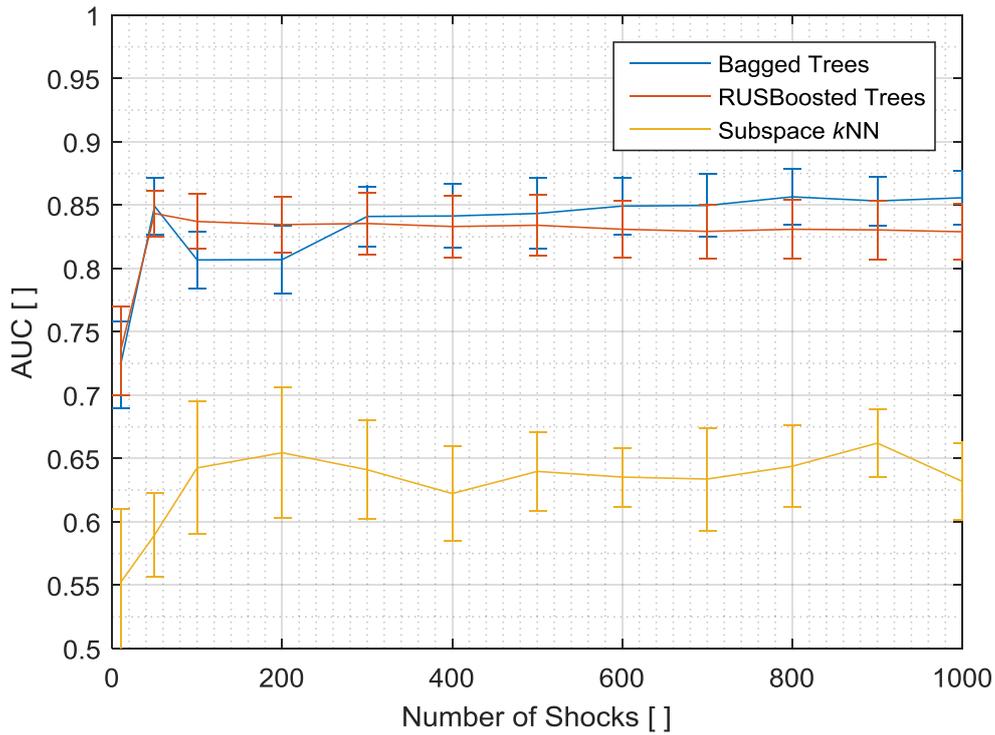


Figure 62: Ensemble classifiers' performance in function of the number of shocks in training dataset size (10 shocks per 100 s; 95 % confidence interval)

7.5.6 Support Vector Machine

Support Vector Machine (SVM) algorithms are powerful classifiers that only work when the data have only two classes. So they can be used as shock detectors for instance. The mechanics of SVM is as follows; during the learning phase, the algorithm finds a hyperplane that maximise the distance between the two classes' data points represented in all their dimensions (predictors). As for the k NN, there are 36 dimensions. This hyperplane is defined as:

$$D(\mathbf{x}) = \sum_{i=1}^n \hat{\alpha}_i y_i (\mathbf{x} \cdot \mathbf{x}') + \hat{b}, \quad (\text{eq. 26})$$

where (\mathbf{x}_i, y_i) , for $i=1, \dots, n$ are, respectively, the predictors matrix and class vector of the n training data points and $(\mathbf{x} \cdot \mathbf{x}')$ is the matrix predictors inner product. The parameters $\hat{\alpha}_i$ are found by maximising the function:

$$L(\alpha) = \sum_{i=1}^n \alpha - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j), \quad (\text{eq. 27})$$

constrained by

$$\sum_{i=1}^n y_i \alpha_i = 0 \quad (\text{eq. 28})$$

$$0 \leq \alpha_i \leq C/n, \quad \text{for } i = 1, \dots, n,$$

where C is a regulation parameter.

The parameter \hat{b} is used to scale the Support Vector (\mathbf{x}_s, y_s) such as only the first class exists in $D(\mathbf{x}) \geq 1$ and only the second class exists in $D(\mathbf{x}) \leq -1$:

$$\hat{b} = y_s - \sum_{i=1}^n \hat{\alpha}_i y_i (\mathbf{x}_i \cdot \mathbf{x}_s) . \quad (\text{eq. 29})$$

Figure 63 shows the SVM process for a two dimension classification problem. The zone between -1 and 1 regroups the nonseparable data points. These points are classified using $D(\mathbf{x}) = 0$ as the boundary line.

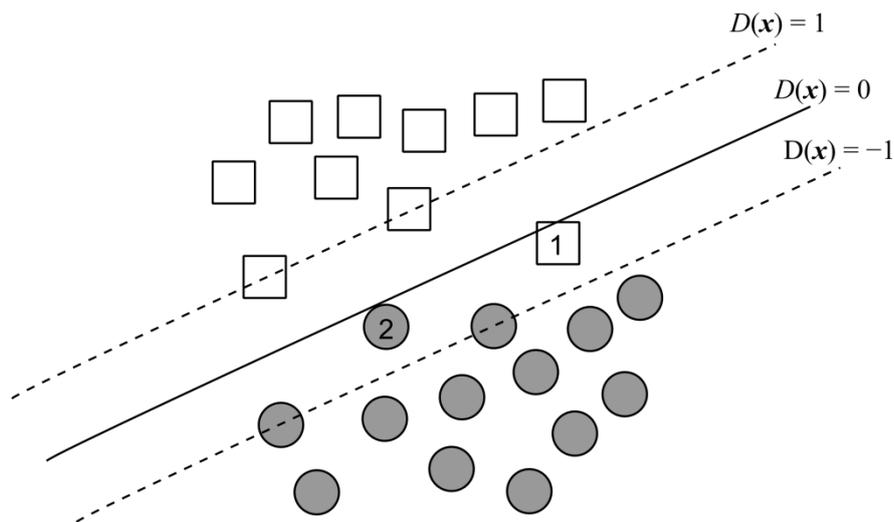


Figure 63: SVM process, 1 and 2 are nonseparable and 1 is also misclassified

There are many cases where it is impossible to solve (eq. 26) in its direct form because no single hyperplane can separate the data classes. For example, the dataset shown in Figure 64 cannot be divided by a single hyperplane because the cluster of circles is surrounded by squares.

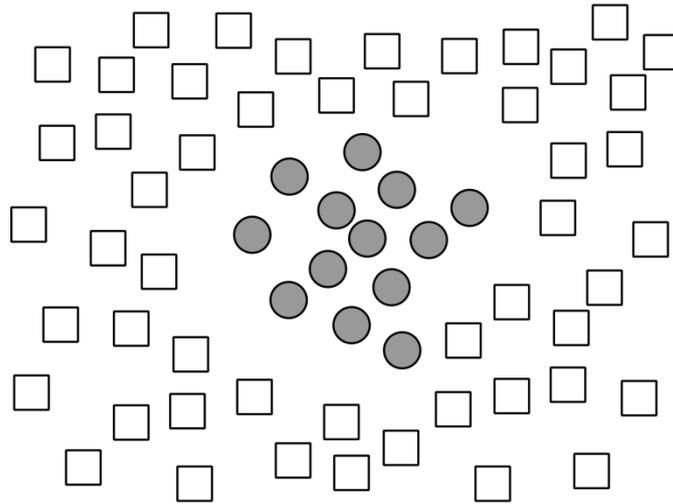


Figure 64: dataset that cannot be divided by a single hyperplane and cannot be classified with linear SVM

This type of problem can be solved using nonlinear transformation by means of Kernel functions which transform the data into a space where the classes are more distinct. This process is done by replacing the inner product in the linear hyperplane definition by a nonlinear Kernel function such as:

$$D(\mathbf{x}) = \sum_{i=1}^n \hat{\alpha}_i y_i G(\mathbf{x} \cdot \mathbf{x}') + \hat{b} . \quad (\text{eq. 30})$$

The more common Kernel functions used on SVM space are polynomial functions of order d ,

$$G(\mathbf{x}, \mathbf{x}') = [\mathbf{x} \cdot \mathbf{x}' + 1]^d \quad (\text{eq. 31})$$

and the Gaussian functions also known as radial basic functions,

$$G(\mathbf{x}, \mathbf{x}') = \exp \left\{ -\frac{|\mathbf{x} - \mathbf{x}'|^2}{\sigma^2} \right\} , \quad (\text{eq. 32})$$

where σ^2 defines the function's width. Figure 65 shows that widths greater than 48 give the best shock detection performance. For more mathematical detail on SVM and Kernel functions development refer to Kinani and Oudadess (2010), Cherkassky and Mulier (2007).

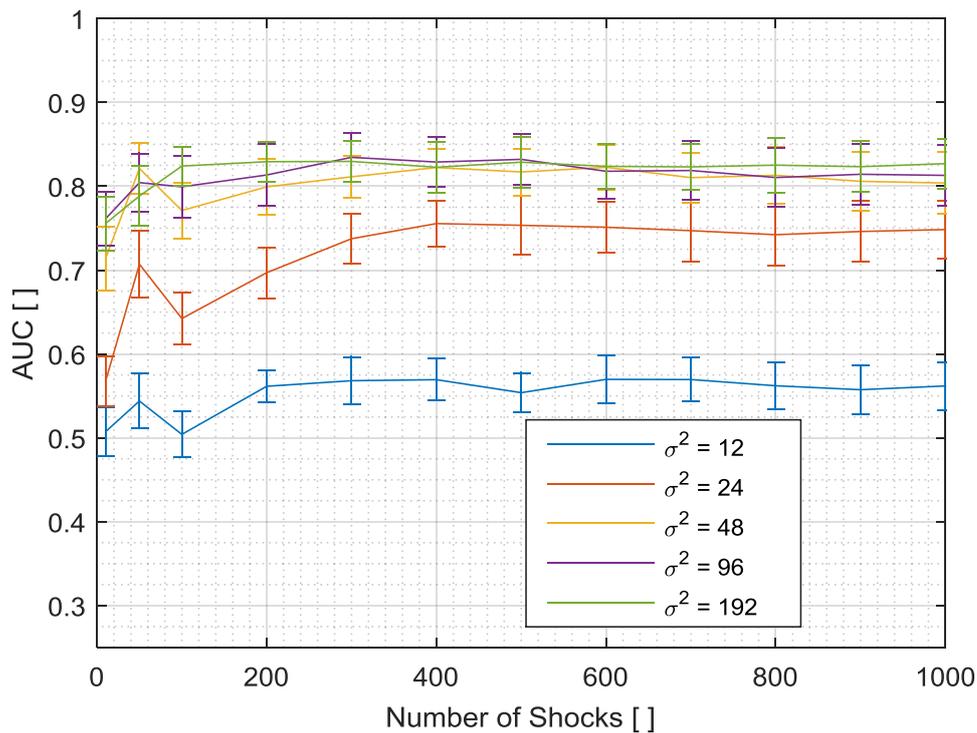


Figure 65: Gaussian function's width performance in function of the number of shock in training dataset size (10 shocks per 100 s; 95 % confidence interval)

There are no direct physical relations known between the predictors used to analyse RVV signals which means there is no *a priori* preferred Kernel transform to detect shocks. As seen in Figure 66, linear (*i.e.* no Kernel transformation) and polynomial SVMs have poor detection performance. Their AUC values highly depend on the training dataset and drop below 0.5 at several training dataset size. This suggests that these classifiers are based on chance rather than on statistics. Conversely, the Gaussian SVM has high AUC values that quickly plateau when a training dataset of more than 300 shocks (1000 s) is used.

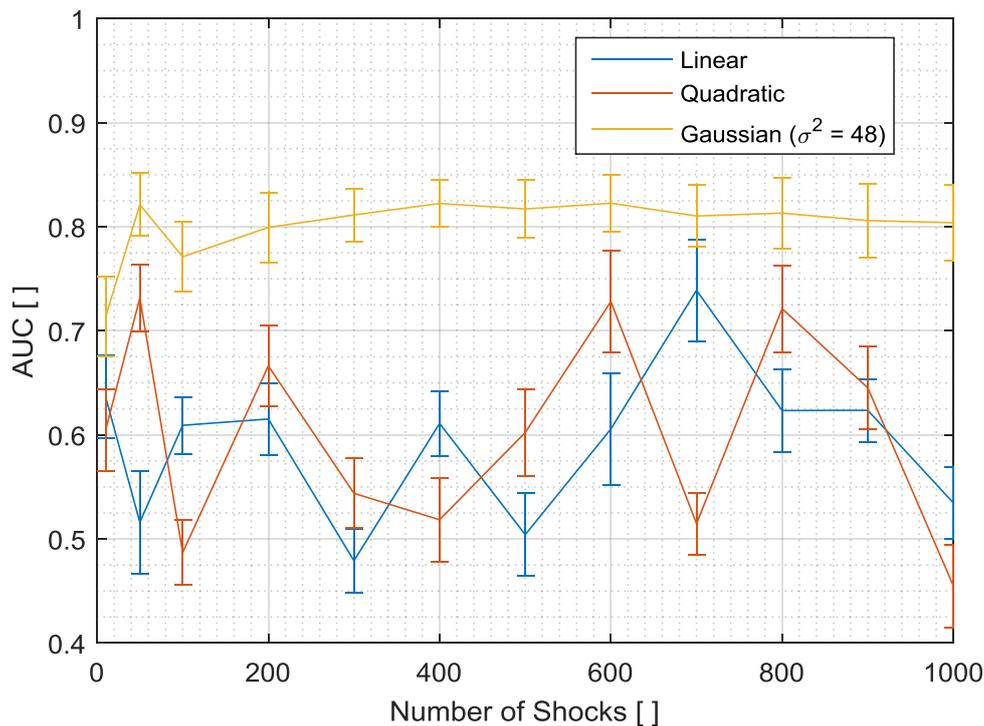


Figure 66: SVM classifiers' performance in function of the number of shock in training dataset size (10 shocks per 100 s; 95 % confidence interval)

Classification speed and memory size of SVM classifiers depend on the boundary size between both classes. When a longer training dataset is used it is more likely to increase this boundary, but there is no direct linear relation. For example the Gaussian SVM trained with a 10 shocks dataset was able to classify a 300 s signal in 1.4 s while the Gaussian SVM trained with a 300 shock dataset classified the same signal in 49 s^v. The memory requirement varies from 9 MB to 90 MB per 100 s of training dataset size, which is relatively low compared to the other classifier types.

7.5.7 Neural Networks

Neural Networks are a broadly used classifier that mimics how the brain works. This algorithm links the predictors (the inputs) to the classes (outputs) using multi-layered networks called hidden layers (Figure 67). The number of hidden layers required depends on the complexity of the classification problem. It can go from one to hundreds of layers. Neural networks are powerful algorithms with high accuracy but they require a large training dataset especially when an extensive number of hidden layers is used. This limits their usage for RVV analysis and transients detection.

^v *Ibidem* ii

According to preliminary investigation, a 1000 shocks (10,000 s) training dataset was not sufficient to train Neural Networks with 1 to 10 hidden layers. None of the trained Neural Networks could accurately detect a single transient event. More complex Neural Network could probably be accurate but that would require a much larger training datasets and perhaps more hidden layers, which could imply several weeks of computation. These extensive computation requirements are probably caused by the large number of predictors used for the RVV analysis. The other classifiers aforementioned do not have these extensive requirements and provide more accurate outcomes. Therefore, the Neural Networks are left out of this assessment.

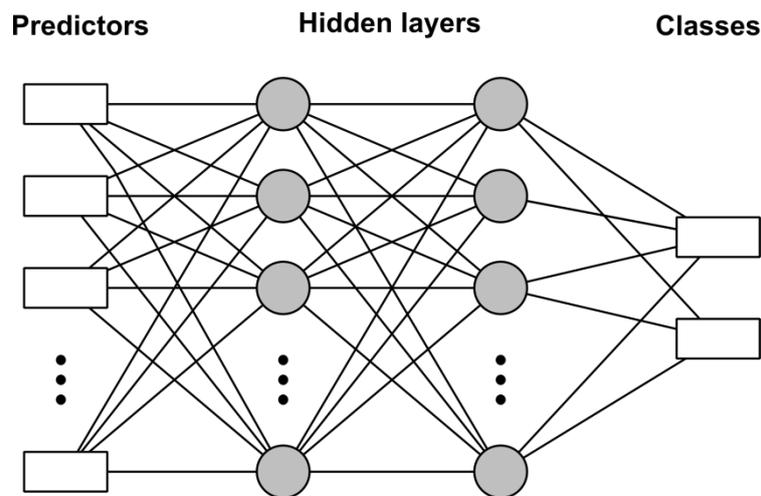


Figure 67: Neural Network with 2 hidden layers and 2 classes (outputs)

7.6 Classifiers Comparison

The analysis of each classifier group shows that Decision Tree, k NN, ensemble and SVM classifiers are suitable to detect shocks buried in RVV signals. Based on the first assessment, the best classifiers of each group are: 20 splits Decision Tree; Euclidean 100NN with no weighting; Bagged and RUSboosted ensemble algorithm; and the Gaussian SVM ($\sigma^2 = 48$). Their detection performances are summarised in Table 3.

Table 3: best performing classifiers from each suitable classifier group, prediction time based on 500 s validation signal and the memory requirement is for the optimal training dataset size

	Average AUC []	Prediction time ^{vi} [s]	Optimal training dataset size [Number of shocks]	Memory requirement [MB]
20 splits Decision Tree	0.81	0.049	200	0.02
Euclidean 100NN	0.84	243	300	8.1
Bagged Ensemble	0.85	20	300	520
RUSBoosted Ensemble	0.83	7.5	50	1.5
Gaussian SVM (48)	0.83	49	300	3

The 20 splits Decision Tree is the fastest classifier and needs the smallest amount of memory but it has one of the lowest AUC. As none of the other classifiers have very extensive prediction time and memory requirement, these advantages are not necessarily important.

The Euclidean 100NN has an above average AUC but it is the slowest classification speed. It can analyse a 500 s signal in 243 s which is about half the signal duration. This relatively slow computation speed would therefore have a significant effect on the analysis of very large signal database.

The Bagged Ensemble has the highest AUC but requires the largest memory requirement (520 MB), which is insignificant for modern computers.

The RUSBoosted Ensemble classifier requires the smallest training dataset. It has an average AUC, fast classification speed and low memory requirement. As opposed to the Bagged Ensemble, this small training dataset requirement could facilitate the RUSBoosted Ensemble classifier implementation on real RVV signals.

The Gaussian SVM classifier has versatile performances, it has: an average AUC; a prediction time five times faster than the Euclidean classifier; an optimal training dataset slightly over the average.

7.7 Conclusion on Machine Learning

Machine learning algorithms produce powerful classifiers that can detect shocks buried in RVV signals. Their main advantage compared to more classical approaches is that they use more than one predictor to support their prediction. To detect shocks, the classifiers were trained using

^{vi} *Ibidem* ii

36 predictors based on the best analysis methods introduced in the previous chapters. As a large training dataset is required to construct a machine learning algorithm, the learning and classification process can be time consuming. To reduce this time, the processing was optimised by reducing the sampling frequency of the predictor and the memory usage by using single-precision instead of double-precision numbers. This allows the algorithms to be trained in less than two hours per classifier using a dataset composed of 1000 shocks buried in 50,000 s long^{vii} signals.

A detection enhancement algorithm was also developed to improve the classifiers' detection accuracy based on the physical nature of the shocks input. This algorithm ensures that the detections last at least 1.4 s, which is the longest impulse function used in the synthetic RVV signals. Any detection shorter than 0.14 s (10% of the longest impulse) is also discarded as it is too short to represent realistic shocks. These values are based on the vehicle model presented at Chapter 4 and when applied in practice, they may be changed to match the dynamic characteristic of the vehicle in question.

Four parameters were used to assess the classifier performance: the AUC, prediction speed, optimal training dataset length and memory requirement. The AUC is calculated on the ROC curve where the sensitivity (y-axis) is defined as the ratio between the number of detection segments that overlap at least 75 % of shock duration over the number of shocks (eq. 20) and the fall-out (x-axis) is defined from the specificity (eq. 21).

Based on these parameters four out of the seven types of machine learning classification algorithms covered in this chapter were found suitable to analyse RVV signals: the Decision Tree, *k*NN, Ensemble and SVM. From those, the best classifiers were defined. The Bagged Ensemble was found to have the best average AUC (0.85) but it needs the largest training dataset (600 shocks) and memory requirement (520 MB). The 20 splits Decision Tree is by far the faster (0.049 s) but it has the lower AUC (0.81). The performances of the Euclidean 100NN, RUSBoosted and Gaussian SVM were similar and laid between the Bagged Ensemble and Decision Tree.

There is no single classifier that really stands out from this assessment. The five best performing classifiers have similar merits and limitations. These classifiers need to be compared in detail before determining which one should be used to detect shocks in RVV signals. Prior to this, each classifier needs to be optimised to ensure it is compared at its full potential.

^{vii} *Ibidem* ii

Chapter 8: ROAD VEHICLE VIBRATION CLASSIFIERS

Five classifiers were identified as well suited to detect shocks buried in RVV signals: Decision Tree, Euclidean 100NN, Bagged Ensemble, RUSBoosted Ensemble and Gaussian SVM. Preliminary assessment of their classification performances on RVV showed similar results and no algorithm clearly stood out. The classifiers have similar accuracy (AUC within 0.81 and 0.85). The prediction speed of the Decision Tree is more than 150 times faster than any other classifiers, but this does not justify its selection because computing time is not a major issue in RVV analysis and the slowest classifier (Euclidean 100NN) performed reasonably quickly for a post-processing application. There are differences within training dataset requirements but this would only be a valid selection criterion if the classifiers were trained on *in-situ* dataset (*i.e.* real RVV measurement) where defining the real position of shocks could be difficult and time consuming.

In order to select the best classifier, more specific tools need to be used to compare accuracy beyond the AUC value. Prior to this more comprehensive comparison, each classifier algorithm is optimised further using these new accuracy measurements as reference. This will ensure that more accurate RVV shock classifiers are identified.

8.1 Classifier Accuracy

The AUC value is a convenient accuracy measurand because it integrates the classifier sensitivity and specificity dependence in a single value (Bradley, 1997, Ling et al., 2003). However, this simplicity is also an important shortcoming because it could lead to false comparison (Powers, 2012). The potential risk of using AUC to assess accuracy comes from the ROC curve. This curve represents the classifier's sensitivity as a function of its fall-out (*i.e.* the true detection rate as a function of the false detection rate). The AUC represents the area under this curve, which gives its global performance but neglects any local behaviour. For instance, two classifiers can have the same AUC but clearly different ROC shapes as shown in Figure 68. Classifier A has better sensitivity at low fall-out value but reaching 100 % detection rate results to a very high fall-out. On the other hand, classifier B has a poor sensitivity at low fall-out but

surpasses classifier A's sensitivity for any fall-out value over 0.23. Classifier A is better for an application where low false detection rate is required and classifier B is better for an application where high detection rate is more important regardless of the false detection rate.

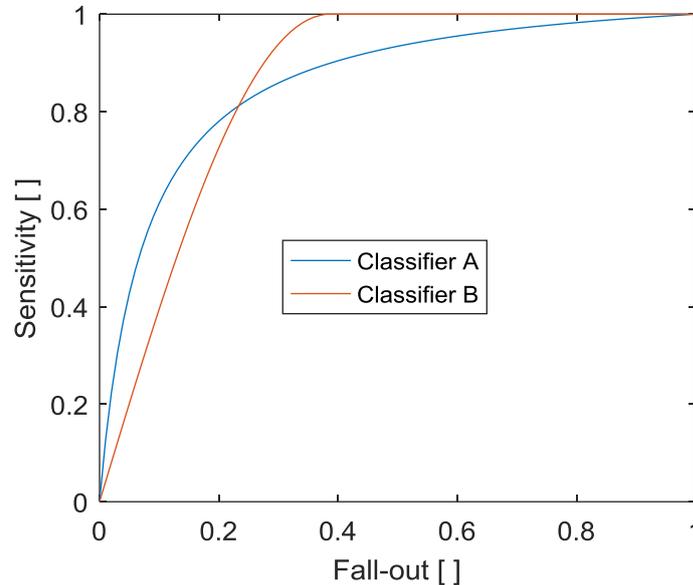


Figure 68: ROC curve of two classifiers with same AUC values (0.80)

8.1.1 Optimal Operation Point

This leads to an important concept in detection theory, the Optimal Operation Point (OOP). Classifiers give the probability that a class exists. However, even if the probability is more than 50 %, it does not necessarily mean that this class will be detected. For example, one could use a 1 % fall-out classifier to detect a very rare genetic anomaly present in one in a billion persons, but this would be a very poor classifier choice. Within ten thousand samples, about one hundred will be found positive when it is very likely that none of them is a true positive because the likelihood of this anomaly is one over one billion. One could assume that 1 % fall-out is very good, but not if it is used to detect very unlikely events. The Bayesian theory and decision cost functions are used to overcome this important shortcoming by finding the optimal trade-off between the fall-out and the sensitivity for a specific classifier and application. This trade-off is known as the OOP. There are a number of different OOPs which can be found on the ROC curve using different criteria.

8.1.1.1 Bayes' Criterion

The Bayes' criterion finds the OOP of a classifier based on the *a priori* probability and the cost of each decision. For a binary detection, a classifier bases its predictions on two hypotheses (H_0 or H_1). H_0 is called the null hypothesis or the “no shock” class for RVV application and H_1 is called the alternate hypothesis or “shock” class. The respective *a priori* probabilities of

occurrence of these hypotheses are P_0 and P_1 . The binary case has four possible decision types which have their respective cost:

1. decide H_0 when H_0 is true with a cost value of C_{00} ;
2. decide H_0 when H_1 is true with a cost value of C_{01} ;
3. decide H_1 when H_0 is true with a cost value of C_{10} ;
4. decide H_1 when H_1 is true with a cost value of C_{11} .

From the cost values and the *a priori* probability, the OOP is where the ROC curve derivative equals to:

$$\frac{dP_D}{dP_F} = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}, \quad (\text{eq. 33})$$

where the sensitivity is represented by the probability of detection, P_D , and the fall-out by the probability of false alarm, P_F . More information on the Bayes' criterion and its development is presented in Appendix D.

8.1.1.2 Minimax Criterion

There is a fundamental shortcoming with the Bayes' criterion that limits its usage on the RVV signal; the *a priori* probability of shocks is unknown. In the case of RVV, this probability cannot be estimated *a priori* because it depends on the road condition. By way of an example, some unexpected roadworks can significantly increase the likelihood of shocks on a road that should have normally been smooth. In such a case, the Minimax criterion can be used to define the OOP without using the hypothesis *a priori* probabilities. This point is where the derivative of the ROC curve equals to:

$$\frac{dP_D}{dP_F} = \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})}. \quad (\text{eq. 34})$$

More information on the Minimax criterion and its development is presented in Appendix E.

In many situations it is also difficult to evaluate the cost of missed and false detections. For instance, what would be the cost of missing a shock in RVV signal? Should this cost be higher than making a false shock detection? These questions are rhetorical because they depend on too many parameters and assumptions. One solution to this issue is to give the same cost for missed and false detections and to attribute no cost for correct detection. This approach is called the Maximum-Likelihood Estimator (MLE) where the OOP is defined as:

$$\frac{dP_D}{dP_F} = 1. \quad (\text{eq. 35})$$

8.1.1.3 Neyman-Pearson Criterion

The Neyman-Pearson criterion is another way to find the OOP without knowing the *a priori* probability and without attributing any decision cost. This criterion is fairly simple. It determines the OOP by fixing the fall-out to a certain level and using the point where the ROC curve crosses this level. Limiting the maximum level of false detection (fall-out) can be seen as fixing the significance level of the detection. For example for a 10 % significance level or fall-out, the operation point on the ROC is where the fall-out equals 0.1. The selection of the significance level depends on what is found acceptable for a given application which is often based on an arbitrary selection. A 10% value was used for shock detection purposes as it results to a reasonable amount of false detections. As shocks do not need to be defined under ‘strong evidence’, the value is twice the significance level proposed by Parsons (2007) for critical analysis.

8.1.2 OOP Selection

The MLE and the Neyman-Pearson criteria can be used to compare the OOPs of the five selected classifiers from Chapter 7. Those criteria are based on the ROC curve which depends on the size of the validation dataset. For application on stationary processes, the size of the validation dataset varies generally within 10 % and 25 % of the size of the training dataset (Rogers and Girolami, 2011, Shalev-Shwartz and Ben-David, 2014). However, much bigger validation datasets are required for nonstationary processes in order to cover the variation of the signal’s statistical moments in time. For instance, the assessment of the detection performance of shocks buried in RVV signals must consider a large variety of shock amplitudes and random vibration intensities combinations which cannot be done using a short validation dataset.

8.1.2.1 Validation Dataset Uncertainty

The relative classifiers performance assessment of Chapter 7 was performed using a validation dataset composed of 500 shocks which is more than the optimal number of shocks required in a training dataset for all selected classifiers (within 50 and 300 shocks, Table 3). A large validation dataset was used to ensure an accurate validation without *a priori* knowledge of the learning dataset size requirement. This excessively large dataset could, however, induce unnecessary and extensive computing time.

The dataset size can be optimised by computing the AUC value of the five selected classifiers for different validation dataset lengths (Figure 69). The stochasticity of these validation datasets means the AUC is a probabilistic value and has an inherent uncertainty. The AUC’s uncertainty is quantified by separating the validation datasets into five shocks segments (≈ 50 s). AUC values are computed for every segment. The overall AUC value and its confidence intervals are calculated using the n segments’ average, μ , and standard deviation, σ :

$$\text{c.i.} = \mu \pm t_{\alpha, (n-1)} \frac{\sigma}{\sqrt{n}} \quad (\text{eq. 36})$$

where $t_{\alpha, (n-1)}$ is the Student's t -distribution for a confidence level, α .

Figure 69 shows that the classifiers reach an AUC plateau when the validation dataset contains about 100 shocks which is approximately a third of the optimal training dataset used to train most classifiers. The confidence intervals show that it is possible to significantly distinguish the AUC value within some classifiers.

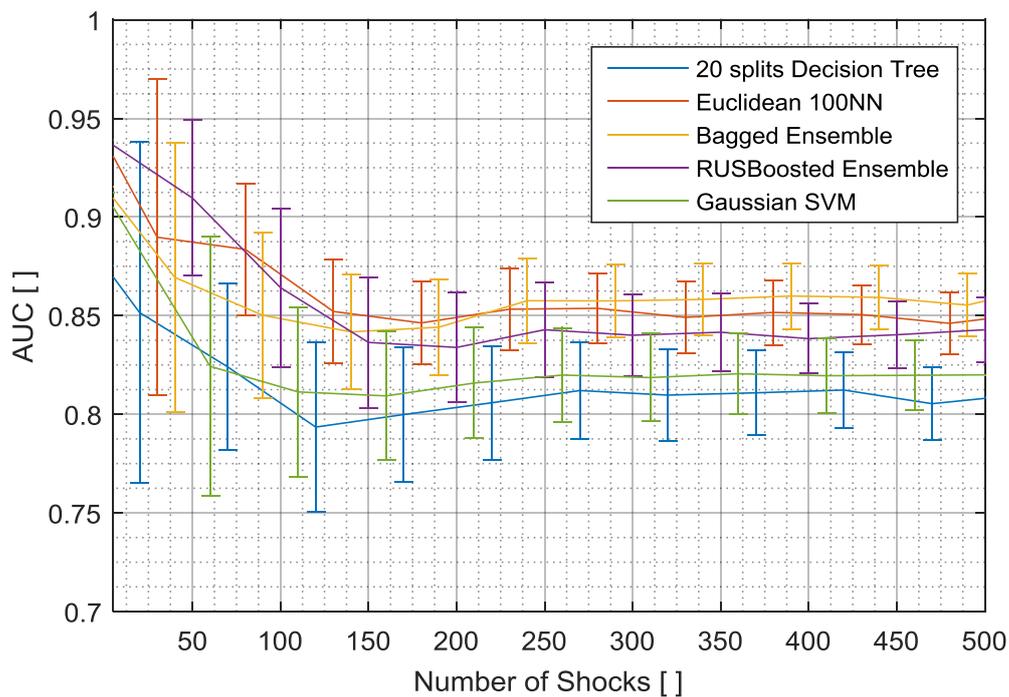


Figure 69: AUC value as a function of the number of shocks buried in the validation dataset (10 shocks per 100 s; 95 % confidence interval)

Since a longer dataset is composed of more segments of 10 shocks (100 s), their uncertainty necessarily decreases with the validation dataset's size up to an asymptotical minimum (Figure 70). The presented classifiers tend towards the same minimal confidence interval range value at 0.02 (± 0.01). This level of certainty is obtained for every classifier with a validation dataset of at least 350 shocks.

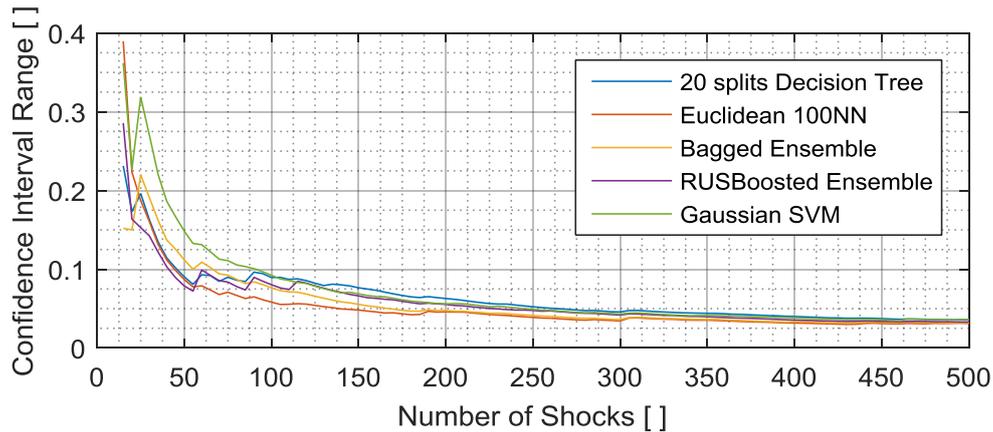


Figure 70: AUC 95 % confidence interval range as a function of the validation dataset's size

8.1.2.2 OOPs Comparison

Two OOPs were calculated for the five selected classifiers using the MLE and the Neyman-Pearson criteria on a 500 shocks dataset. These OOPs are presented in Figure 71, where the MLE OOPs are circled in red and the Neyman-Pearson OOPs are positioned where the ROC curves cross the 0.1 fall-out (dashed-line).

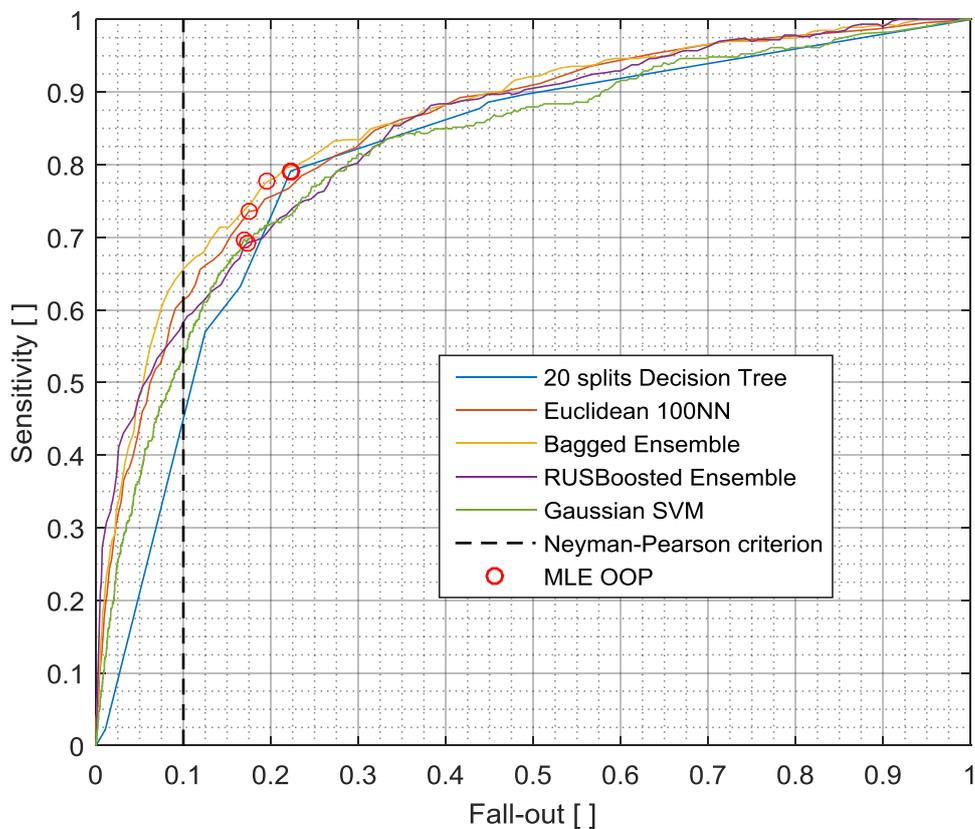


Figure 71: ROC curves and OOP of the five selected classifiers

The OOPs' coordinates (fall-out and sensitivity) and the AUC values from this validation dataset are presented in Table 4. These results lead to diverse conclusions on the classifiers' performance. The slope of the Decision Tree ROC curve increases at a slower rate than the other ROC curves which gives it the lowest AUC value (0.81); the highest MLE OOP fall-out (0.22) and the lowest Neyman-Pearson sensitivity (0.45). In the other hand, this classifier has the highest MLE OOP sensitivity (0.79).

The Euclidean 100NN, RUSBoosted Ensemble and Gaussian SVM all have the lowest MLE fall-out (0.17). Among these three, the Euclidean 100NN has the highest MLE and Neyman-Pearson sensitivities (0.74 and 0.61 respectively) and AUC value (0.85) which makes it the best classifier of the group. The Bagged Ensemble has the best Neyman-Pearson OOP with a sensitivity of 0.65. It also has the highest AUC value (0.86).

The 20 splits Decision Tree can be considered the best classifier for its optimal sensitivity, or, if a low fall-out value is required, the Bagged Ensemble would be the best classifier. At this point, there is no reason to why one classifier should be chosen over the others and more information is required to assess the classifiers prediction performance.

Table 4: OOPs comparison within the MLE, Neyman-Pearson (set at 10 % fall-out) and AUC values for the five selected classifiers using a 500 shocks validation dataset

	MLE		Neyman-Pearson	AUC
	Sensitivity	Fall-out	Sensitivity	
20 splits Decision Tree	0.79	0.22	0.45	0.81
Euclidean 100NN	0.74	0.17	0.61	0.85
Bagged Ensemble	0.78	0.20	0.65	0.86
RUSBoosted Ensemble	0.69	0.17	0.58	0.84
Gaussian SVM	0.70	0.17	0.54	0.82

8.1.3 Shock Amplitudes Distribution

The ROC curve and OOP coordinates give a good indication of the true and false detection rates of the classifiers. However, they do not provide any insight into the quality of their detections. RVV simulation accuracy depends on the characteristics of each of its modes. As explained previously, once the shocks are detected in a RVV signal, they can be extracted and characterised independently. The other modes can then be characterised in the remaining signal without their parameters being affected by the shocks.

The main parameter used to characterise shocks is the maximum absolute amplitude which defines the severity of the shocks. The classifiers' detection quality can be assessed by

comparing the maximum absolute acceleration distributions for the shocks really present in the signal (called real shocks for simplicity's sake) and the detections (true and false detections) using both the MLE and Neyman-Pearson criterion (Figure 72 left). These distributions represent the maximum absolute acceleration values of each shock superimposed on the signal (real distribution) and each shock detection (detection distribution). Comparing the real and the detected shock amplitude distributions gives more insight about the classifiers' detection quality than the sensitivity and fall-out values because it indicates which shock intensities are more likely to be misclassified. The RMS value of the signal is indicated on the figure to give a relative indication of shock intensities.

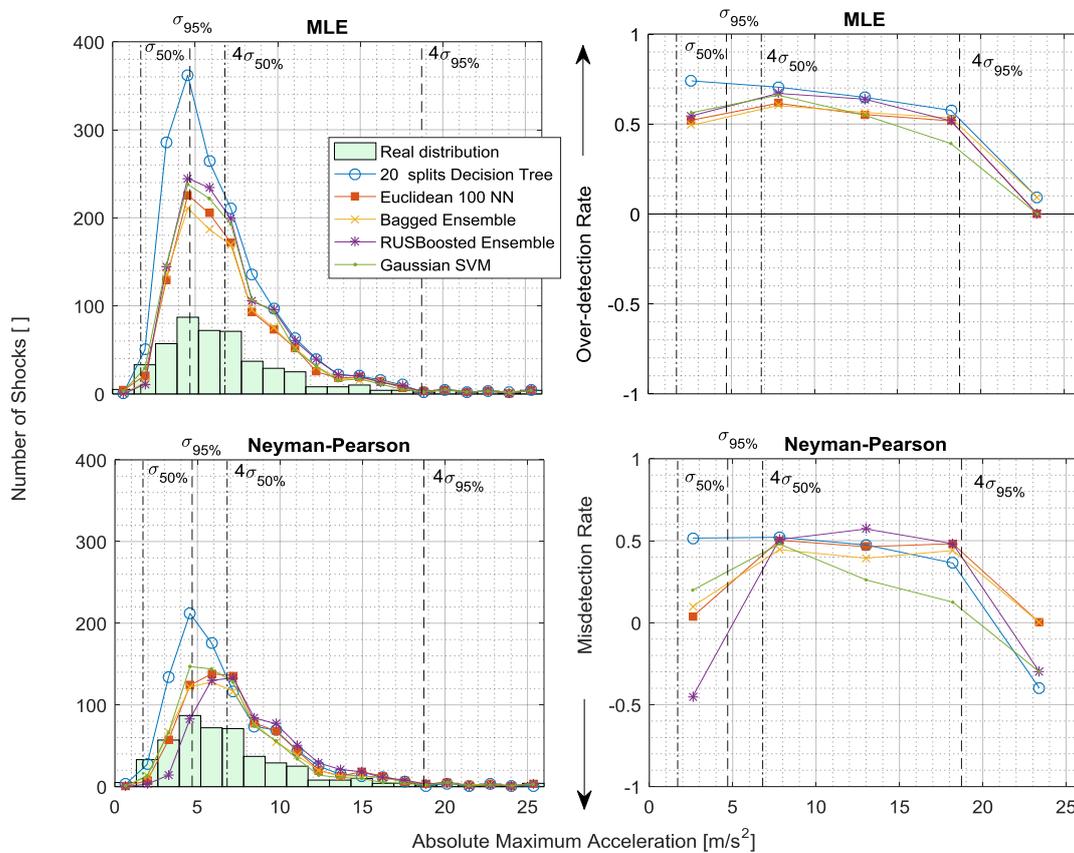


Figure 72: maximum absolute acceleration distributions for all the selected classifiers using MLE and Neyman-Pearson criteria (left) and Misdetction/Over-detection graph based on these distribution (right), $\sigma_{50\%}$, and $\sigma_{95\%}$ are, respectively, the median and 95th percentile RMS value of the stationary Gaussian random segments composing the signal

Maximum acceleration distributions give a complete representation of the detection accuracy. However, the shape of the distributions makes the analysis of the high acceleration region difficult. This is because the number of high amplitude shocks is too small in comparison to the total number of shocks and the discrepancies within the classifiers' distribution cannot be easily

seen in that region. To ease the analysis, the error between the detection and the real distributions is presented on a Misdetection/Over-detection graph (Figure 72 right). This graph averages the difference between the number of detections and the real shocks distributed into five bins (acceleration ranges). When a bin has more detections than real shocks the Over-detection rate is calculated by dividing the difference by the number of detection. On the opposite, when a bin has more shocks than detections, the Misdetection rate is calculated by dividing the difference by the number of real shocks. The lower the value is for both the Over-detection and Misdetection rate, the better the classification is.

The significance of the shock amplitude is relative to the intensity of the signal. As presented in Figure 73, shocks with maximum absolute amplitude inferior to four times the RMS value (standard deviation, σ , for zero mean signals) of the underlying random Gaussian signal are barely noticeable and have a little effect on RVV. As RVV are nonstationary signals, the intensity cannot be assessed with a single RMS value but it can be assessed with the distribution RMS value of the stationary Gaussian random segments composing the signal (in this case a Rayleigh distribution, as presented in Figure 12). So to visualise the relative importance of the shocks buried in RVV signal, the median, $\sigma_{50\%}$, and the 95th percentile, $\sigma_{95\%}$, RMS values and four times their respective values, $4\sigma_{50\%}$ and $4\sigma_{95\%}$, are presented on the shock amplitude distribution and Misdetection/Over-detection plots.

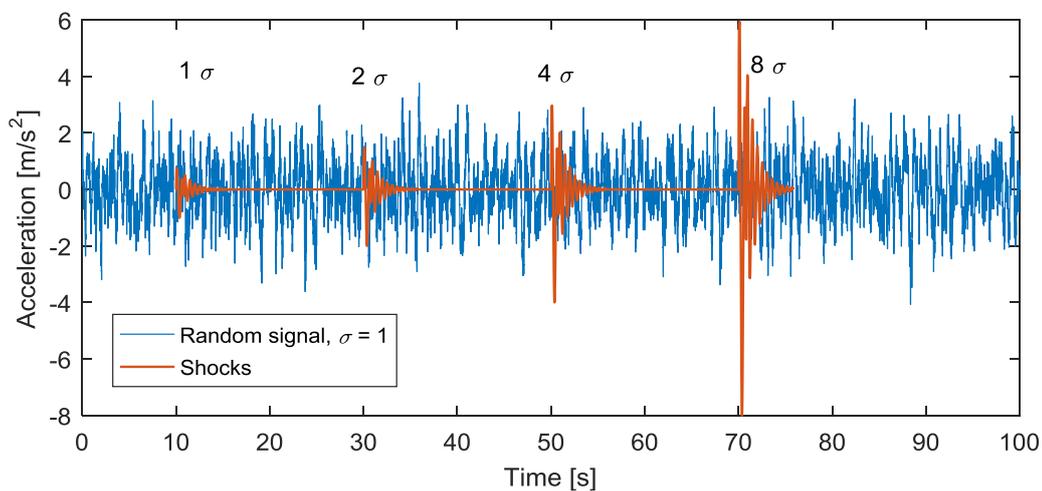


Figure 73: shock maximum absolute significance relative to the underlying random Gaussian signal RMS value (σ)

The maximum amplitude distributions of the MLE criterion have the same shape for all the classifiers, except for the Decision Tree, which has much higher values between 0 and 5 m/s^2 ($\approx \sigma_{95\%}$) (Figure 72 top). Up to 19 m/s^2 ($\approx 4\sigma_{95\%}$), almost all the over-prediction rates are above 0.5 which means that more than half of the detections below 19 m/s^2 ($\approx 4\sigma_{95\%}$) are false. Above 19 m/s^2 ($\approx 4\sigma_{95\%}$), the MLE distributions are closer to the real distribution and the over-detection

rate is within 0 and 0.1 suggesting that the detections are more accurate for high amplitude shocks.

The Neyman-Pearson criterion creates smaller discrepancies between the detections and the real distributions especially at low acceleration. Between 0 and 5 m/s² ($\approx \sigma_{95\%}$), the shape of the distributions varies significantly between classifiers. The error of the distributions goes from an over-detection rate of 0.5 (Decision Tree) to a misdetection rate of 0.5 (RUSBoosted Ensemble). Similar to the MLE criterion, the 5 m/s² to 19 m/s² ($\approx \sigma_{95\%}$ to $\approx 4\sigma_{95\%}$) region is over represented but the over-detection rates are slightly lower than the MLE criterion, especially for the Gaussian SVM classifier. There are also some misdetections above 19 m/s² ($\approx 4\sigma_{95\%}$) for the Decision Tree (over-detection rate of 0.4), RUSBoosted Ensemble and Gaussian SVM classifiers (both an over-detection rate of 0.3). The Euclidean and Bagged Ensemble have a perfect detection score in this acceleration region ($\approx \sigma_{95\%}$ to $\approx 4\sigma_{95\%}$). In the RVV analysis context, the latter misdetections predominantly affect the detection quality. This suggests that the Euclidean 100NN and Bagged Ensemble classifiers detect the more important shocks more accurately (highest maximum amplitude) when used with the Neyman-Pearson criterion.

8.1.4 Pseudo Energy Ratio/Fall-Out (PERFO) curve

Comparison of the distribution of the maximum absolute accelerations of the real shocks and the detections reveals a lot of details on the classifiers' accuracy and detection quality. However this is not a convenient approach as the distribution comparison is qualitative rather than quantitative. To improve this, the detections' maximum absolute acceleration correctness can be reduced to a single number.

Since the importance of detecting shocks increases with their amplitude, the classifiers' accuracy and quality can be assessed as the ratio between the detection and the shocks' maximum absolute acceleration pseudo energy, calculated as:

$$R = \frac{\sum[\text{Detections' Max Absolute Acceleration}]^2}{\sum[\text{Shocks' Max Absolute Acceleration}]^2}. \quad (\text{eq. 37})$$

As for sensitivity, the pseudo energy ratio depends of the classifier's detection threshold and directly affects the fall-out level. The relationship between the pseudo energy ratio and the fall-out is presented in Figure 74. In contrast to the ROC curve, the Pseudo Energy Ratio/Fall-Out (PERFO) curve does not end at the coordinate (1, 1) and is not monotonic. This is because reducing the detection threshold increases the length of signal segments considered as shocks. As the segments become longer, they may include more than one shock. However, only the absolute acceleration of each segment is considered in the PERFO calculation, such that only the shock with the highest amplitude is assessed. Therefore, the larger the detection segments

become, the more shocks that are potentially left out of this analysis which explains why the pseudo energy ratio drops after a certain point. From Figure 74, it can be seen that this drop appears for fall-out from between 0.3 and 0.5 depending on the classifier. As the classifiers operate under this fall-out level, it does not affect the current analysis.

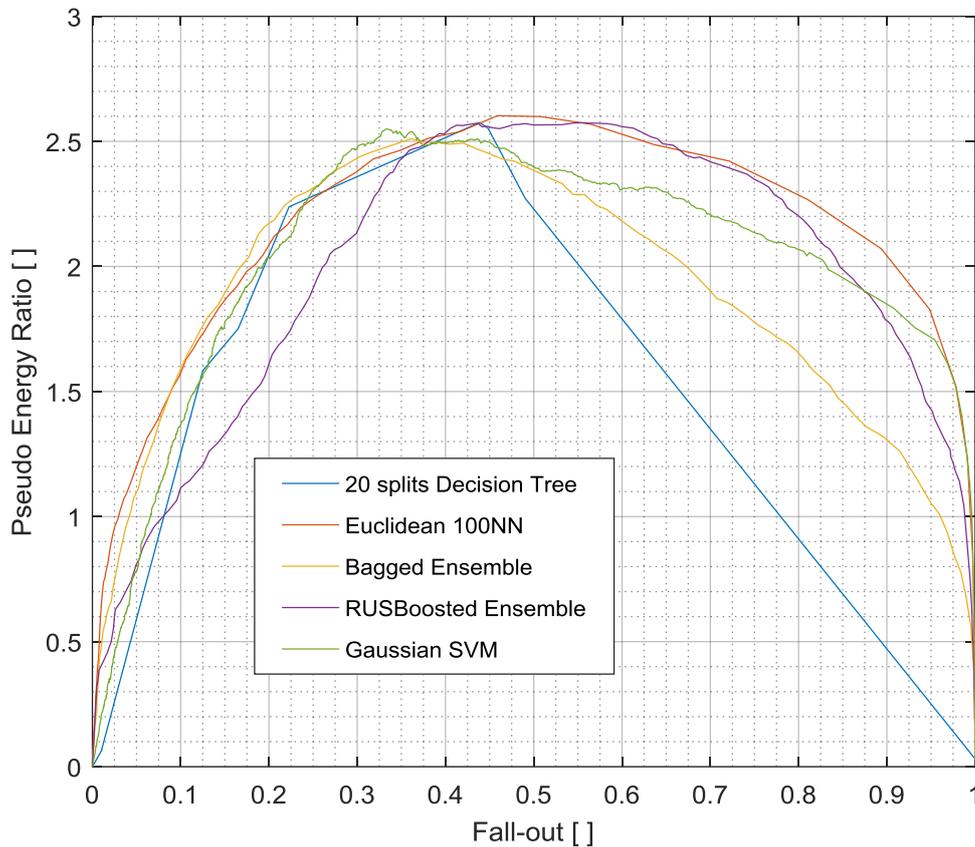


Figure 74: PERFO curve for the selected classifiers

The detection quality can be assessed by observing where the PERFO curve has a pseudo energy ratio of one which represents the operation point where the pseudo energy of the detection and the actual shocks in the signal are equal. Since the pseudo energy is calculated from the square of the maximum absolute amplitude, the lower the fall-out value is at this point, the fewer high amplitude false detections that are present in the classification. Hence, the PERFO curve measures how accurately the correct amount of pseudo shock energy is detected in a signal.

8.1.5 PERFO Criterion

As seen on the enlargement of the PERFO curve in Figure 75, there is not necessary an exact discrete point on the curve that corresponds to a pseudo energy ratio of one, because the curve is calculated from a finite number of possible classification threshold values. The PERFO curve is

therefore linearly interpolated to find the fall-out value that corresponds to a pseudo energy ratio of one. This point can then be used as a new criterion to define the classifiers' OOP which will be called the PERFO criterion. The fall-out value for each classifier at this new OOP is given in Table 5.

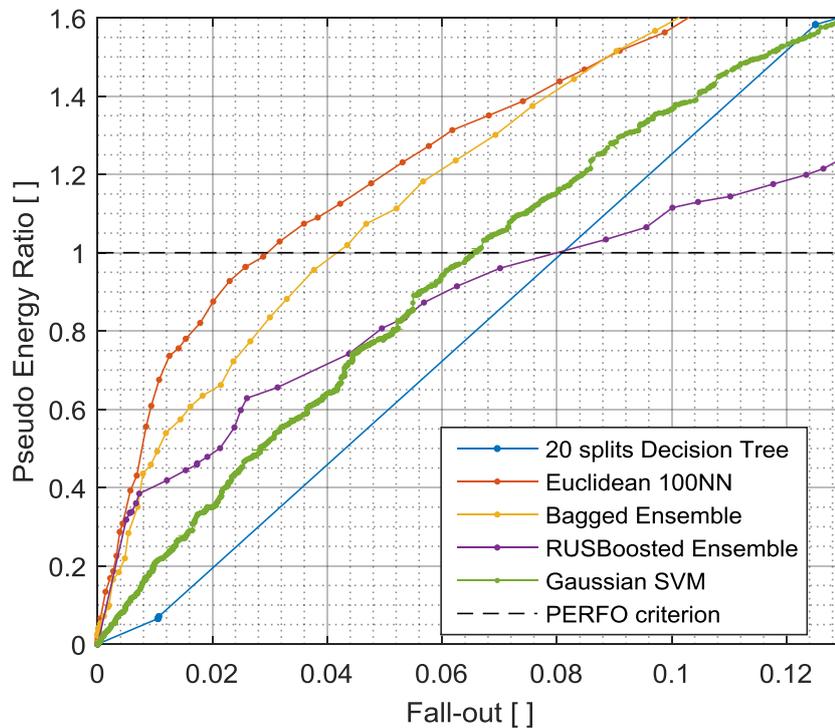


Figure 75: enlargement of PERFO curve presented at Figure 74, where the dots correspond to the discrete PERFO values

Table 5: OOPs comparison within PERFO criterion and the “classical criteria” presented in Table 4 (*i.e.* the MLE and Neyman-Pearson) for the selected classifiers using a 500 shocks validation dataset

	PERFO	MLE	Neyman-Pearson	
	Fall-out	Sensitivity	Fall-out	
			Sensitivity	
20 splits Decision Tree	0.08	0.79	0.22	0.45
Euclidean 100NN	0.03	0.74	0.17	0.61
Bagged Ensemble	0.04	0.78	0.20	0.65
RUSBoosted Ensemble	0.08	0.69	0.17	0.58
Gaussian SVM	0.07	0.70	0.17	0.54

The PERFO criterion effectively defines OOPs for shocks detection application compared to the classical criteria (*i.e.* MLE and Neyman-Pearson). The maximum absolute acceleration

amplitudes' distributions of the detections using the PERFO criterion are close to the real shocks (Figure 76). For most shock amplitudes, there are small discrepancies within the distributions at high amplitude compared with the other criteria.

The 20 splits Decision Tree and the RUSBoosted ensemble have the highest PERFO criterion fall-out values (0.08). It has a maximum misdetection rate of 0.15 for shocks below 8 m/s^2 ($\approx \sigma_{50\%}$) and a maximum over-detection rate of 0.37 within 8 m/s^2 and 19 m/s^2 ($\approx \sigma_{50\%}$ and $\approx 4\sigma_{95\%}$). However, it missed 40 % of the shocks above 19 m/s^2 ($4\sigma_{95\%}$), which is where the most important shocks are.

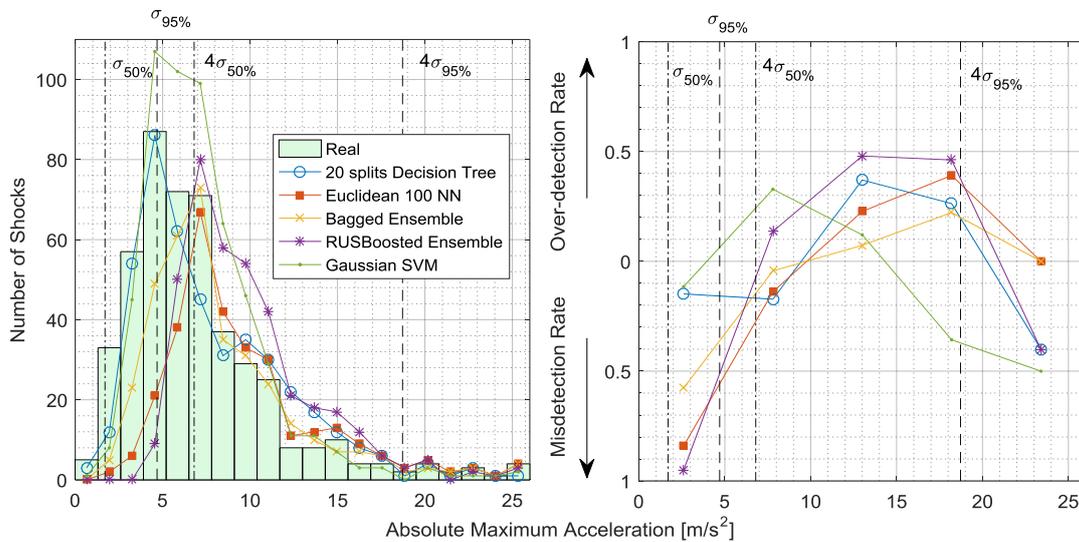


Figure 76: maximum absolute acceleration distributions for all the selected classifiers using the PERFO criterion (left) and Misdetction/Over-detection graph based on these distributions (right), $\sigma_{50\%}$, and $\sigma_{95\%}$ are, respectively, the median and 95th percentile RMS value of the stationary Gaussian random segments composing the signal

The Euclidean 100NN is the classifier with the lowest PERFO criterion fall-out and its low fall-out value (0.03) suggests it also has an accurate maximum amplitude distribution, especially above 5 m/s^2 ($\approx \sigma_{95\%}$). It has a high misdetection rate (0.84) below 5 m/s^2 ($\approx \sigma_{95\%}$), but this zone is the least important for RVV shocks analysis. Where it is most important (*i.e.* above 19 m/s^2 , $\approx \sigma_{95\%}$), over-detection rate is null.

The Bagged Ensemble has a better Misdetction/Over-detection curve than the Euclidean 100NN even if its PERFO criterion fall-out is slightly higher (0.04). Misdetection and over-detection rates are lower than the Euclidean 100NN at every point. Its misdetection rate below 5 m/s^2 ($\approx \sigma_{95\%}$) is 0.58 and its maximum over-detection rate is 0.22 within 13 m/s^2 and 19 m/s^2 ($\approx 3\sigma_{95\%}$ and $\approx 4\sigma_{95\%}$).

The RUSBoosted Ensemble has the worst performance for almost all acceleration bands. It has a misdetection rate of 0.95 below 5 m/s^2 ($\approx \sigma_{95\%}$) and then reaches a maximum over-detection rate of 0.48 within 8 m/s^2 and 13 m/s^2 ($\approx 4\sigma_{50\%}$ and $\approx 3\sigma_{95\%}$). It also missed 40 % of the shocks above 19 m/s^2 ($\approx 4\sigma_{95\%}$).

The Gaussian SVM has a different pattern than the other classifiers. It mostly over-detects shocks below 13 m/s^2 ($\approx 3\sigma_{95\%}$), and then misdetects them above this level. Its maximum misdetection rate (0.5) occurs above 19 m/s^2 ($\approx 4\sigma_{95\%}$) which is a major shortcoming for the RVV application.

8.1.5.1 PERFO Uncertainty

As for the ROC curve, the variance of a PERFO curve depends on the length of the validation dataset used for its computation. The longer the validation dataset is, the more shocks it includes, which means the classifier decisions based on chance are averaged out and the PERFO curve is closer to its real values. This implies that the validation dataset size also affects the PERFO OOP and the fall-out value associated with it. The PERFO criterion fall-out value as a function of the number of shocks buried in the validation dataset is shown in Figure 77. As for the AUC, the confidence intervals are calculated using (eq. 36) and dataset segments of five shocks. These confidence intervals are relatively small and allow the discrimination of most classifiers' fall-out value. Only the RUSBoosted Ensemble can be confounded with the 20 splits Decision Tree and the Gaussian SVM.

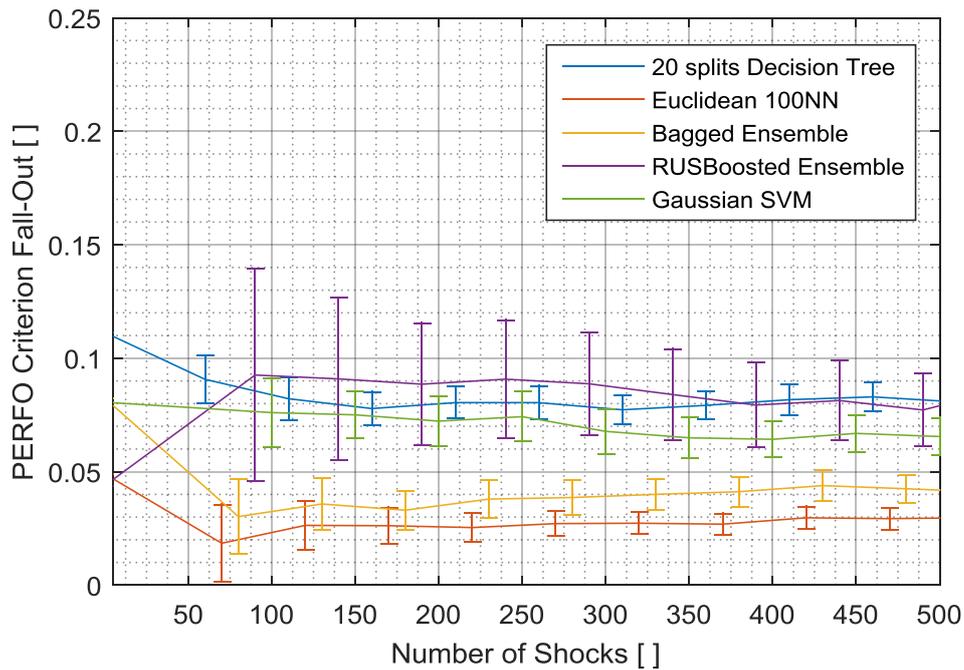


Figure 77: PERFO criterion fall-out value as a function of the number of shocks buried in the validation dataset (10 shocks per 100 s, 95 % confidence interval)

As expected, the uncertainty decreases as a function of the number of shocks included in the validation dataset. This relationship is shown at Figure 78. Above 350 shocks, the reduction in the confidence interval is very minimal excepted for the RUSBoosted Ensemble, which benefits from longer validation dataset. However, even with this extra precision, this classifier has the highest uncertainty. It appears that 350 shocks is an adequate validation dataset size as the difference between the other classifier at this point is more settled which suggests that these classifiers reach a 'steady state'.

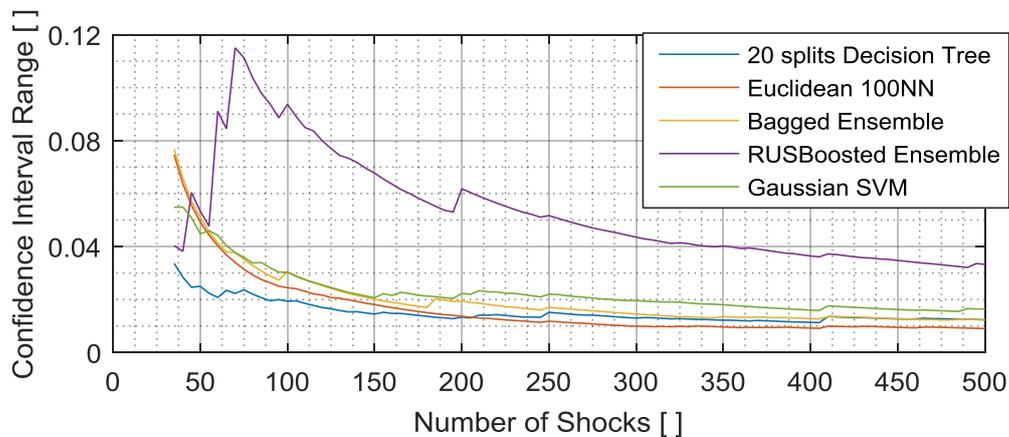


Figure 78: PERFO criterion fall-out 95 % confidence interval range as a function of the validation dataset size (10 shocks per 100 s)

8.1.5.2 Threshold Uncertainty

The PERFO criterion calibrates the classifiers' OOP and provides a detection threshold that gives the same pseudo energy level as those of the real shocks buried in the signal. By definition, this calibration is made by finding the detection threshold that corresponds to the pseudo energy ratio of one on the PERFO curve. As the PERFO curve is based on a stochastic dataset, it is important to analyse the effect of the PERFO criterion variability on the detection threshold. But first, the different definitions of the threshold have to be addressed.

The Decision Tree, *k*NN and Bagged Ensemble classifiers use the posteriori probability to make their predictions. This value goes from zero to one and, for perfectly calibrated classifiers, a class should be predicted when its posteriori probability is above 0.5. However, this is generally not the case and the decision threshold has to be calibrated to fit the application.

The RUSBoosted Ensemble classifier is slightly different because its predictions are based on a weighted vote on the hypothesis tests made from its ensemble of decision trees (Seiffert et al., 2008). Hence, its vote values are not constrained to values between 0 and 1 and can be within 0 and ∞ . For the RUSBoosted Ensemble used in this chapter the vote values vary from 1.8 to 3.8. The higher the vote value is, more likely the class exists. As there is no direct probability attached to the vote value, this type of classifier does not have a default threshold and has to be calibrated.

The SVM classifiers base their prediction on the distance from the Support Vector, which is scaled in such a way that there are only first class samples at a distance greater than 1 and only second class samples at a distance less than -1 . The zone within 1 and -1 is composed of both classes. For a perfectly calibrated classifier, the boundary line (Support Vector) defines where both classes have the same likelihood. The threshold calibration defines the distance from the Support Vector that should be used to make the classification decisions. This threshold value is theoretically within -1 and 1.

In order to compare the variability of these three different types of decision threshold, their uncertainty is normalised by their range R such as:

$$\bar{U} = t_{\alpha,(n-1)} \frac{\sigma}{R\sqrt{n}} \quad (\text{eq. 38})$$

where $t_{\alpha,(n-1)}$ is the Student's *t*-distribution for a confidence interval, α ; σ is the standard deviation and; n is the number of 10 shocks segments in the validation dataset.

This uncertainty decreases and becomes an asymptotical value as the validation dataset's size increase (Figure 79). The 20 splits Decision Tree reaches this asymptote within less than

150 shocks. The other classifiers need a longer validation dataset to reach it (around 400 to 450 shocks) which is more than the confidence interval plateau of the PERFO criterion fall-out. This means that a very long validation dataset is necessary to obtain the best estimate of the real optimal detection threshold. This could be caused by the linear interpolation in the OOP calculation.

The 20 splits Decision Tree, Bagged Ensemble and RUSBoosted Ensemble classifiers have a very small uncertainty on the threshold value which is less than 2.5 % of their range. The Gaussian SVM and Euclidean 100NN have slightly more uncertainty with respectively 5.7 % and 6.7 % of their range. All the classifiers have less than 10 % uncertainty of their range when using a dataset of at least 150 shocks.

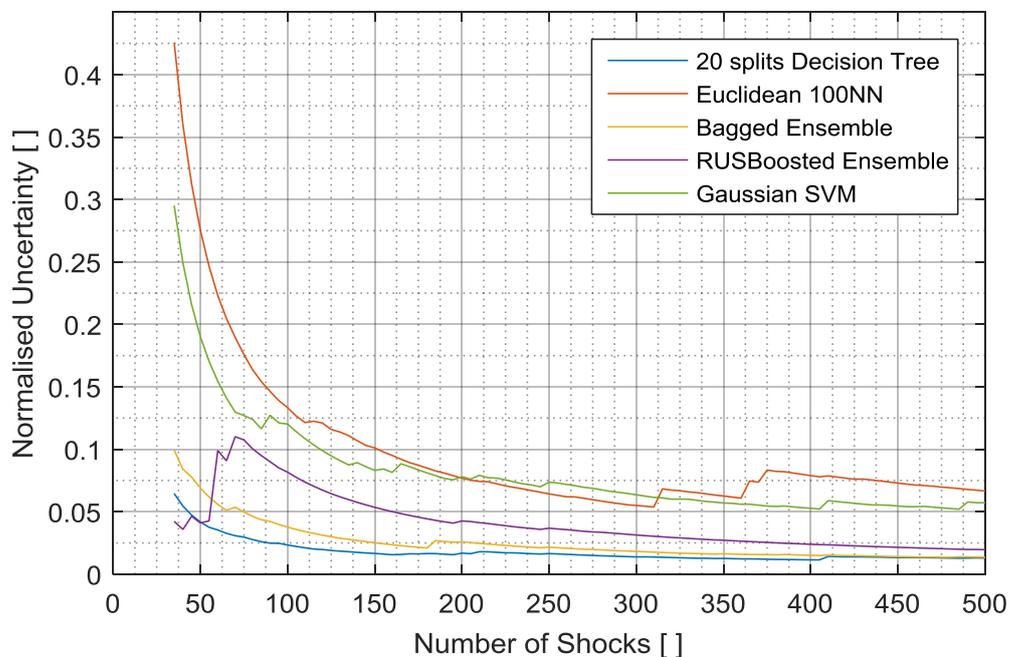


Figure 79: normalised uncertainty on the detection threshold defined with the PERFO as a function of the validation dataset size (10 shocks per 100 s, 95 % confidence interval)

8.2 Classifier Optimisation

The PERFO criterion fall-out can be used to optimise the classifiers because it is a single value which accurately assesses shock detections performance in RVV signals. Each classifier has different parameters which can be adjusted to improve detections. These parameters were reviewed in Chapter 7 and preliminary adjustments were made to assess the potential of each classifier. These adjustments are refined and optimised again using PERFO criterion fall-out to produce the best detection performances. The training and validation datasets used for the optimisation have 300 and 350 shocks respectively (10 shocks per 100 s). These datasets

provide a good trade-off between accuracy and processing time, which is essential in this iterative optimisation process. It is assumed that the optimised parameters should not affect the PERFO criterion uncertainty for each classifier. The fall-out values are therefore presented with the 350 shocks confidence interval range of their respective classifier type (Figure 78).

8.2.1 Decision Tree

Decision Tree algorithms are controlled by one main parameter: the number of splits or decisions. Figure 80 shows the PERFO criterion fall-out for Decision Trees with up to 82 splits (maximum number of splits created by the algorithm for this learning process). The fall-out value rapidly drops at three splits to a fall-out of 0.07. This means that based on only the first three decision tests (splits), the equivalent of pseudo shocks energy can be detected with a fall-out of only 7 %.

Adding more splits only slightly refines the detection accuracy. The minimum fall-out (0.05) is obtained at 25 splits, but considering the uncertainty, any number of splits between 22 and 29 can be considered optimal. Above 29 splits, the Decision Tree's performance worsens and the fall-out significantly increases to up to 0.22. This suggests that the 30th and above splits of the tree overfit the training data and deteriorate the classifier's predictions.

Interestingly, the PERFO criterion fall-out has far less variability than the AUC. As presented in section 7.5.3, the uncertainty associate with AUC values of the 20 and 80 splits Decision Trees is so important that cannot their AUC cannot be distinguished (Figure 59). The difference between AUC is smaller than their confidence interval range which makes this difference insignificant. The PERFO criterion gives more discriminating power as the difference between the 20 and 80 splits Decision Trees' PERFO fall-out is more than 15 times bigger than the confidence interval bound. This means the number of splits affects high amplitude detections the most as it has a greater effect on the PERFO curve compared to the ROC curve.

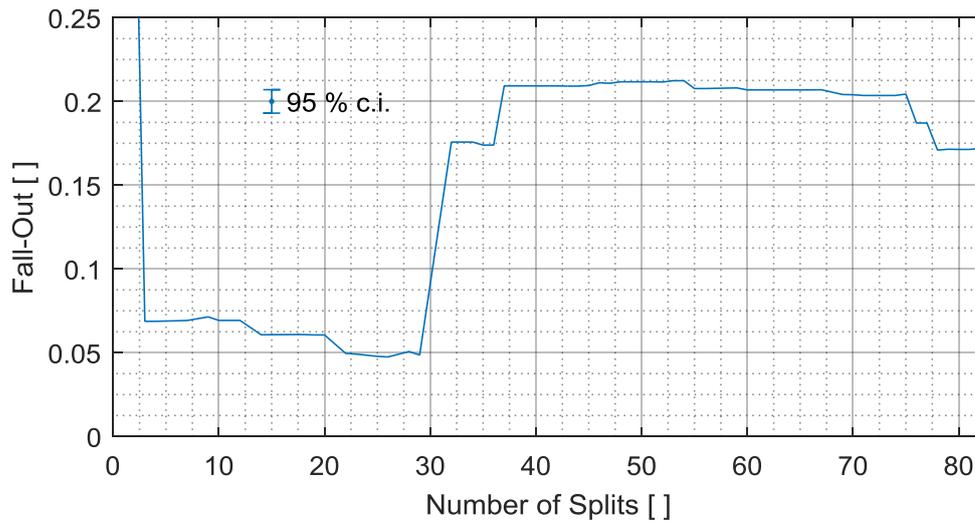


Figure 80: fall-out at an energy ratio of one as a function of the number of Decision Tree splits

8.2.2 k -Nearest Neighbours

There are two parameters that define the k NN classifiers: the number of nearest neighbours, k , and the type of distance used to compare the new samples. As seen in Figure 81, both the Euclidean and Cosine distance have a similar fall-out value as a function of the number of nearest neighbours which corroborate the conclusion of section 7.5.4 where the AUC values as a function of the training dataset's size of both distances followed the same trend. Therefore, there are no benefits in using the cosine distance instead of the more common and simpler straight-line defined by the Euclidean distance.

The fall-out value quickly stabilises at 0.04 above 20 nearest neighbours which suggests that within the uncertainty any k NN classifiers greater than 20 provide maximal detection accuracy. The 100NN used in the previous section was therefore already optimal.

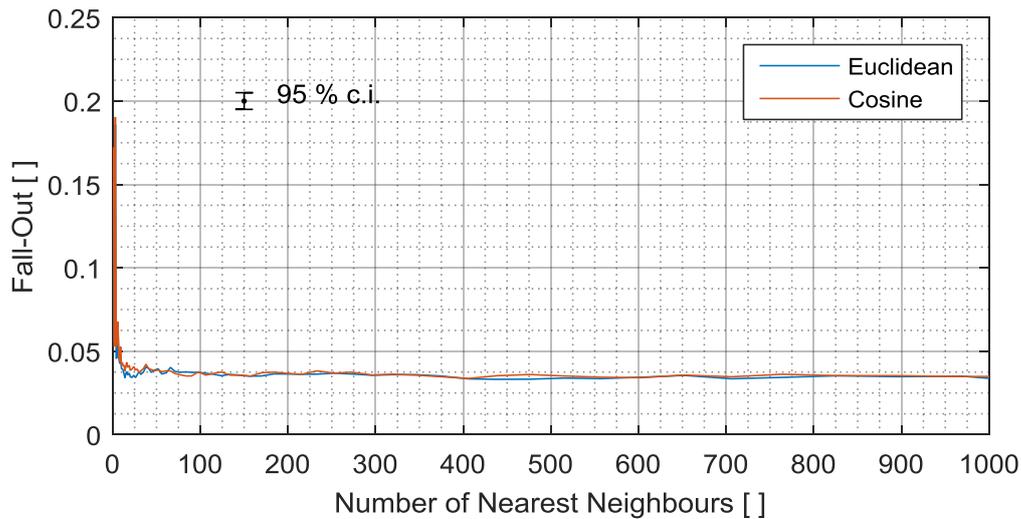


Figure 81: fall-out at an energy ratio of one as a function of the number of nearest neighbours for the Euclidean and Cosine distances

8.2.3 Ensemble

The Ensemble algorithms use several classifiers trained from different segments of the signal to detect shocks. The classification accuracy increases as a function of the number of classifiers. The optimal number of classifiers is obtained at the point where increasing the ensemble does not improve the performance.

For shock detection applications, the optimal number of classifiers is relatively small (Figure 82). The Bagged Ensemble slightly improves by using 10 to 150 classifiers. Its PERFO criterion fall-out value decreases from 0.05 to 0.04. Afterwards, the fall-out is almost unaffected by the number of classifiers. The Bagged Ensemble is composed of Decision Trees, which have a minimal fall-out of 0.05 when used in its direct form. So the combination of the predictions of more than 150 Decision Trees improves the prediction of a single Decision Tree.

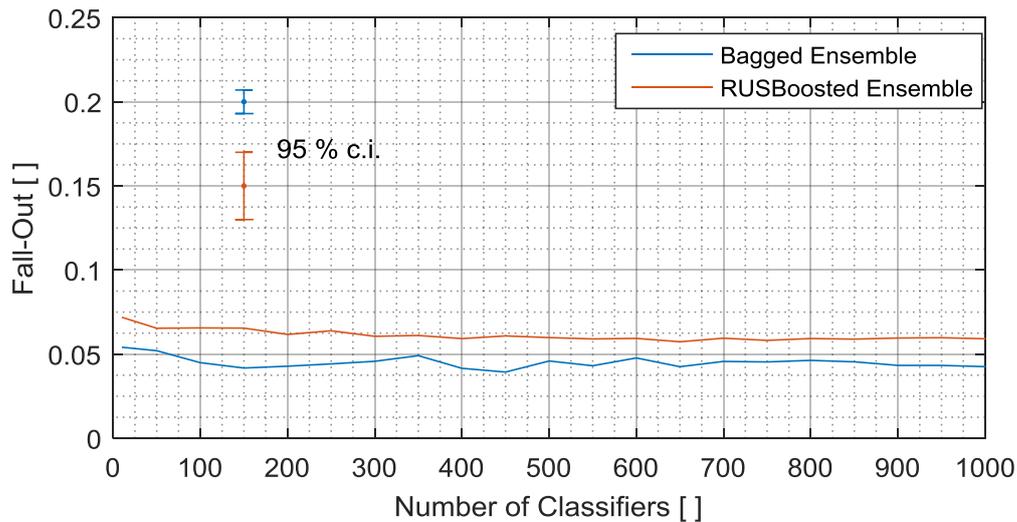


Figure 82: fall-out at an energy ratio of one as a function of the number of ensembles used with the Bagged and RUSBoosted Ensemble classifiers

The effect of the number of classifiers used in RUSBoosted Ensemble is insignificant as it is smaller than the confidence interval of its PERFO criterion fall-out (Figure 82). This means that increasing the RUSBoosted Ensemble's size does not significantly improve its shock detection performance. Being also composed of Decision Trees, this ensemble algorithm even exacerbates the detection of its subcomponent with an average fall-out of 0.06 considering that the Decision Tree has a minimum of 0.05. The RUSBoosted Ensemble should therefore not be considered as a viable RVV signals classifier.

8.2.4 SVM

As shown in section 7.5.6, from the Kernel tested functions, only the Gaussian functions provide accurate SVM classifiers. These functions are controlled by one parameter known as the Gaussian coefficient, or function's width. The preliminary optimisation in this section shows that the AUC value tends to a maximum at a Gaussian coefficient above 48.

However, as seen in Figure 83, the PERFO criterion fall-out shows a different trend. The fall-out rapidly decreases at a Gaussian coefficient above 25. From 100, the fall-out stabilises around 0.03 up to a Gaussian coefficient of 1150, where it starts to increase.

The discrepancy between the AUC and PERFO criterion optimisations can be explained by better detections of high amplitude shocks with a Gaussian coefficient within 48 and 100 which has a greater impact on the PERFO curve than the ROC curve.

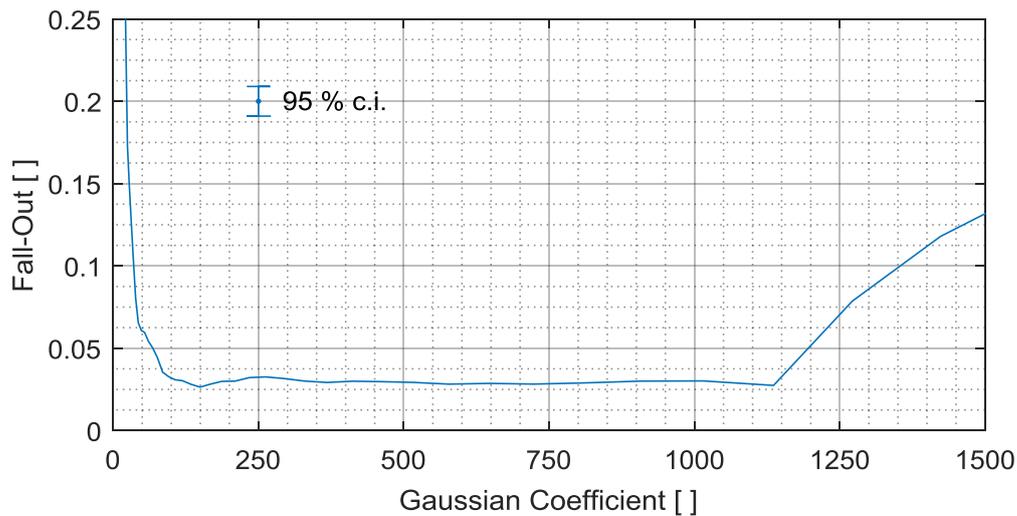


Figure 83: fall-out at an energy ratio of one in function of SVM Gaussian coefficient

8.2.5 Optimal Classifiers

A summary of optimised parameters of the best performing classifiers is presented in Table 6. There is an improvement in the PERFO criterion fall-out for the Decision Tree and the Gaussian SVM. The optimised Euclidean k NN's fall-out value is slightly higher than the value presented in Table 4; this is caused by the calculation's uncertainty. Considering this uncertainty, there are now almost no significant differences within the optimised classifiers' fall-out value.

Table 6: best performing classifiers after optimisation

Classifier	Optimisation		PERFO criterion
	Parameter	Value	Fall-out
Decision Tree	Number of splits	22-29	0.05
Euclidean 20NN	Number of NN	> 20	0.04
Bagged Ensemble	Number of classifiers	> 150	0.04
Gaussian SVM	Gaussian coefficient	100-1150	0.03

Optimised classifiers can be compared using their maximum absolute acceleration distribution and Misdetection/Over-detection graph (Figure 84). For the purpose of this comparison, the minimum optimisation parameter was used, which in most cases gives the simplest classifier. The optimised classifiers have distributions that are different to those of the classifiers evaluated before. The detections are much more precise at high acceleration shock amplitude. Above 19 m/s^2 ($\approx 4\sigma_{95\%}$), all the classifiers have a maximum over-detection rate below 0.37 and the maximum misdetection rate below 0.25. The drawback is that they missed most of the shocks

below 7 m/s^2 ($\approx 4\sigma_{50\%}$) where the misdetection rates are within 0.67 and 0.93. This shows that the optimisation is very specific to the shocks detection application.

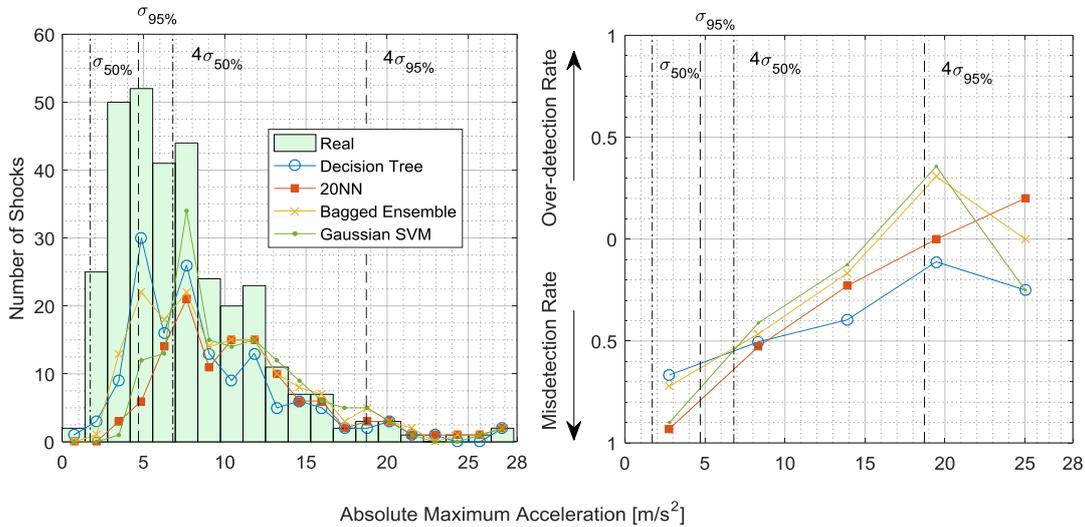


Figure 84: maximum absolute acceleration distributions for the optimised classifiers using the minimum optimisation parameters (left) and Mis-detection/Over-detection graph based on these distributions (right)

It is, once again, difficult to compare the performance of the optimised classifiers because their performances are similar. The Decision Tree misdetects shocks in every acceleration bandwidth, which suggests it did not obtain a pseudo energy ratio of one. This can be explained by the interpolation in the PERFO criterion process. Because the Decision Tree classifier uses a limited number of threshold values, there is not necessarily an actual threshold value that gives a pseudo energy ratio of one (Figure 75). Therefore, the threshold value interpolated from the PERFO criterion does not provide the correct energy ratio. The other classifiers do not have this issue because they have a sufficient number of threshold values.

The 20NN classifier only over-detected (ratio of 0.2) the shocks above 25 m/s^2 ($\approx 5\sigma_{95\%}$), which means it would provide the most severe shock representation. The Bagged Ensemble provides the least severe detections, with a maximum over-detection rate of 0.3 m/s^2 at 19 m/s^2 ($\approx 4\sigma_{95\%}$), and a perfect score above 25 m/s^2 ($\approx 5\sigma_{95\%}$). The Gaussian SVM has a worse Mis-detection/Over-detection curve than the Bagged classifier at every bandwidth, so even if its PERFO criterion fall-out value is inferior to any other classifiers, its performance based on the shock distribution suggests that it is not the best classifier for shock detection.

8.3 Conclusion on Classifiers

This chapter presented the different criteria used to find the OOP of classifiers, which give the optimal detection sensitivity for a certain application. According to the maximum absolute

acceleration distribution of the detections, the conventional criteria, *i.e.* the MLE and Neyman-Pearson criteria, produced classifiers that over-detect shocks, especially at low acceleration.

Considering that the high acceleration shocks are more important to detect for a RVV analysis purpose, a new criterion was developed based on the pseudo energy of the detections and the real shocks. This criterion called PERFO gives more accurate detections than the conventional criteria. The maximum absolute acceleration distributions of the detections made with the PERFO criterion follow more closely the real shock distribution.

Using the PERFO criterion, all the classifiers were optimised. It was found that the RUSBoosted Ensemble classifier does not improve the performances of its fundamental classifier (the Decision Tree). Having no improvement, the RUSBoosted Ensemble is therefore not considered as a potential shock classifier. From the remaining classifiers, the 20NN and Bagged Ensemble have the best detection performances.

The optimised classifiers can detect relatively high amplitude shocks ($> 4\sigma_{50\%}$) buried in synthetic RVV signals with a high degree of accuracy and it is now important to evaluate how their detections perform on real RVV signals. An assessment of the classifier using acceleration signal measured on a vehicle travelling on different road will give a complete appreciation performance of machine algorithms at detecting shocks.

Chapter 9: MODEL VALIDATION

The previous chapters shown how machine learning classifiers can accurately detect shocks buried in synthetic Road Vehicle Vibration (RVV) signals. Using the pseudo energy ratio between the shocks and prediction (PERFO criterion), classifiers can be optimised to precisely detect the high amplitude shocks with low fall-out values. It is now important to assess how these classifiers perform on real RVV measurements.

The minimum number of shocks required to perform supervised machine learning is substantial. As shown in Chapters 7 and 8, 300 shocks are needed for the classifier's training plus a further 350 shocks for the validation and calibration to be optimal. Therefore, to teach classifiers uniquely from real RVV signals, measurement must be performed on a vehicle passing over 650 well identified shocks. Classifiers are specifically trained for a certain vehicle payloads. Therefore for real-live application, it is unrealistic to retrain the classifier with such cumbersome training datasets every time the vehicle payload changes. Long measurements can be made on road vehicles. However, it is much more complicated to survey long sections of roads in order to accurately identify the position of the shocks.

To overcome this limitation, a learning method integrating both real and synthetic signals to train the classifier is postulated (Figure 85). This method requires two datasets of RVV measurements to be made on the same vehicle which are: (1) the training measurement dataset where the position of the shocks is not identified and (2) the validation measurement dataset where the position of the shocks is identified. The training measurement dataset has a long duration (3500 s) and is performed on open roads representing normal travelling routes. This measurement dataset is used to characterise the vehicle's dynamics from which a synthetic signal is generated. The RVV synthetic signal is nonstationary and contains a sufficient number of shocks needed for optimal classifier training. This synthetic signal provides a training dataset specific to the vehicle payload. Once the classifier training is completed, the classifiers are validated using the validation measurement datasets. These measurement datasets are shorter than the training measurement dataset and are performed on road circuits where the position of

the shocks is identified. This validation is essential to calibrate the classifier's threshold and to define the validity of the classification. The calibrated classifiers can then be used to identify shocks in other measurements made with the vehicle.

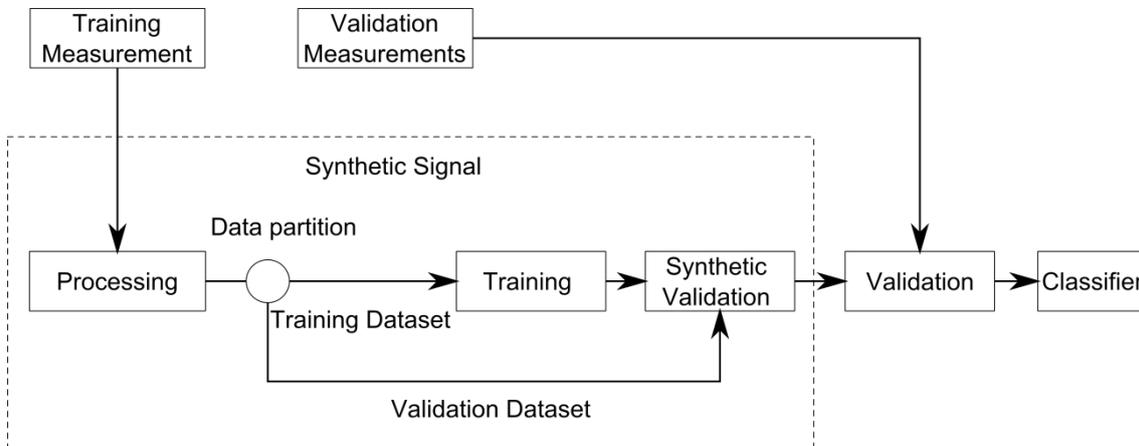


Figure 85: learning flow chart to detect shocks buried in real RVV signal

9.1 RVV Measurement

All the measurement datasets required to train and validate shock classifiers were performed with the same small utility vehicle, Mitsubishi Triton (UTE). The vehicle has a kerb mass of 1555 kg and can transport a maximum payload of 1165 kg. During the measurement, the only payload was the weight of two persons sitting in the cabin (≈ 150 kg). The vehicle has a front suspension composed of coil springs with telescopic dampers and a rear suspension composed of leaf springs and telescopic dampers. Its wheelbase and track width are 3 m and 1.5 m respectively.

A Slam Stick X accelerometer (Midé Technology, USA) was fixed on the chassis of the vehicle near the left leaf spring rear mount. This is approximately 1.5 m away from the Centre of Gravity (CG) of the vehicle. As recommended by Long (2016), this position away from the centre of gravity reduces the effect of wheel-base filtering which can otherwise create anti-resonance drops in the measured spectrum. The signal was sampled at 1024 Hz and the accelerometer has a range of ± 250 m/s². Only the vertical acceleration was measured as this is the main axis affected by shocks.

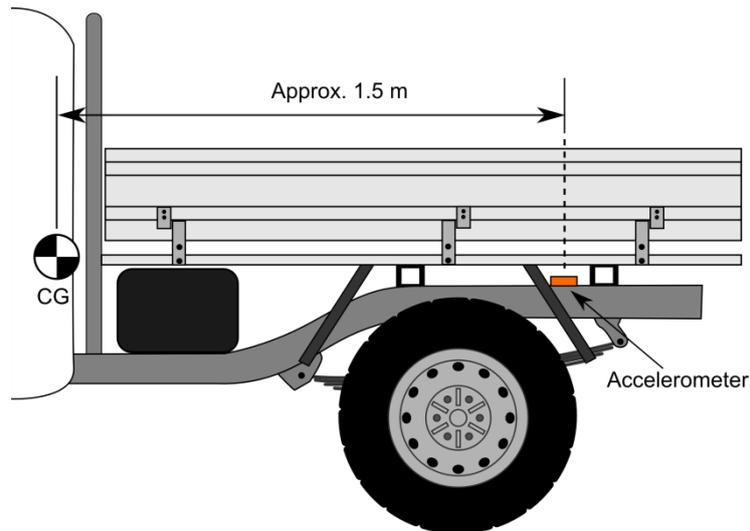


Figure 86: position of the accelerometer on the vehicle rear end, drawing not to scale

9.1.1 Training Measurement Dataset

The objective of the training measurement dataset is to characterise the vehicle's dynamics during normal operation. The measurement was performed on 52.8 km of public roads in the west of Melbourne, Australia (Figure 87), which represents a mix of typical Australian motorways, urban roads and country roads. The information on the vehicle speed and location was recorded with a GPS (Table 7). The measurement duration was close to one hour (3500 s). The vehicle had a mean speed of 55 km/h with a standard deviation of 29.8 km/h. It reached a maximum of 103 km/h. The experimenters occupying the vehicle manually attempted to count the shock occurrences during the acquisition. A shock was counted when both experimenters agreed that the vehicle encountered a shock. The number of manually-identified shocks during the measurement was 117. The overall RMS value of the vehicle vertical acceleration was 1.6 m/s^2 .

Table 7: summary of the measurement datasets

Measurement	Road Type	Time [s]	Speed [km/h]			Distance [km]	Number of shocks []	RMS [m/s ²]
			Mean	Standard deviation	Max			
Training dataset	Public road	3500	55.6	29.8	103	52.8	117	1.6
Validation dataset 1	Asphalt circuit	680	38.7	8.5	67.1	7x1	42	1.5
Validation dataset 2	Off road circuit	450	19.5	6.8	29	10x0.25	40	2.5

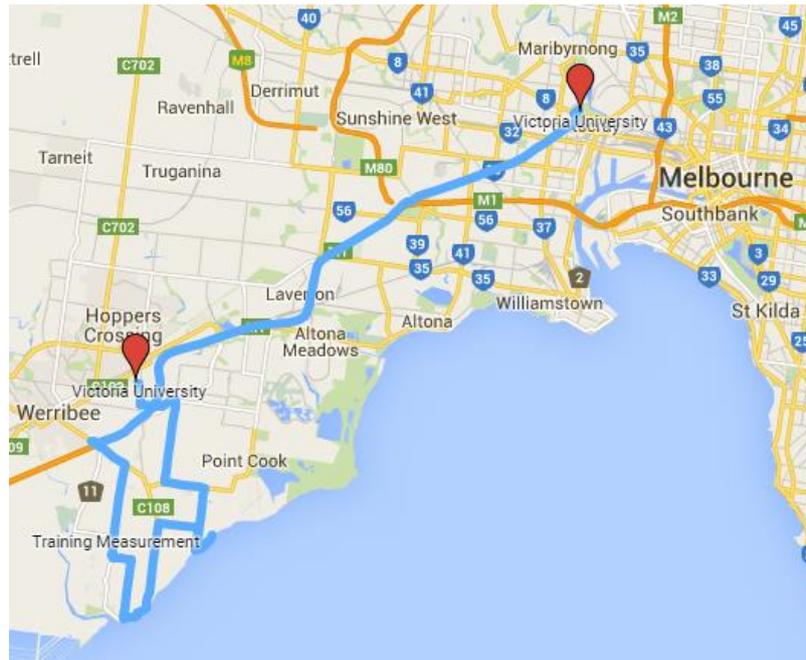


Figure 87: map of the training measurement dataset

9.1.2 Validation Measurement Datasets

The objective of the validation measurement datasets is to calibrate the classifier's Optimal Operation Point (OOP) and to assess how reliably shocks are detected. Therefore, the location of the shocks encountered during the measurements needs to be clearly identified. To do so, the measurements were performed on closed circuits. As the classifier needs to function with varying RMS levels (nonstationary components), asphalted and off-road circuits were used. These circuits were located at Victoria University's Werribee campus.

The validation measurement dataset 1 was performed on a 1 km circuit of smooth asphalted road exempt from any shocks (Figure 88). Six steel plates were fixed on the road pavement to generate shocks when in contact with the left wheels of the vehicle (same side as the accelerometer). Shocks 2 and 5 were composed of a 10 mm high by 80 mm long steel plate which was considered as a small shock. Shocks 3 and 4 were composed of a 15 mm high by 100 mm long steel plate which was considered as a medium shock. Shocks 1 and 6 were composed of a 20 mm high by 150 mm long steel plate which was considered as a big shock.

The vehicle was driven around the asphalted circuit seven times for a total of 42 shocks (Table 7). The measurement duration was 680 s. The shocks density (≈ 6.2 shocks per 100 s) is slightly inferior to the one used in the synthetic signal (10 shocks per 100 s). The vehicle had an average speed of 38.7 km/h (standard deviation and maximum of 8.5 and 67.1 km/h respectively). The RMS level of the vertical acceleration was 1.5 m/s^2 which is similar to the value measured with the training measurement dataset.

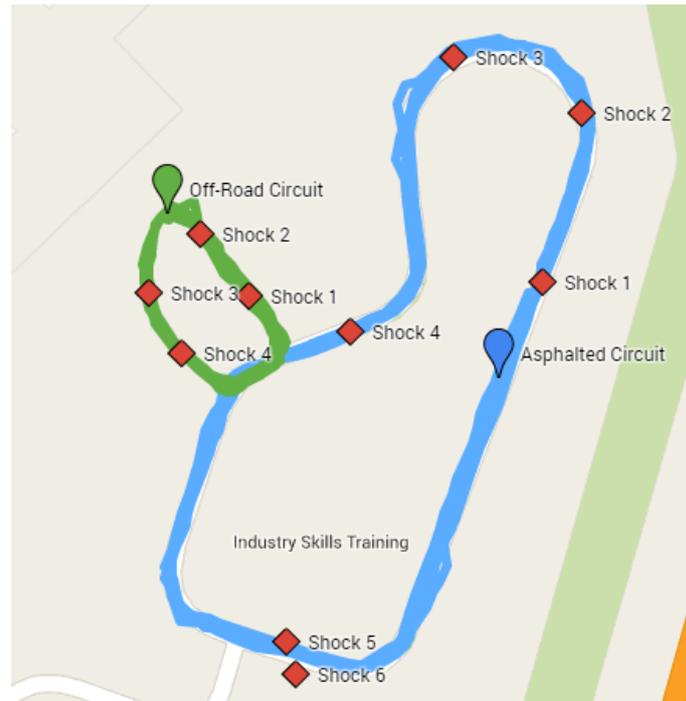


Figure 88: map of the validation measurement datasets and the shock positions

The validation measurement dataset 2 was performed on a 250 m off road circuit (Figure 88). The surface was composed of rough uneven gravel and rocks. Four artificial shocks were added to the circuit but there could have been other aberrations on the surface that may have generated unidentified shocks. For the first five laps, shocks 2 and 4 were composed of a 15 mm high by 80 mm long steel plate and for the remaining five laps they were replaced by a 48 mm high by 98 mm long lumber beam in order to increase the shocks amplitude. This change was made because these shocks were perceived by the experimenters during the measurement as not severe enough for shock detection purpose. Shocks 1 and 3 were composed of a 20 mm high by 150 mm long steel plate through the 10 laps of the measurement.

The vehicle went around the off-road circuit ten times (2.5 km) for a total of 40 shocks (Table 7). The measurement duration was 450 s. The shocks density (≈ 8.9 shocks per 100 s) is close to the one used for the synthetic signal (10 shocks per 100 s). The vehicle had an average speed of 19.5 km/h (standard deviation and maximum values of 6.8 km/h and 29 km/h respectively). The RMS level of the vertical acceleration was 2.5 m/s^2 , which is representative of the severity of the off-road circuit.

9.2 Synthetic Training Signal

It is impractical to perform supervised classifier learning on RVV measurement because the required number of perfectly identified shocks is too large. To overcome this issue, a synthetic RVV signal can be created using the vehicle's dynamic characteristics identified from the

learning measurement dataset. The nonstationary component of the RVV signal was created using the vehicle's vibration spectrum and the statistical distribution of the RMS whereas the transient component was created slightly differently using the Random Decrement signature of the vehicle (Figure 89).

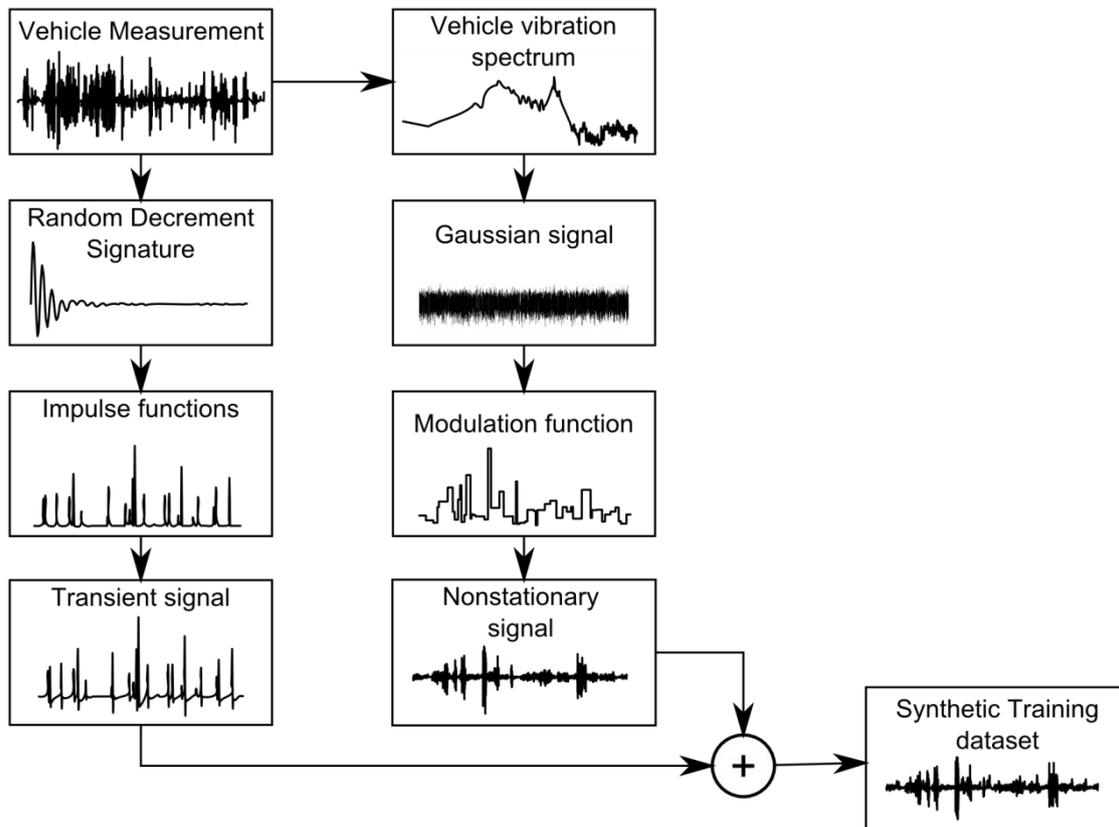


Figure 89: flow chart of the synthetic signal creation from the learning dataset

9.2.1 Nonstationary Component

The nonstationary component of RVV was modelled with a series of Gaussian signals of different RMS values. To represent a realistic spectrum, the Gaussian sequences were created from the average Power Density Spectrum (PDS) calculated on the training measurement dataset (Figure 90).

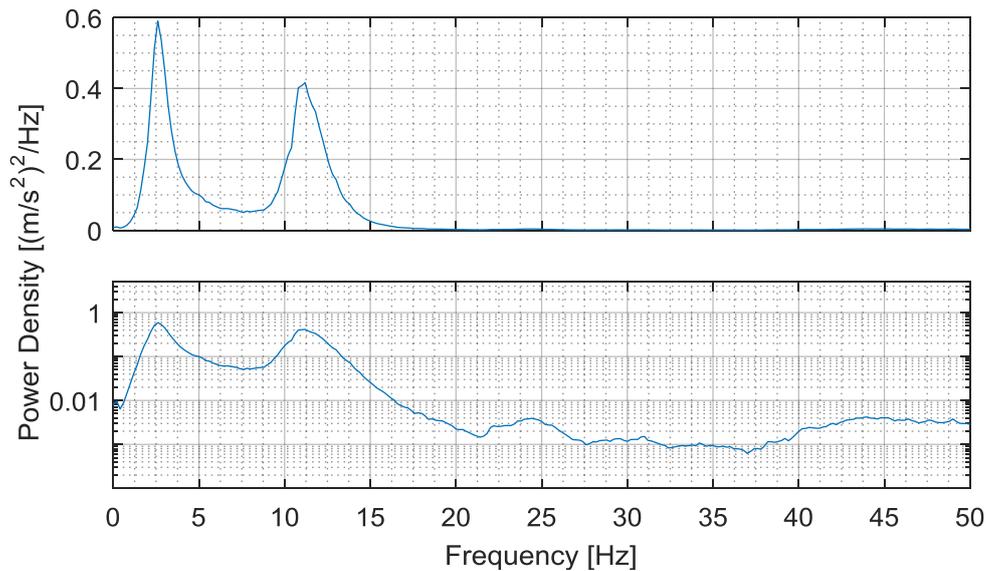


Figure 90: PDS ($\Delta f = 0.2$ Hz, Hanning window, 1405 averages) of the training measurement dataset (linear scale at the top and semi-log scales at the bottom)

To create a Gaussian time signal with the same spectrum as the vehicle, the PDS was then scaled into an amplitude (acceleration) spectrum with a uniformly-distributed random phase. The frequency resolution of the spectrum was interpolated to match the desired time signal duration. Once interpolated, the spectrum was transformed to the time domain with the Inverse Fast Fourier Transform (IFFT).

The RMS value of the resulting signal was then modulated in a random fashion to generate a nonstationary signal. The modulation levels were scaled to match the RMS level of the measurement, which was 1.6 m/s^2 . Samples of the training measurement dataset and the equivalent nonstationary signal are presented in Figure 91. More detail on how the synthesise RVV nonstationary signals are given in Chapter 4.

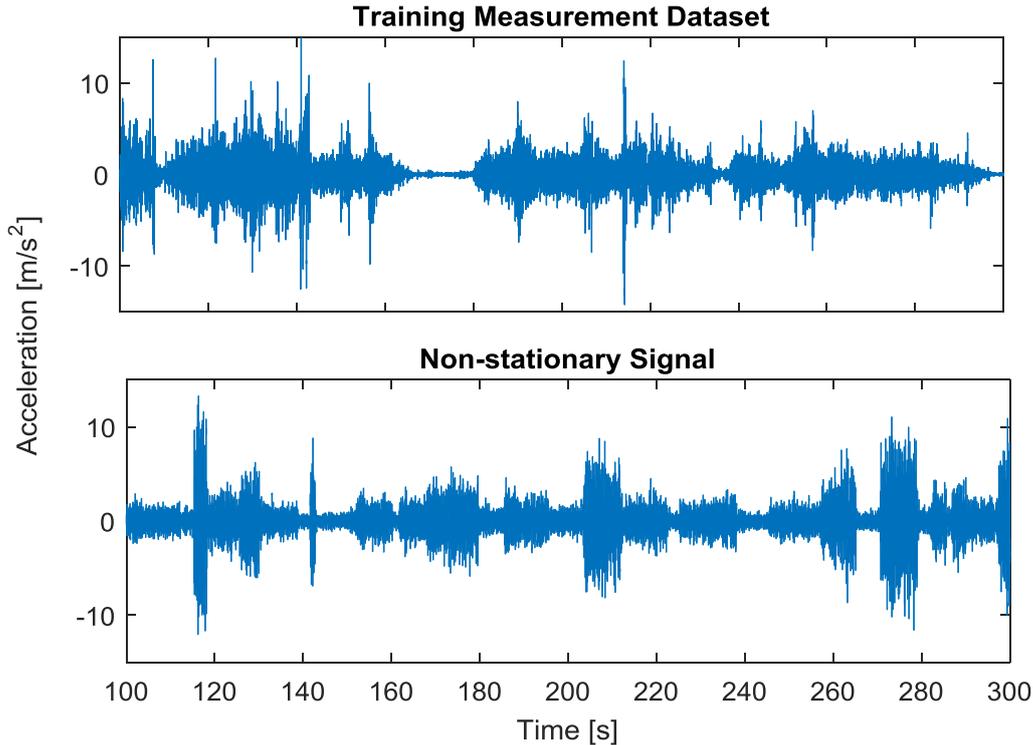


Figure 91: typical segment of the training measurement dataset and nonstationary synthetic signal

9.2.2 Transient Component

The RVV signal synthesiser presented in Chapter 4 generates shocks from the impulse response of the vehicle based on its FRF. However, the excitation input was unknown during *in-situ* RVV measurement, so the vehicle's FRF cannot be calculated using conventional methods based on auto and cross-correlation functions. Therefore, an operational modal analysis approach had to be used.

One proven method to estimate the impulse response of a road vehicle using *in-situ* measurement is the Random Decrement Signature (Ainalis, 2014, Milliken et al., 2001). The Random Decrement Signature is based on a convolution process but in contrast with the auto-correlation function, it is not calculated on every time sample. Instead, the Random Decrement Signature uses a triggering condition to average the product of the auto-convolution such as:

$$D_{xx} = \frac{1}{N} \sum_{i=1}^N x(t_i + \tau_w)x(t_i), \quad (\text{eq. 39})$$

where $x(t_i)$ is the part of the signal that meets the triggering condition and N is the number of averages. For instance, the triggering condition can be set on the zero-crossing where the Random Decrement Signature will be calculated every time the signal crosses zero. The triggering conditions can be alternatively set on a threshold value and on the signal slope

polarity. The advantage of the Random Decrement Signature over the auto-correlation function is that the function is scaled so it has a constant amplitude which is proportional to the impulse response of the vehicle. The signature's scale can be adjusted with the trigger condition.

The optimal parameters to estimate the impulse response of the vehicle using Random Decrement Signature were studied by Ainalis (2014). The trigger was set on zero-crossing with no specific slope polarity or minimal number of points between the averaging. The triggering condition provided around 30,000 averages on the training measurement dataset which provides an accurate impulse response estimation. A 2 s window was used for the signature length. The measurement dataset was filtered with a low-pass filter at 25 Hz (5th order Butterworth) in order to estimate only the first two natural frequencies of the vehicle (as per the analytical FRF model used in Chapter 4).

The Random Decrement Signature of the vehicle is presented in Figure 92. The 2 s window was sufficient to allow complete decay of the signature. The FFT of the signature gives a scaled estimation of the vehicle's FRF (Figure 93). As seen in the figure, unsprung mass damping (first peak) is high compared to the sprung mass damping (second peak) which is unlike the FRF model used in Chapter 4. This is because the vehicle's suspension is designed to have an optimal damping with a certain payload and the measurements were performed without any significant payload. This means that with a heavier payload, the natural frequency of the sprung mass would decrease and the damping of the suspension would be larger than the critical damping of the system resulting in a FRF closer to the model used in Chapter 4. Nevertheless, this technicality does not affect the machine learning process because the classifiers will be trained using this new vehicle model.

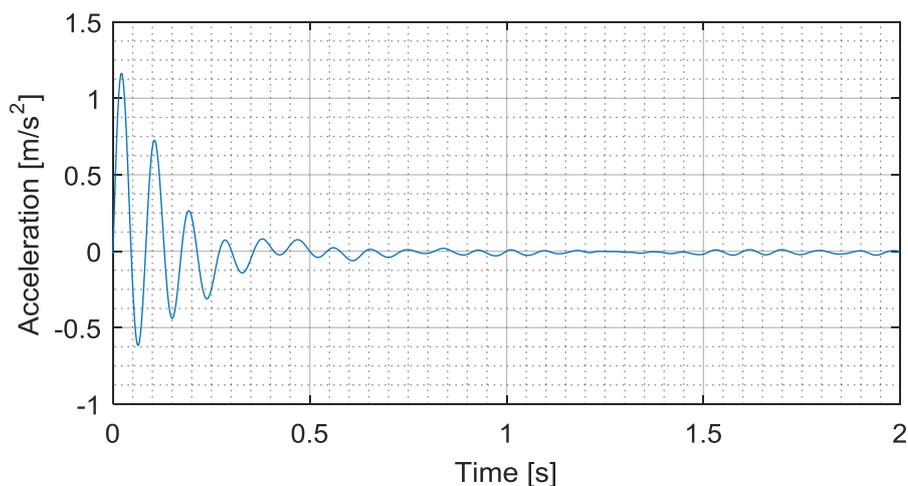


Figure 92: Random Decrement Signature of the vehicle

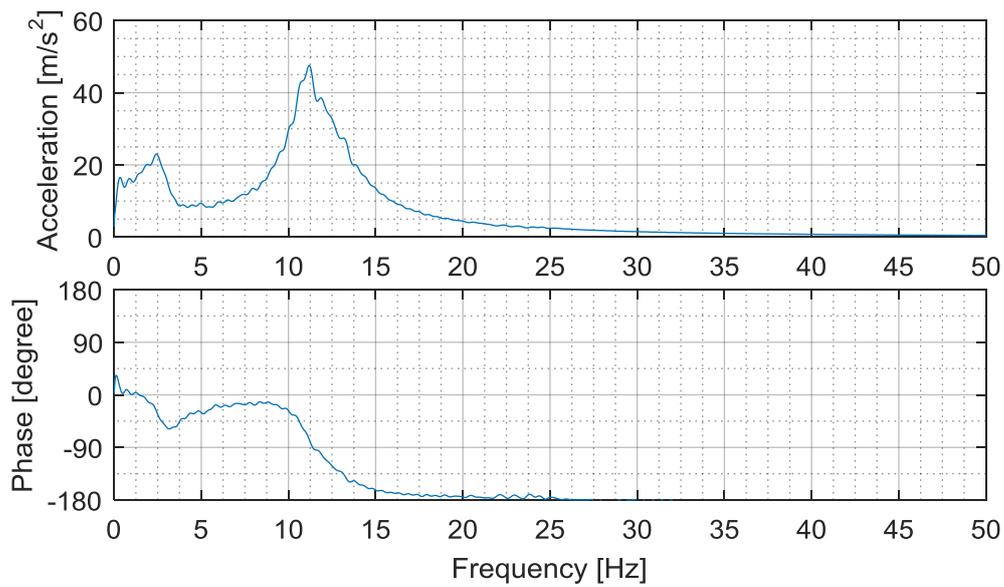


Figure 93: acceleration spectrum of the Random Decrement Signature of the training measurement dataset

To generate shocks, the Random Decrement Signature was convoluted with a sequence of randomly distributed half-sine shock (impulse) functions with the same characteristics (duration and amplitude) as those used for the classifiers' training in Chapter 7. This convolution created a series of randomly scaled and randomly occurring impulse responses which was then superimposed onto the nonstationary signal (Figure 94). The position of the shock functions was used to index the shocks during the learning process.

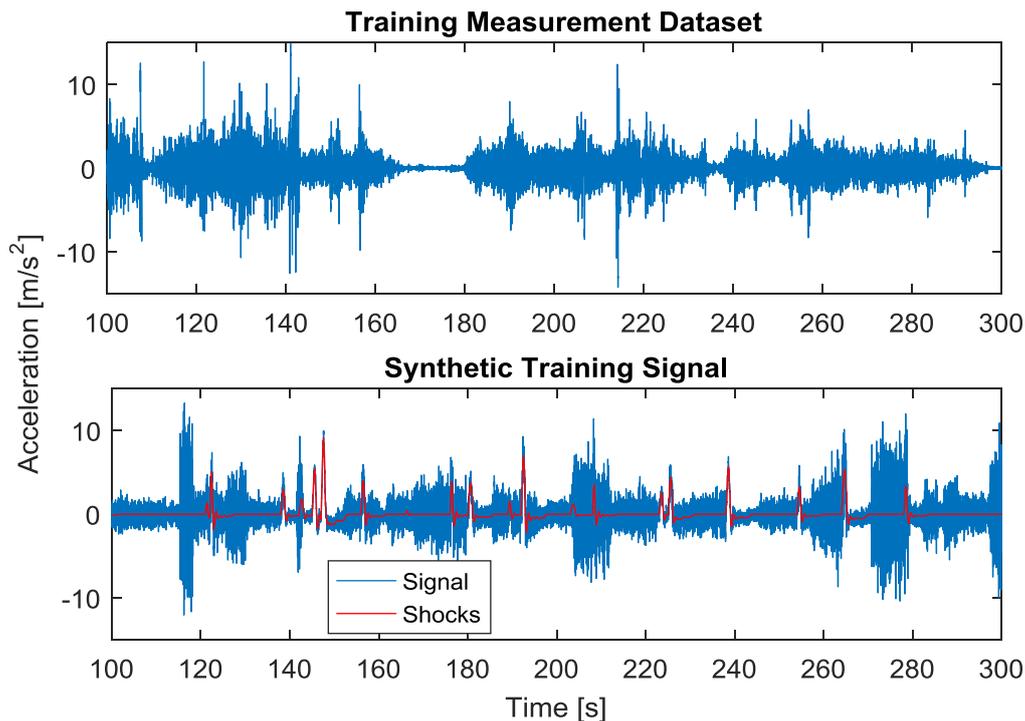


Figure 94: typical segment of the training measurement dataset and synthetic training signal (nonstationary and shocks)

9.3 Classifier Training

Decision Tree, 20NN, Bagged Ensemble and SVM classifiers were trained with the synthetic training signal using their optimal parameters identified in Chapter 8. A signal composed of 300 shocks (10 shocks per 100 s) was used for the classifiers' learning and another signal with 350 shocks was used to build PERFO curves and to calibrate the classifiers' OOP.

The synthetic training signal was processed into predictors presented in Section 7.2 (such as moving statistic, HHT and DWT). The learning process was sped up by setting the predictors' format to a single-precision floating point (32 bits). Preliminary analysis shows that a factor of 32 instead of 64 used in Section 7.3 gives a better shock detection sensitivity for these specific measurement datasets. This may be caused by the vehicle's FRF. Since the sprung mass is more damped than the unsprung mass, the impulse response of the vehicle is dominated by the unsprung mass which has a natural frequency around 11 Hz which falls over the 8 Hz Nyquist frequency of the 64 times decimated dataset. The hypothesis is therefore that the sampling frequency of predictors must be at least twice the preponderant impulse response frequency.

9.3.1 Synthetic Calibration

Under ideal conditions, synthetic RVV signals would be sufficient to fully train, calibrate and validate the classifiers. However, as described in the next section this does not seem to be possible as the OOPs defined with this synthetic calibration were not sufficiently sensitive to

detect the real measurement shocks. The synthetic calibration approach is therefore presented to compare the classifiers' learning with the previous learning performance made in Chapter 8 from a generic vehicle model.

The synthetic calibration provides ROC curves with slightly higher AUC values than the classifiers trained in Section 8.2.5 (Table 8). Considering the variability of the ROC curves, both learning approaches have similar performances.

Table 8: classifiers' performance comparison between the learning made in Section 8.2.5 (Generic) and the synthetic validation (Vehicle)

Classifier	AUC		PERFO Fall-out	
	Generic	Vehicle	Generic	Vehicle
Decision Tree	0.81	0.86	0.05	0.08
Euclidean 20NN	0.84	0.86	0.04	0.17
Bagged Ensemble	0.85	0.92	0.04	0.08
Gaussian SVM	0.83	0.89	0.03	0.08

The classifiers' OOP was defined using the PERFO curves computed for the synthetic dataset (Table 8 and Figure 95). The Decision Tree, Bagged Ensemble and Gaussian SVM classifiers have similar PERFO fall-out values, around 0.08, which is about twice the values seen in Section 8.2.5. The 20NN classifier has a higher PERFO fall-out of 0.17, which is more than four times its corresponding value in Section 8.2.5.

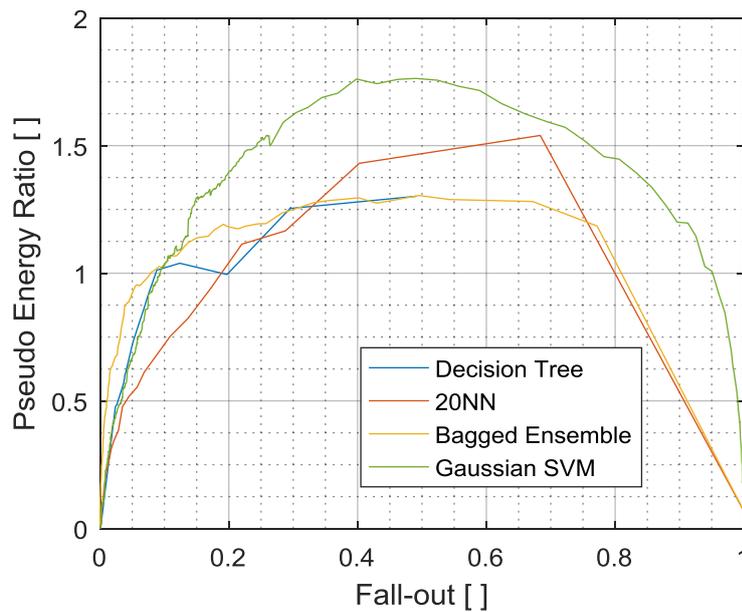


Figure 95: PERFO curves of the synthetic validation dataset

The interesting point is that even if the 20NN classifier has the highest PERFO fall-out, the distribution of its detections is the most accurate in most of the acceleration bandwidths (Figure 96). For the highest bandwidth, only the Bagged Ensemble classifier has a lower error. Compared with the previous learning (Section 8.2.5), the classifiers have a lower misdetection rate across the acceleration range.

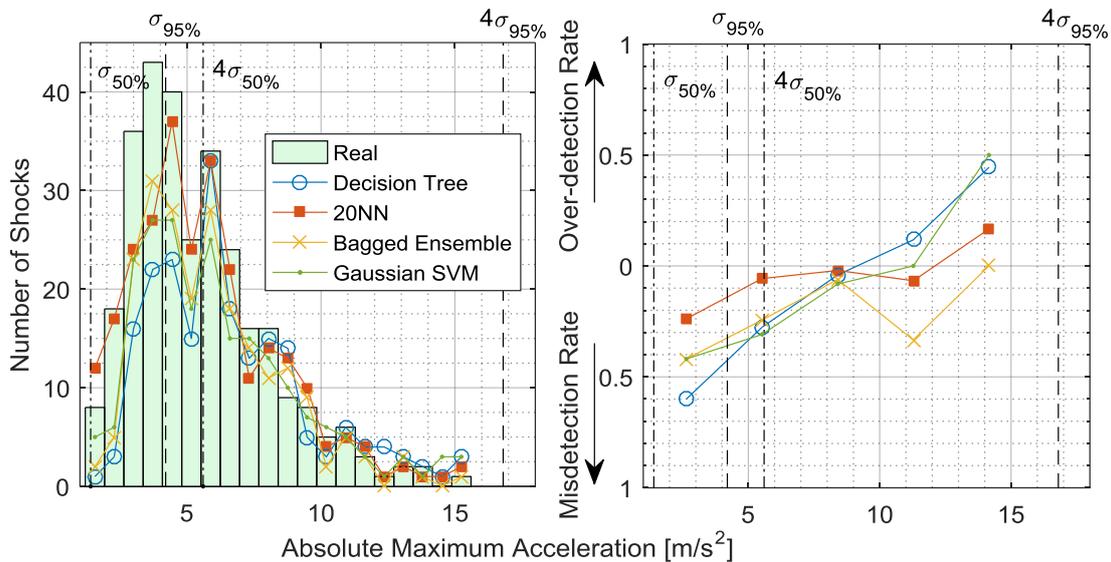


Figure 96: maximum absolute acceleration distributions of the detections and shocks of the synthetic validation dataset (left) and Mis-detection/Over-detection graph based on these distributions (right), $\sigma_{50\%}$, and $\sigma_{95\%}$ are, respectively, the median and 95th percentile RMS value of the stationary Gaussian random segments composing the signal

9.4 Validation on Real Measurements

Validation using a synthetic validation dataset gave an appreciation of how the classifiers were successfully trained but it did not give any insight into their application on real measurement datasets. This ultimate assessment can be performed using the same validation procedure presented in the previous section on the validation measurement datasets.

One of the first observations from the validation measurement datasets application is that the OOP calibration with a synthetic dataset is inadequate. It appeared that the detection threshold defined from the synthetic dataset is not sensitive enough for real shock detection. Using this threshold, each classifier detected less than 10 shocks in all the validation measurement datasets (composed of more than 80 shocks). The classifiers needed to be recalibrated using the PERFO criterion on the validation measurement datasets.

Before looking at the PERFO curves, it is important to look at the ROC curves to evaluate the actual sensitivity of the classifiers. The PERFO curve is a good tool to assess the detection algorithm's maximum amplitude with respect to real shocks' maximum amplitude. However, it does not indicate if the detections are true or false. For instance, a classifier could have a pseudo energy ratio of one just by guessing where the shocks are without using any significant statistics. This is where the ROC curves have an important role because they represent the relationship between the sensitivity and the fall-out which gives a good appreciation of the classifiers' detection accuracy.

To assess the detection accuracy, ROC curves were computed with the validation measurement dataset made on the asphalted circuit (Figure 97). This circuit has potentially fewer unidentified shocks than the off-road circuit so it provides more accurate ROC curves. These ROC curves lead to very interesting results. The detection of the Decision Tree, 20NN and Bagged Ensemble classifiers are not better than a classifier based on random guessing (chance). They have AUC values of 0.49, 0.43 and 0.53 respectively. Considering the AUC's uncertainty calculated in Chapter 7, these values are equivalent to 0.5, which represents detecting shocks by guessing. Only the Gaussian SVM provides accurate detections with an AUC of 0.94, which is a 6 % increase from the synthetic validation.

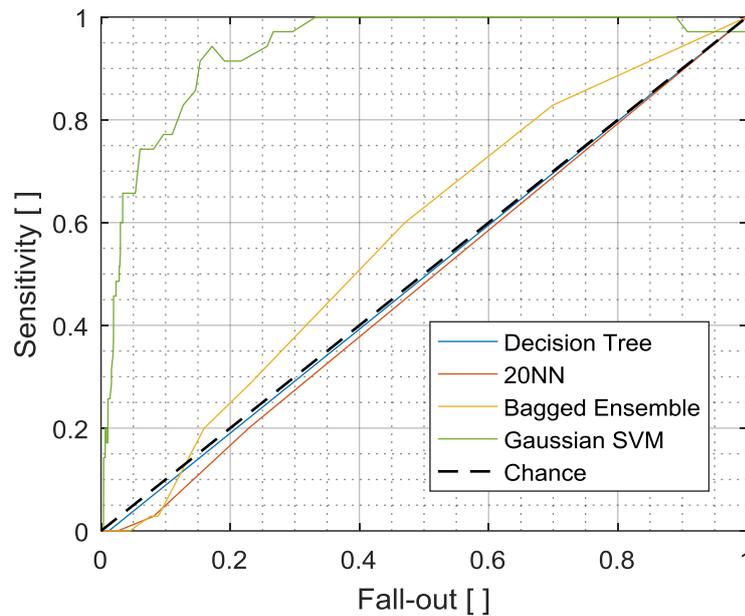


Figure 97: ROC curves of the validation measurement dataset made on the asphalted circuit

This is very interesting because all the classifiers have AUC values well over 0.5 when validated with a synthetic dataset. This means that learning from a synthetic dataset cannot be adapted for a real measurement dataset for the Decision Tree, 20NN and Bagged Ensemble classifiers; only the Gaussian SVM can manage this adaptation. This could be because the Gaussian SVM is the only nonlinear classifier in this evaluation. It is therefore possible that real shocks have behaviours that cannot be handled by linear classifiers. Thus, there is no need to assess the linear classifiers any further and only the Gaussian classifier is included for the remainder of the analysis.

The Gaussian SVM's PERFO curve was computed for both validation measurement datasets (Figure 98). The off-road curve reaches a pseudo energy ratio of one at a higher fall-out value (0.06) than the asphalted curve (0.05). This may be caused by the unidentified shocks present on the off-road circuit that could have been detected. The small difference between both values suggests that there may not be too many unidentified shocks and the off-road measurement dataset is reliable for validation purposes.

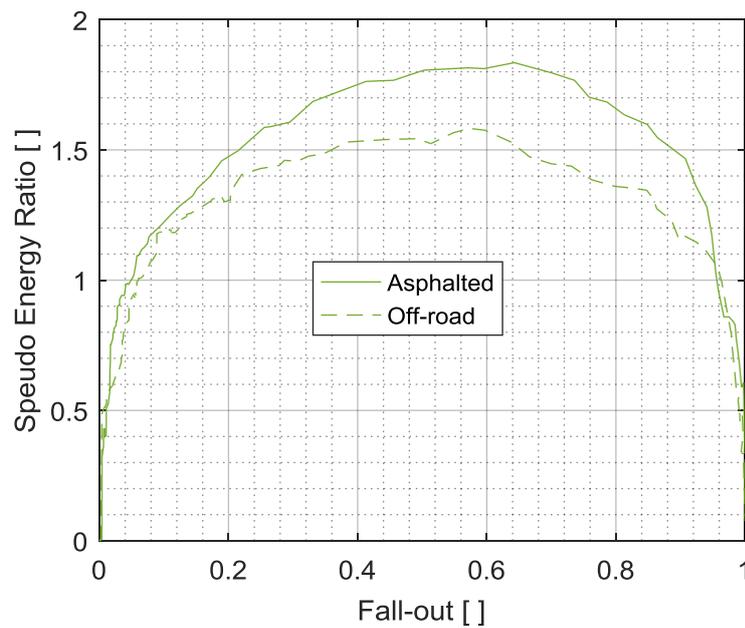


Figure 98: Gaussian SVM's PERFO curve of the validation measurement datasets

The OOP calculated from the PERFO criterion gives a detection threshold for the asphalted and off-road circuit of -0.96 and -0.87 respectively. The SVM represents shocks in the positive space and the “no-shocks” in the negative space, so the smaller the detection threshold is, the more shock sensitive the classifier is. In theory, scores under -1 represent the space where the “no shock” events of the training dataset are separated from the shocks. The detection threshold being very close to this value suggests that the shocks buried in the synthetic dataset were too easily identifiable. In other words, the artificial shocks superimposed on the synthetic nonstationary signal were too severe. This also explains why the synthetic calibration was not sensitive enough to detect real shocks.

The difference between threshold values from both circuits is not great enough to affect the detection distributions. For both validation measurement datasets, the maximum amplitude distribution remains the same using either detection thresholds (Figure 99 and Figure 100). There are almost no errors on the high amplitude shocks encountered on the asphalted circuit. This high amplitude accuracy was expected as this circuit has a smooth pavement which does not generate high amplitude vibration that could be misclassified as shocks. The same level of high amplitude detection accuracy is almost achieved on the off-road measurement where only one shock above 16 m/s^2 is misdetected. The low amplitude detections on the asphalted circuit are also very good with a maximum misdetection rate of 0.25 in the low amplitude range.

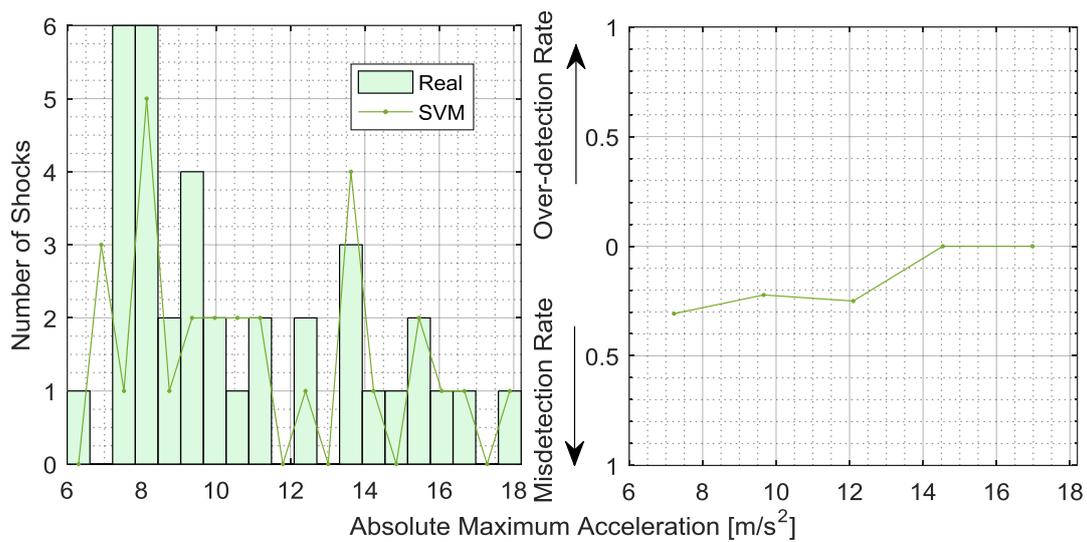


Figure 99: maximum absolute acceleration distributions of the detections and shocks of the asphalted circuit validation measurement dataset (left) and Misdetecion/Over-detection graph based on these distributions (right)

As the off-road measurement inherently contains more background vibration, its low amplitude detections are less accurate than its asphalted counterpart. Nonetheless, detection is still within the range of the synthetic validations (Figure 100). The maximum misdetection rate is 0.6 at the lowest acceleration bandwidth and the maximum over-detection rate is 0.3 at the second bandwidth. Considering the relative importance of detecting high amplitude shocks, the Gaussian SVM classifier performs very well on real RVV measurements.

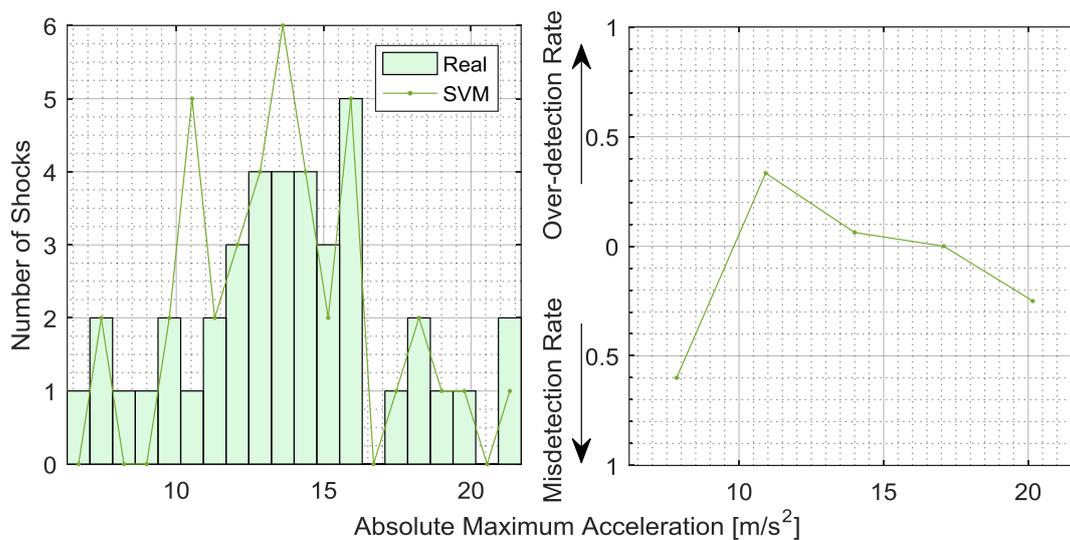


Figure 100: maximum absolute acceleration distributions of the detections and shocks of the off-road circuit validation measurement dataset (left) and Misdetection/Over-detection graph based on these distributions (right)

Another way to appreciate the accuracy of shock detections is to carry-out direct observation in the time domain. Figure 101 shows the location of the detected shocks in the asphalted circuit measurement dataset. The classifier missed shock number 5 four times out of seven, it was a ‘small’ shock positioned 10 m before a big shock (shock 6). Other than that, it seems as if there are no patterns in the false and misdetection in this measurement dataset.

The interpretation of the off-road measurement dataset’s detection in the time domain (Figure 102) is interesting as the signal is noisier and the shocks do not clearly stand out as on the asphalted circuit. Most of the false-detections occurred in the last segment of the lap where the circuit surface was rougher and where there were more likely to be unidentified shocks.

There are very few misdetections in the measurement dataset. Most of those occurred in the first five laps where shocks 2 and 4 were physically smaller than for the remaining five laps. Each of those smaller shocks was missed twice during the first five laps. This is remarkable because these shocks were detected most of the time even though these two smaller shocks were perceived as not severe enough for shock detection purposes by the experimenters present in the vehicle during the measurement. This suggests that the Gaussian SVM classifier could be better than humans at detecting shocks. The two other misdetections are the shocks numbered 3 and 4 (in their more severe state).

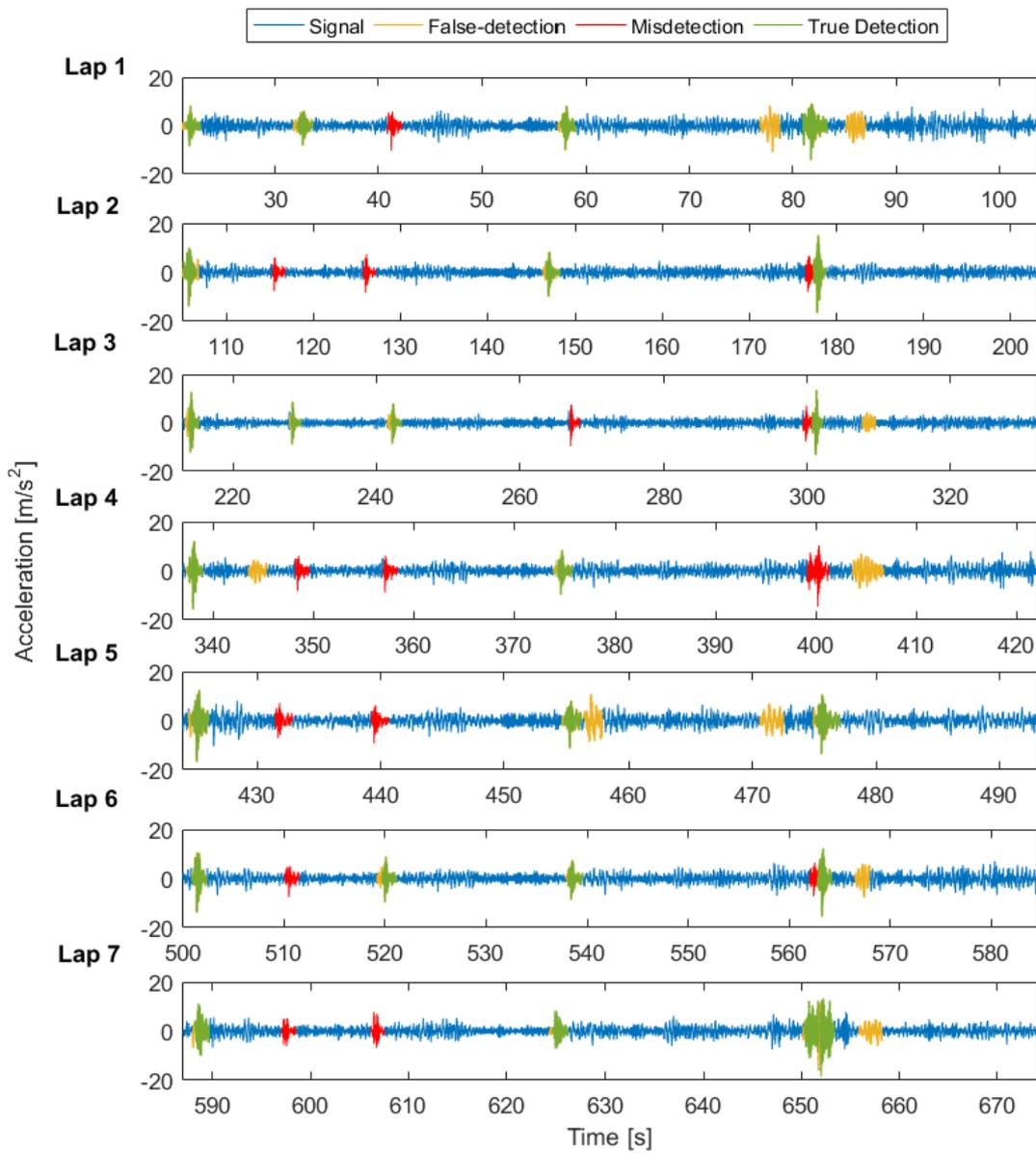


Figure 101: position of the detections and shocks in the asphalted circuit validation measurement dataset, the false-detection, misdetection and true detection are based on the artificial shocks placed on the circuit

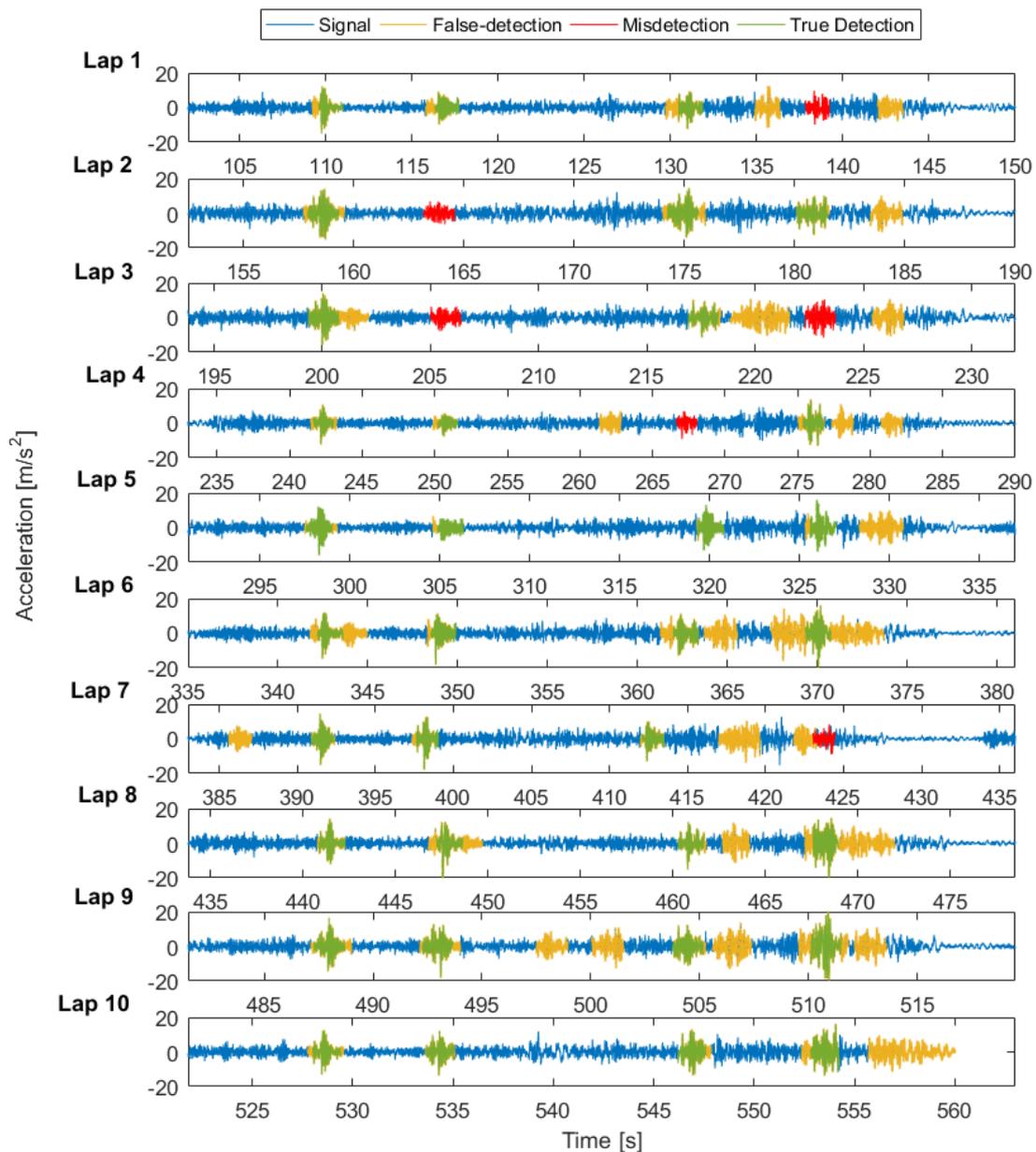


Figure 102: position of the detections and shocks in the off-road circuit validation measurement dataset, misdetection and true detection are based on the artificial shocks placed on the circuit

9.5 Classifier Application

Shock detections on validation measurement datasets being a success, the Gaussian SVM classifier can be applied to measurements made on the vehicle with the same payload. The machine learning approach presented in this chapter allowed the identification of shocks buried in the training measurement dataset. The classifier being trained on a synthetic reconstruction of the training measurement dataset, the classifier has not *a priori* any knowledge of the signal, there is no risk of bias of the detection made on the training measurement dataset. This is an

important benefit to this machine learning approach; the classifier can be trained and applied on the same signal (after a calibration from a short validation measurement dataset).

The distribution of the absolute maximum acceleration of the shocks identified in the training measurement dataset is presented in Figure 103. Since this is a counting process, the distribution should in theory follow a Poisson distribution function. However, the overrepresentation of shocks in the 0 to 1.2 m/s² and 8 m/s² to 9.2 m/s² bands distorts the Poisson fit and a Gaussian distribution function offers a better fit. This being said, the 160 shocks identified in the measurement dataset may not be sufficient to clearly describe the type of statistical distribution for the shock amplitudes. Probability Density Functions (PDF) of longer measurement datasets (*i.e.* with more shocks) should ideally be analysed to accurately model the shock amplitude distribution.

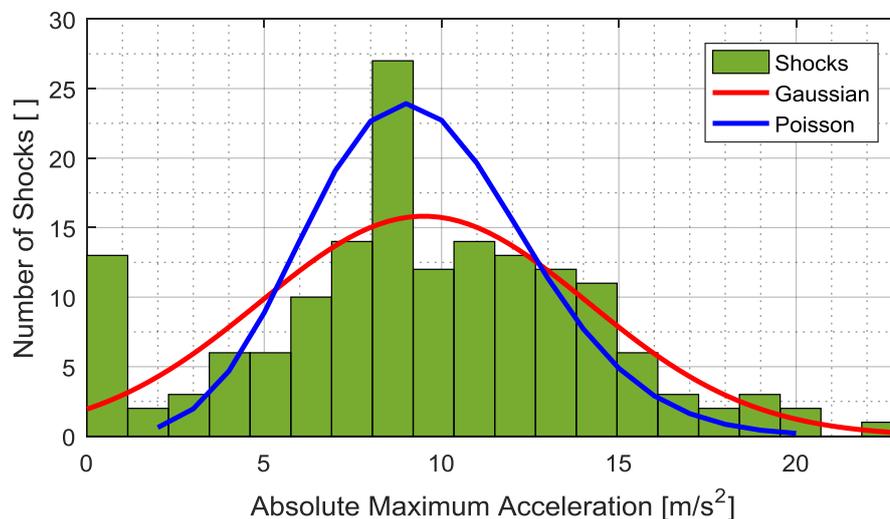


Figure 103: maximum absolute acceleration distributions and Gaussian fit of the shock detections made by the Gaussian SVM classifier on the training measurement dataset

The total number of shocks detected by the Gaussian SVM classifier (160) is 37 % more than the manual estimate (count) made during the measurement (117). This difference corroborates the humans' low shock sensitivity observed on the off-road validation measurement dataset. This relatively small difference also suggests that the classifier's calibration was adequate.

The positions of the detected shocks within the signal are shown in Figure 104. It is worth noticing that the classifier does not base its detection on a sudden acceleration peak, as some of these peaks are not identified as shocks. For instance, the vehicle traversed four consecutive speed bumps between 790 and 880 s and only the first bump (as encountered by the front wheel) was detected as a shock (Figure 105). As those speed bumps were passed at relatively low speed (within 19 km/h and 24 km/h according to the GPS data), their duration was longer than the

maximum duration included in the learning. To include such events in the detections, the synthetic training signal should include shock pulses of larger width and higher amplitude to recreate the vehicle's response to speed bumps.

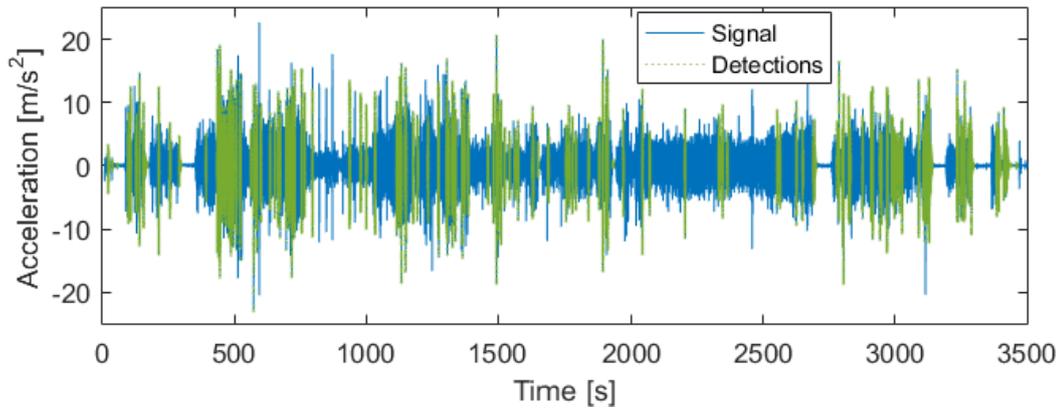


Figure 104: shock detections made by the Gaussian SVM classifier on the training measurement dataset

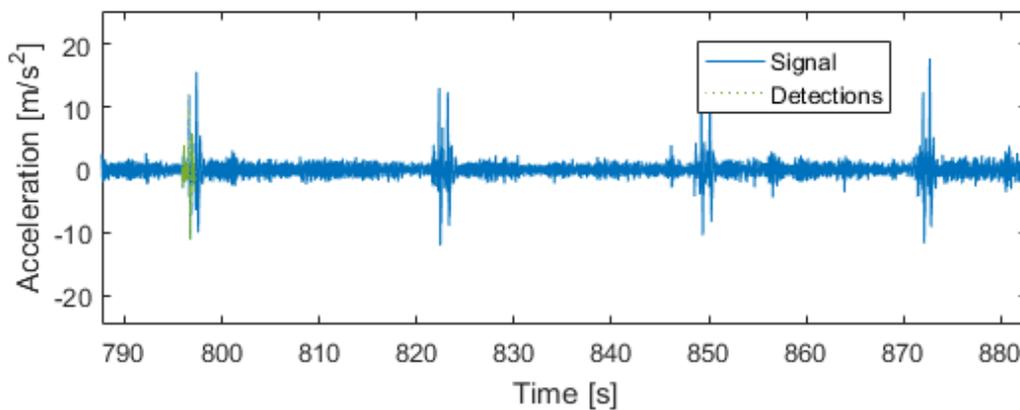


Figure 105: segment of the training measurement dataset composed of four speed bumps

It was mentioned in Chapter 4 that shocks can also be described in terms of distributions of the impulse durations and time interval between occurrences. Since the machine learning process fixes the detection duration at 1.4 s for a typical RVV, the impulse duration cannot be analysed. Nevertheless, the fixed detection duration does not affect the time interval between the shocks and its PDF can be computed from the analysis (Figure 106). The distribution has a similar shape to the distribution of the duration of RVV Gaussian segments studied by Rouillard (2007a) and presented in Section 4.1. The same hyperbolic function (eq. 9) was curve fitted to the distribution:

$$p(d) = \frac{C}{\sinh(kd)}$$

where d is the segment duration and C and k empirical constants. These constants were estimated to $c=1.235$ and $k=0.0153$ by a least mean square regression. As for the absolute maximum acceleration distribution, this represents only one RVV signal and a large sample size should be studied before drawing any conclusion on the time interval between shocks distribution.

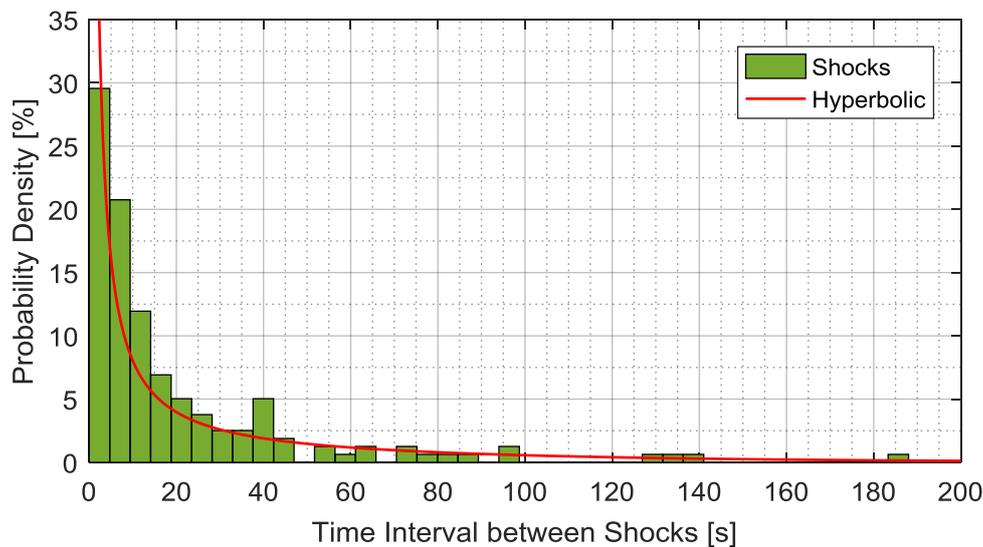


Figure 106: time interval between the shock detections distribution made by the Gaussian SVM classifier on the training measurement dataset and hyperbolic fit

9.5.1 Spectral Analysis

Spectral analysis of RVV is broadly used to define efficiency of protective packaging. So it is interesting to look at the effect of the shocks on a RVV spectrum. The vibration created from the road pavement profile (nonstationary component) and the shocks do not have the same spectrum. The road pavement profile has a Brownian spectrum (ISO-8608), *i.e.* its amplitude decays in function of the frequency squared, which generates a greater vehicle response at low frequency. Shocks are different. They are created by short impulses or quasi-Dirac functions which have a response on a frequency bandwidth inversely proportional to their time width. For instance, short impulses can generate a vehicle response on a larger bandwidth.

In an ideal scenario, this implies that RVV signals without shocks could have a lower spectral amplitude at high frequency than signals including shocks. However, this phenomenon cannot be observed in Figure 107 where the shock detections have been removed from the training measurement dataset recorded on public roads. The PDS of the signal with shocks has about a 25 % higher response at 2.5 Hz and 11 Hz than the one sans-shocks. Except of these peaks, no other discrepancies are observed even on a logarithmic scale. This is due to the dynamic

behaviour vehicle including the enveloping effect of the pneumatic tire over the road aberrations.

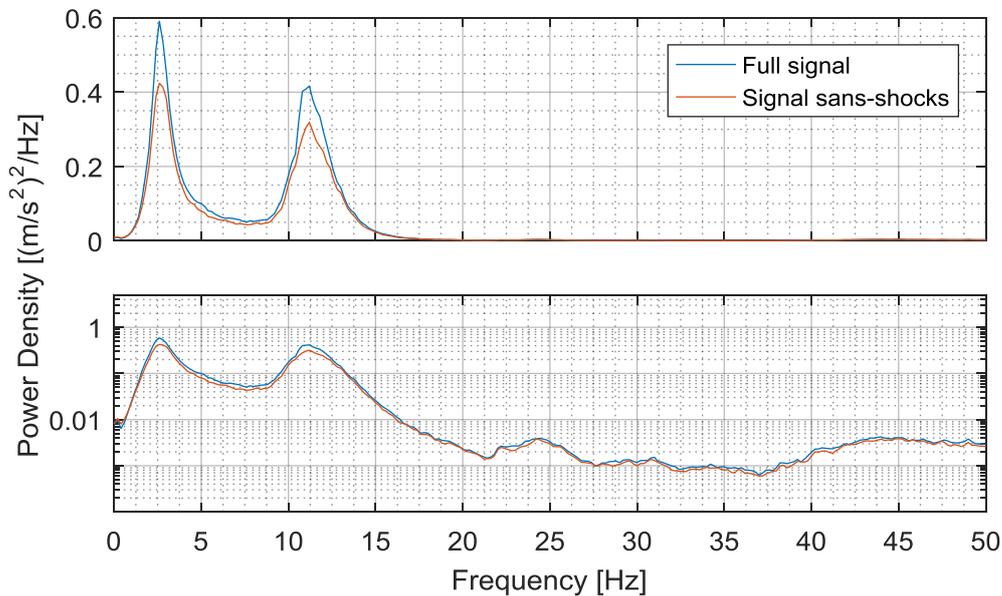


Figure 107: PSD of the training measurement dataset with and without shocks (respectively 1405 and 1228 averages on a linear and log scale, $\Delta f = 0.2$ Hz, Hanning window)

9.6 Future Validation Work

It was shown that the Gaussian SVM classifier can detect shocks on real RVV measurements, but there are still more validations to be undertaken before applying it to real life distribution problems. For instance, the limits of the learning have to be defined and these questions must be addressed: how can the learning be extended? Can the same learning be applied to the same type of vehicle? To different payloads? Which type of shocks are detected (shocks duration)?

The calibration limitations and requirements also have to be addressed. For instance, it may be possible to apply the same learning to different vehicles, but each of them may require a specific calibration. The optimal validation measurement dataset must also be defined as the type and number of shocks may affect the complexity of the calibration.

This more comprehensive validation requires well-defined application contexts such as specific route distributions, transport vehicles and shipments. This would be the final step before the classifier implementation to model the RVV. Such work is however beyond the scope of the thesis, which is to explore how the different modes of RVV can be identified. This opens opportunities to researchers to apply this machine learning approach to study RVV in more depth.

9.7 Conclusion on Model Validation

This chapter has demonstrated that at least one machine learning classifier can be used to detect shocks buried in real RVV measurements, namely, the Gaussian SVM. This classifier was trained from a synthetic dataset which was created from the dynamic behaviour of a specific vehicle. The vehicle's dynamics were characterised with a training measurement dataset made during normal operation. Once trained, the classifier was calibrated and validated using measurement datasets made on closed circuits composed of artificially added shocks. From the classifiers optimised in Chapter 8, only the nonlinear classifier (Gaussian SVM) could accurately detect shocks and had an AUC above 0.5 (AUC = 0.94).

As the training measurement dataset was not directly used in the classifier's training, the training signal was analysed with the Gaussian SVM. The analysis revealed that the distribution of the shocks' absolute maximum acceleration can be described with a Gaussian function. It was also shown that the classifier was not totally sensitive to the shocks created by speed bumps, which could more likely be changed by using a different synthetic training dataset.

Chapter 10: CONCLUSIONS

This thesis addressed the fundamental piece of the puzzle for packaging optimisation which is to characterise the modes of the vibration induced to freight during road transport. This characterisation is complex because the RVV (Road Vehicle Vibration) is composed of three different modes: *i.e.* the nonstationary random, transient and harmonic components. These modes are described by different statistics and mathematical models so they must be characterised separately. The problem is that the modes cannot be measured individually, as they coexist and are confounded in a single acceleration signal. Therefore, they must be identified and extracted from the signal prior to their characterisation. The main achievement of the thesis was to integrate many signal analysis methods in a machine learning algorithm to identify RVV modes. This proved the main thesis hypothesis stated in Chapter 3 that the distinct modes (or components) constituting the vibration produced by road vehicles can be reliably detected and separately identified.

This novel approach has important advantages over the current RVV simulation methods because it detects the shocks present in RVV signals. The shocks and nonstationary random vibration can then be analysed separately. Packaging can therefore be optimised based on more reliable design criteria because the shocks and nonstationary random vibration can be more accurately characterised. Unnecessary cushioning material will be excluded in their design without the risks of damaging product thanks to realistic RVV simulation.

One useful outcome of the thesis is the development of a RVV signal synthesiser. This tool was developed for the specific purpose of evaluating various signal analysis techniques as well as to train and validate the machine learning algorithms. The synthesiser uses a comprehensive RVV model where the contribution of each RVV mode can be purposely defined and individually controlled. Beyond its applications in algorithm validation, the signal synthesiser can be directly used in a numerical simulation model or to drive a vibration table in a physical simulation. Using the characterisation of each mode for typical transport route, the synthesiser can create a simulation signal that represents typical dynamic load of the typical route.

The review of the existing signal analysis methods showed their benefits and shortcomings and it turned out that there is no universal method which can identify the different RVV modes. The reviewed methods were found to complement each other. For instance, the Hilbert-Huang Transform (HHT) can detect nonstationary components but lacked sensitivity for shocks. On the other hand, the Discrete Wavelet Transform (DWT) detects shocks reasonably well, but it struggles with nonstationary components. It is from this complementarity that machine learning came into form. The ability of machine learning to analyse significant volume of data was perfectly suited to integrate many analysis methods to create a reliable shock classifier.

As a first attempt to use machine learning to identify the RVV modes, the RVV signals were simplified and only the two principal modes were considered (*i.e.* the nonstationary and transient components). The shock classifiers were evaluated and optimised following the best machine learning practices. The existing evaluation methods gave a fair insight on the shocks classification performance, but a more specific method was developed to ensure the classifiers provide even better predictions. This new evaluation method is a significant contribution to machine learning as it could be applied on other classification problems where the detections have different relative importance.

The four best performing machine algorithms were applied on real RVV measurement. The objective of the real vehicle implementation is twofold: the first part is to validate that classification on real RVV measurement is possible and second is to assess how machine learning process can be applied to real distribution environment context. This is the keystone of the thesis as it confirms that the proposed shock detection method works and that it can be applied to real measurement. The application is not as straight forward as one may believe. The machine learning algorithms need a large training dataset to base its predictions. As the location of the shocks encounter on the road is very difficult to precisely define, RVV measurements cannot be used to train the algorithm. To overcome this difficulty, the machine learning training was made from synthetic RVV signals generated to represent a specific vehicle using an unlimited training dataset from limited validation measurements (Figure 108). The implementation scheme showed very good classification accuracy with the Gaussian Support Vector Machine (Gaussian SVM) algorithm. As a result of the method presented in the thesis, Gaussian SVM algorithm could now be used on a larger measurement size in order to characterise actual distribution routes and create better RVV simulation.

Outside packaging optimisation applications, this research also has a significant contribution to knowledge. The developed modelling method itself will be useful in mixed-modes signal processing which have a variety of applications outside packaging such as in advance control, noise filtering and stochastic analysis. Nonstationary, transient and harmonic components are

already well studied in other domains than distribution packaging, but they are rarely considered altogether. For instance, in radio-frequency signal processing, harmonic components are extracted from noise (Gaussian components) but nonstationary and transient components are not present in the signal. Therefore, identifying these three different modes in the same signal will be new in signal processing.

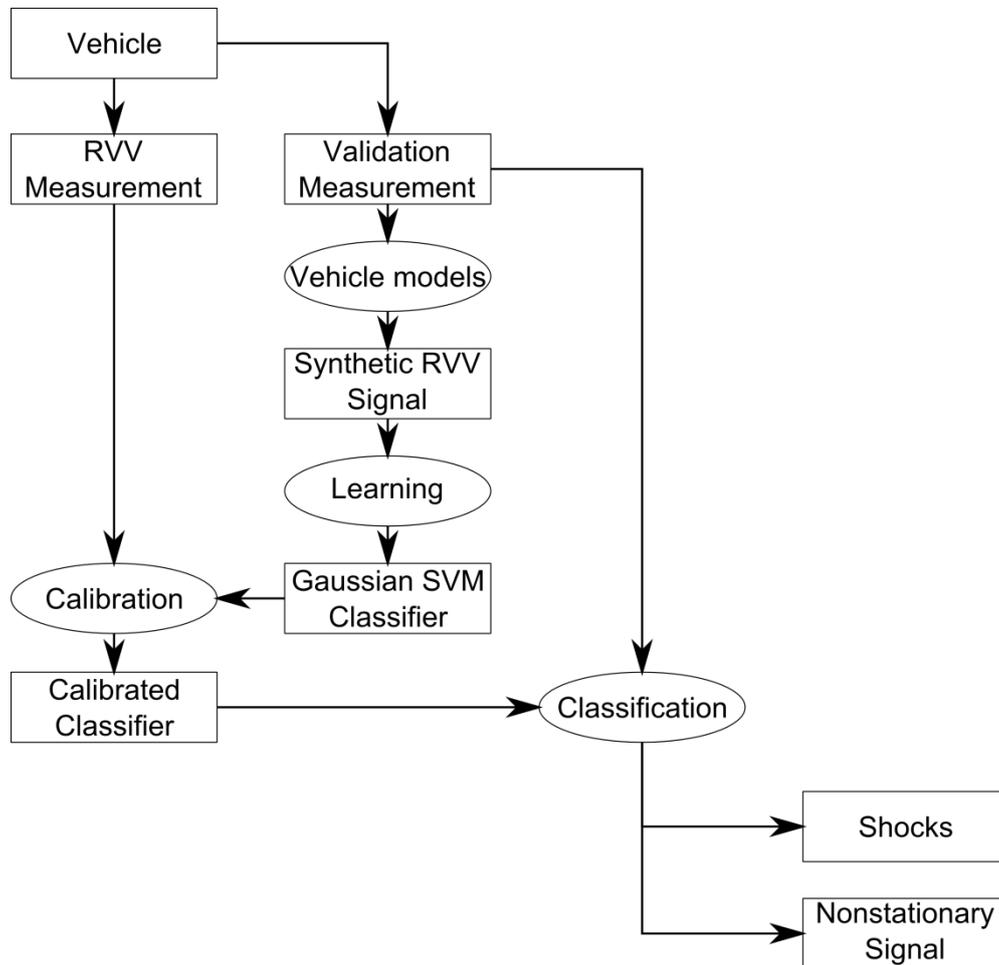


Figure 108: flow chart on machine learning classification implementation to real vehicle

Another outcome of the thesis is a better understanding of the RVV. Nowadays, there is very little information on the number and severity of the impacts incurred in typical journey because most of the tools used to analyse road degradation and roughness, such as the International Roughness Index, do not specifically take into account transient events (Bruscella et al., 1999). The relative importance of the harmonic components with the rest of the signal remains also unknown and only Charles (1993) pointed out this issue without really developing on it. The mixed modes model will provide answers to this gap which could have an important effect on packaging optimisation as other applications such as vehicle dynamic, passengers' comfort and road maintenance.

REFERENCES

- AINALIS, D. T. 2014. *Estimating the Dynamic Characteristics of Road Vehicles Using Vibration Response Data*. Victoria University.
- APC 2015. Australia's Packaging recycling Rate. Australian Packaging Covenant.
- ASTM-D999 2008. Standard Test Methods for Vibration Testing of Shipping Containers. West Conshohocken, PA: ASTM International.
- ASTM-D3580 2010. Test Methods for Vibration (Vertical Linear Motion) Test of Products. West Conshohocken, PA: ASTM Standard.
- ASTM-D4728 2012. Test Method for Random Vibration Testing of Shipping Containers. West Conshohocken, PA: ASTM International.
- AUSTROADS 2005. RoadFacts 2005: An Overview of the Australian and New Zealand Road Systems. Sydney: Austroads.
- AYENU-PRAH, A. & ATTOH-OKINE, N. 2009. Comparative study of Hilbert–Huang transform, Fourier transform and wavelet transform in pavement profile analysis. *Vehicle System Dynamics*, 47, 437-456.
- BENDAT, J. S. & PIERSOL, A. G. 2011. *Random data: analysis and measurement procedures*, Hoboken, New Jersey, USA, John Wiley & Sons.
- BOX, G. E. & COX, D. R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252.
- BRADLEY, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30, 1145-1159.
- BRUSCELLA, B. 1997. *The analysis and simulation of the spectral and statistical properties of road roughness for package performance testing*. Victoria University of Technology.
- BRUSCELLA, B., ROUILLARD, V. & SEK, M. 1999. Analysis of road surface profiles. *Journal of Transportation Engineering-Asce*, 125, 55-59.
- BUREAU OF INFRASTRUCTURE, T. A. R. E. B. 2010. Road freight estimates and forecasts in Australia: interstate, capital cities and rest of state. Canberra ACT.
- CARLSTEIN, E. G., MULLER, H.-G. & SIEGMUND, D. Change-point problems. 1994 1994 Hayward, California, USA. Institute of Mathematical Statistics.

- CEBON, D. 1989. Vehicle-generated road damage: a review. *Vehicle system dynamics*, 18, 107-150.
- CEBON, D. 1999. *Handbook of vehicle-road interaction / David Cebon*, Lisse ; Abingdon : Swets & Zeitlinger.
- CHARLES, D. 1993. Derivation of Environment Descriptions and Test Severities from Measured Road Transportation Data. *Journal of the IES*, 36, 37-42.
- CHEN, X., WANG, M., ZHANG, Y., FENG, Y., WU, Z. & HUANG, N. E. 2013. Detecting Signals from Data with Noise: Theory and Applications. *Journal of the Atmospheric Sciences*, 70, 1489-1504.
- CHERKASSKY, V. & MULIER, F. M. 2007. *Learning from data: concepts, theory, and methods*, New Jersey, USA, John Wiley & Sons.
- CHONHENCHOB, V., SINGH, S. P., SINGH, J. J., STALLINGS, J. & GREWAL, G. 2012. Measurement and Analysis of Vehicle Vibration for Delivering Packages in Small-Sized and Medium-Sized Trucks and Automobiles. *Packaging Technology and Science*, 25, 31-38.
- CONSULTING, O. 2015. Global Protective Packaging Market - By Regions and Vendors - Market Trends and Forecasts (2014 - 2020).
- FAWCETT, T. 2004. ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31, 1-38.
- FLANDRIN, P., RILLING, G. & GONCALVES, P. 2004. Empirical mode decomposition as a filter bank. *Signal Processing Letters, IEEE*, 11, 112-114.
- FOX, T. 2013. Global food: waste not, want not. London: Institute of Mechanical Engineers.
- GARCIA-ROMEU-MARTINEZ, M.-A., ROUILLARD, V. & CLOQUELL-BALLESTER, V.-A. A Model for the Statistical Distribution of Road Vehicle Vibrations. World Congress on Engineering, 2007 2007 London, UK. Citeseer, 1225-1230.
- GARCIA-ROMEU-MARTINEZ, M.-A., SINGH, S. P. & CLOQUELL-BALLESTER, V.-A. 2008. Measurement and analysis of vibration levels for truck transport in Spain as a function of payload, suspension and speed. *Packaging Technology and Science*, 21, 439-451.
- GILLESPIE, T. D. 1985. Heavy truck ride. SAE Technical Paper.
- GRIFFITHS, K. R. 2012. *An Improved Method for Simulation of Vehicle Vibration Using a Journey Database and Wavelet Analysis for the Pre-Distribution Testing of Packaging*. Dissertation/Thesis, University of Bath.
- HARRIS, C. M., PIERSOL, A. G. & PAEZ, T. L. 2002. *Harris' shock and vibration handbook*, McGraw-Hill New York.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. & FRANKLIN, J. 2005. *The elements of statistical learning: data mining, inference and prediction*, New York, USA, Springer.

- HOSOYAMA, A., SAITO, K. & NAKAJIMAM, T. Non-Gaussian Random Vibrations Using Kurtosis. Eighteenth IAPRI World Packaging Conference, 2012 San Luis Obispo (CA), USA. DEStech Publications, Inc, 35.
- HUANG, N. E. & SHEN, S. S. 2005. *Hilbert-Huang transform and its applications*, Singapore, World Scientific.
- HUANG, N. E., SHEN, Z. & LONG, S. R. 1999. A new view of nonlinear water waves: The Hilbert Spectrum I. *Annual review of fluid mechanics*, 31, 417-457.
- HUANG, N. E., SHEN, Z., LONG, S. R., WU, M. C., SHIH, H. H., ZHENG, Q., YEN, N.-C., TUNG, C. C. & LIU, H. H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454, 903-995.
- HUANG, N. E., WU, M.-L. C., LONG, S. R., SHEN, S. S., QU, W., GLOERSEN, P. & FAN, K. L. 2003a. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459, 2317-2345.
- HUANG, N. E., WU, M. L., QU, W., LONG, S. R. & SHEN, S. S. 2003b. Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Applied stochastic models in business and industry*, 19, 245-268.
- HUANG, N. E. & WU, Z. 2008. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46.
- HUBBARD, B. B. 1998. *The World According to Wavelets The Story of a Mathematical Technique in the Making*, Universities Press.
- IEC-60068 2007. Environmental testing – Part 2-6: Tests – Test Fc: Vibration (sinusoidal). Geneva, Switzerland: International Electrotechnical Commission.
- ISO-8608 1995. International Standard 8608: Mechanical Vibration – Road Surface Profiles – Reporting of Measured Data. Geneva: International Organisation for Standardisation.
- ISO-8608 2011. Mechanical vibration — Road surface profiles — Reporting of measured data. Geneva, Switzerland: ISO standard.
- ISO-13355 2002. Packaging. Complete, filled transport packages and unit loads. Vertical random vibration test. London, UK: BSI.
- ISO-13749 2011. Railway applications. Wheelsets and bogies. Method of specifying the structural requirements of bogie frames. Geneva, Switzerland: ISO standard.
- ISTA. *Test Series* [Online]. Michigan: International Safe Transit Association. Available: <https://www.ista.org/pages/procedures/ista-procedures.php> [Accessed 2016-10-05 2016].
- JANIC, M. 2007. Modelling the full costs of an intermodal and road freight transport network. *Transportation Research Part D: Transport and Environment*, 12, 33-44.
- KINANI, A. E. & OUDADESS, M. 2010. *Distribution theory and applications*, Hackensack, NJ : World Scientific.

- KIPP, W. I. 2000a. Developments in testing products for distribution. *Packaging Technology and Science*, 13, 89-98.
- KIPP, W. I. 2000b. Vibration testing equivalence. *International Safe Transit Association, East Lansing MI, USA. Paper presented at ISTACON.*
- KIPP, W. I. ACCELERATED RANDOM VIBRATION WITH TIME-HISTORY SHOCK. IoPP 2001 Annual Membership Meeting, 2001 2001 San Jose, California.
- KIPP, W. I. Random Vibration Testing Of Packaged-Products: Considerations For Methodology Improvement. 16th IAPRI World Conference on Packaging, 2008 2008 Bangkok, Thailand.
- LEE, S.-K. & SON, C.-Y. 2001. Ride comfort analysis of a vehicle based on continuous wavelet transform. *KSME International Journal*, 15, 535-543.
- LEE, S. & WHITE, P. 2000. Application of wavelet analysis to the impact harshness of a vehicle. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 214, 1331-1338.
- LEPINE, J., ROUILLARD, V. & SEK, M. 2014. Using the Hilbert-Huang Transform to Identify Harmonics and Transients within Random Signals. In: AUSTRALIA, E. (ed.) *Australasian Congress on Applied Mechanics*. Melbourne.
- LEPINE, J., SEK, M. & ROUILLARD, V. Mixed-Mode Signal Detection of Road Vehicle Vibration Using Hilbert-Huang Transform. ASME 2015 Noise Control and Acoustics Division Conference at InterNoise 2015, 2015. American Society of Mechanical Engineers, V001T01A002-V001T01A002.
- LING, C. X., HUANG, J. & ZHANG, H. 2003. AUC: a better measure than accuracy in comparing learning algorithms. *Advances in Artificial Intelligence*. Springer.
- LONG, M. 2016. *On the Statistical Correlation between Heave, Pitch and Roll Motion of Road Transport Vehicles*. PhD, Victoria University.
- LU, F., ISHIKAWA, Y., KITAZAWA, H. & SATAKE, T. 2010. Effect of vehicle speed on shock and vibration levels in truck transport. *Packaging Technology and Science*, 23, 101-109.
- LU, F., ISHIKAWA, Y., SHIINA, T. & SATAKE, T. 2008. Analysis of Shock and Vibration in Truck Transport in Japan. *Packaging Technology and Science*, 21, 479-489.
- MAHAJERIN, E. & BURGESS, G. Investigation of package vibration during the repetitive shock test. Proceedings of the World Congress on Engineering, 2010.
- MANDIC, D. 2011. Filter bank property of multivariate empirical mode decomposition. *Signal Processing, IEEE Transactions on*, 59, 2421-2426.
- MAO, C., JIANG, Y., WANG, D., CHEN, X. & TAO, J. 2015. Modeling and simulation of non-stationary vehicle vibration signals based on Hilbert spectrum. *Mechanical Systems and Signal Processing*, 50, 56-69.
- MIL-STD-810F 2000. Environmental Test Methods and Engineering Guides. USA: DEPARTMENT OF DEFENSE TEST METHOD STANDARD.

- MILLIKEN, P., DE PONT, J., MUELLER, T. & LATTO, D. 2001. ASSESSING ROAD-FRIENDLY SUSPENSIONS. *TRANSFUND NEW ZEALAND RESEARCH REPORT*.
- NEI, D., NAKAMURA, N., ROY, P., ORIKASA, T., ISHIKAWA, Y., KITAZAWA, H. & SHIINA, T. 2008. Wavelet analysis of shock and vibration on the truck bed. *Packaging Technology and Science*, 21, 491-499.
- NEWLAND, D. E. 2012. *An introduction to random vibrations, spectral & wavelet analysis*, Courier Dover Publications.
- OESTERGAARD, S. Packaging Goals in Transport Quality. 7th IAPRI World Conference on Packaging Technology and Science, 1991 Utrecht, the Netherlands.
- OGATA, K. 1995. *Discrete-time control systems*, Prentice Hall Englewood Cliffs, NJ.
- OTARI, S., ODOF, S., NOLOT, J. B., VASSEUR, P., PELLOT, J., KRAJKA, N. & ERRE, D. 2011. Statistical Characterization of Acceleration Levels of Random Vibrations during Transport. *Packaging Technology and Science*, 24, 177-188.
- PARSONS, S. 2007. Encyclopedia of Measurement and Statistics. *Reference & User Services Quarterly*, 46, 87-88.
- PENG, Z., TSE, P. W. & CHU, F. 2005. An improved Hilbert–Huang transform and its application in vibration signal analysis. *Journal of Sound and Vibration*, 286, 187-205.
- PLANCHEREL, M. 1933. Sur les formules de réciprocité du type de Fourier. *Journal of the London Mathematical Society*, 1, 220-226.
- POWERS, D. M. W. The problem of Area Under the Curve. Information Science and Technology (ICIST), 2012 International Conference on, 23-25 March 2012. 567-573.
- RICHARDS, D. P. 1990. Review of analysis and assessment methodologies for road transportation vibration and shock data. *Environmental Engineering*, 3, 23-26.
- RILLING, G. & FLANDRIN, P. on the Influence of Sampling on the Empirical Mode Decomposition. ICASSP (3), 2006. 444-447.
- ROGERS, S. & GIROLAMI, M. 2011. *A first course in machine learning*, Florida, USA, CRC Press.
- ROOT, D. 1997. *Six Step Method For Cushioned Package Development* [Online]. Monterey, CA, USA: Lansmont Corporation. Available: www.lansmont.com [Accessed 8 april 2014 2014].
- ROUILLARD, V. 2007a. *On the synthesis of non-Gaussian road vehicle vibrations*. PhD Thesis, Monash University.
- ROUILLARD, V. The synthesis of road vehicle vibrations based on the statistical distribution of segment lengths. Proceedings of the 5th Australasian Congress on Applied Mechanics, 2007 2007b Brisbane, Australia. Engineers Australia, 614.
- ROUILLARD, V. 2008. Generating Road Vibration Test Schedules from Pavement Profiles for Packaging Optimization. *Packaging Technology and Science*, 21, 501-514.

- ROUILLARD, V. 2014. Quantifying the Non-stationarity of Vehicle Vibrations with the Run Test. *Packaging Technology and Science*, 27, 203-219.
- ROUILLARD, V., BRUSCELLA, B. & SEK, M. 2000. Classification of road surface profiles. *Journal of Transportation Engineering-Asce*, 126, 41-45.
- ROUILLARD, V. & LAMB, M. 2008. On the Effects of Sampling Parameters When Surveying Distribution Vibrations. *Packaging Technology and Science*, 21, 467-477.
- ROUILLARD, V. & RICHMOND, R. 2007. A novel approach to analysing and simulating railcar shock and vibrations. *Packaging Technology and Science*, 20, 17-26.
- ROUILLARD, V. & SEK, M. 2013. Creating Transport Vibration Simulation Profiles from Vehicle and Road Characteristics. *Packaging Technology and Science*, 26, 82-95.
- ROUILLARD, V. & SEK, M. A. 2000. Monitoring and simulating non-stationary vibrations for package optimization. *Packaging Technology and Science*, 13, 149-156.
- ROUILLARD, V. & SEK, M. A. 2002. A statistical model for longitudinal road topography. *Road and Transport Research*, 11, 17-23.
- ROUILLARD, V. & SEK, M. A. 2005. The use of intrinsic mode functions to characterize shock and vibration in the distribution environment. *Packaging Technology and Science*, 18, 39-51.
- ROUILLARD, V. & SEK, M. A. 2010. Synthesizing Nonstationary, Non-Gaussian Random Vibrations. *Packaging Technology and Science*, 23, 423-439.
- SEIFFERT, C., KHOSHGOFTAAR, T. M., VAN HULSE, J. & NAPOLITANO, A. RUSBoost: improving classification performance when training data is skewed. *Pattern Recognition*, 2008. ICPR 2008. 19th International Conference on, 2008. IEEE, 1-4.
- SEK, M. A. Optimisation of Packaging Design Through an Integrated Approach to the Measurement and Laboratory Simulation of Transportation Hazards. 12th International Conference on Packaging, 2001 Warsaw, Poland.
- SHALEV-SHWARTZ, S. & BEN-DAVID, S. 2014. *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.
- SHIRES, D. 2011. On the Time Compression (Test Acceleration) of Broadband Random Vibration Tests. *Packaging Technology and Science*, 24, 75-87.
- SINGH, J., SINGH, S. P. & JONESON, E. 2006. Measurement and analysis of US truck vibration for leaf spring and air ride suspensions, and development of tests to simulate these conditions. *Packaging Technology and Science*, 19, 309-323.
- SMALLWOOD, D. O. 2005. Generating non-Gaussian vibration for testing purposes. *Sound and Vibration*, 39, 18-24.
- STASZEWSKI, W. & GIACOMIN, J. Application of the wavelet based FRFs to the analysis of nonstationary vehicle data. *Proceedings-SPIE the International Society for Optical Engineering*, 1997. SPIE INTERNATIONAL SOCIETY FOR OPTICAL, 425-431.
- STEINWOLF, A. & CONNON III, W. H. 2005. Limitations of the Fourier transform for describing test course profiles. *Sound and Vibration*, 39, 12-17.

- SVANES, E., VOLD, M., MØLLER, H., PETTERSEN, M. K., LARSEN, H. & HANSSSEN, O. J. 2010. Sustainable packaging design: a holistic methodology for packaging design. *Packaging Technology and Science*, 23, 161-175.
- THOMAS, F. 2005. Automated Road Segmentation Using a Bayesian Algorithm. *Journal of Transportation Engineering-Asce*, 131, 591-598.
- ULRICH BRAUNMILLER, R. R., THOMAS TROST, TORSTEN SCHREIBER, BERTRAND LEGUE, THOMAS BLOME, 1999. Source Reduction by European Testing schedules (SRETS). Pfinztal, Germany: EC (European Commission) Standards.
- VAN BAREN, P. 2005. The missing knob on your random vibration controller. *Sound and Vibration*, 39, 10.
- VAN BAREN, P. D. 2008. *System and method for simultaneously controlling spectrum and kurtosis of a random vibration*. USA patent application.
- WALLIN, B. 2007. Developing a Random Vibration Profile Standard. *International Safe Transit Association: East Lansing, MI*.
- WEI, L., FWA, T. & ZHE, Z. 2005. Wavelet analysis and interpretation of road roughness. *Journal of Transportation Engineering-Asce*, 131, 120-130.
- WOLFSTEINER, P. & BREUER, W. 2013. Fatigue assessment of vibrating rail vehicle bogie components under non-Gaussian random excitations using power spectral densities. *Journal of Sound and Vibration*, 332, 5867-5882.
- WORLD ECONOMICS JOURNAL 2016. World Economics Global GDP Data. London, UK: World Economics.
- WU, Z. & HUANG, N. E. 2004. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460, 1597-1611.
- WU, Z. & HUANG, N. E. 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1, 1-41.

APPENDIX A: HILBERT-HUANG TRANSFORM

Developed in the late 90s by Huang et al. (1998), the Hilbert-Huang Transform (HHT) is a fully adaptive time-frequency analysis method that requires very few analysis parameter settings. Based on the Hilbert transform, the HHT is a very powerful tool to analyse nonlinear and nonstationary processes using a sum of narrow-band functions of varying instantaneous frequency and amplitude. However, this method does not yet have a strong theoretical foundation, this being the cost of being a highly adaptive data-analysis method. In that context, seven mathematical problems were stated to establish the theoretical framework of the HHT (Huang and Shen, 2005):

1. adaptive data analysis methodologies in general;
2. nonlinear system identification methods;
3. prediction problems for nonstationary processes (end effects);
4. spline problems (best spline implementation for the HHT, convergence and 2-D);
5. optimization problems (the best IMF selection and uniqueness);
6. approximation problems (Hilbert transform and quadrature);
7. miscellaneous questions concerning the HHT.

These problems might take many years to be resolved, though that does not mean the HHT cannot be used until they are fully resolved. The broadly-used Fourier analysis was in the same situation many years ago. Invented in 1807, it was only fully proven more than a hundred years later by Plancherel (1933).

In order to assist judgement on the suitability of the HHT and its limitations, the mechanics of this empirical transform and how it can be used to analyse nonstationary and noisy signals is presented here.

A.1. Hilbert Transform

One method to overcome the Gabor limit principle is to formulate a signal with a varying frequency and amplitude modulation. In that case the frequency becomes the instantaneous frequency as it describes the signal frequency at a specific moment. The instantaneous frequency of a signal $x(t)$ can be computed using its Hilbert transform $y(t)$ which is defined by the following convolution integral:

$$y(t) = \text{H}[x(t)] = \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t-\tau)} d\tau. \quad (\text{eq. 40})$$

The Hilbert transform can be interpreted as a 90° phase shift of the signal, *i.e.* the signal is transformed in the frequency domain where its phase is shifted by 90° and brought back in the time domain. Using the Hilbert transform, the analytic signal can be defined as

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}, \quad (\text{eq. 41})$$

where the magnitude function $a(t)$ represents the envelope of the original signal,

$$a(t) = \sqrt{x^2(t) + y^2(t)}, \quad (\text{eq. 42})$$

and $\theta(t)$ represents the instantaneous phase function of $x(t)$,

$$\theta(t) = \arctan\left(\frac{y(t)}{x(t)}\right). \quad (\text{eq. 43})$$

As the analytic signal $z(t)$ can be represented as a rotating vector, $a(t)e^{i\theta(t)}$, the time derivative of $\theta(t)$ corresponds to the instantaneous frequency of the vector. Thus the “instantaneous frequency”, $f_i(t)$, of $x(t)$ is given by

$$f_i(t) = \left(\frac{1}{2\pi}\right) \frac{d\theta(t)}{dt}. \quad (\text{eq. 44})$$

The instantaneous frequency seems to easily overcome issues defined by the Gabor limit, but there is an important limitation: the Hilbert transform only works on narrow-band signals. This means it only works with signals with positive peaks and negative troughs (*i.e.* the signal crosses zero in between each peak and trough, Figure 109 a) but does not work on broad-band signals (*e.g.* Figure 109 b) which are a much more common type of signal in real applications.

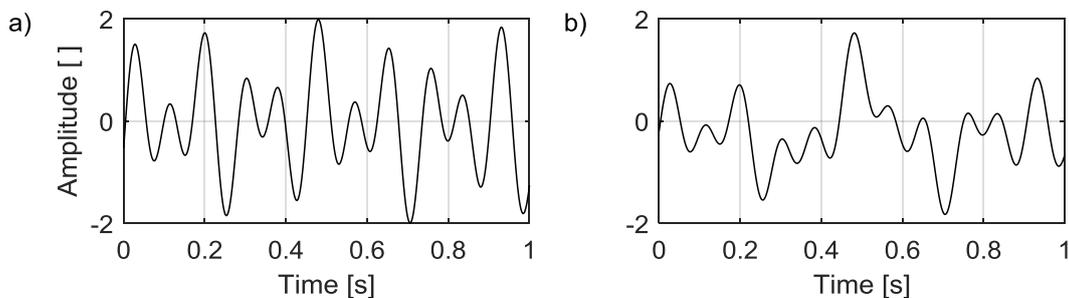


Figure 109: a) narrow-band signal; b) broad-band signal

A.2. Empirical Mode Decomposition

To extend the application of the Hilbert transform to any signal, Huang et al. (1998) proposed the Empirical Mode Decomposition (EMD) technique to decompose a broad-band signal into a sum of narrow-band signals. These narrow-band decompositions, called Intrinsic Mode Functions (IMFs), have the following definitions:

- (1) in the whole dataset, the number of extrema and number of zero-crossings must be either equal or differ by, at most, one;
- (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The IMFs are computed via a sifting process, the first step being to create the upper and lower envelope of the signal by connecting all peaks together and all the troughs together using cubic splines, as shown in Figure 110.

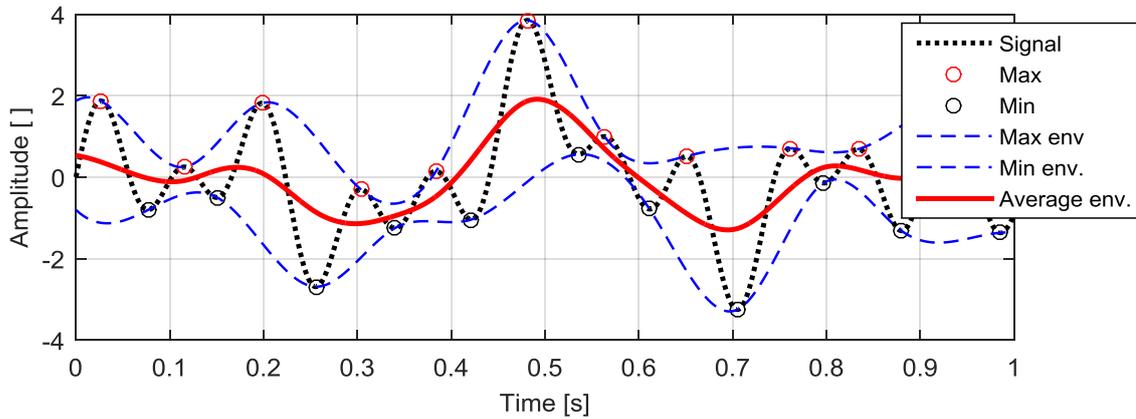


Figure 110: EMD sifting process

The mean value of the upper and lower splines designated m_1 is then subtracted from the signal, $x(t)$, to give the first component, h_1 , called a Proto Mode Function (PMF); *i.e.*,

$$h_1 = x(t) - m_1. \quad (\text{eq. 45})$$

One sift is usually not enough to create a PMF that conforms to the IMF definition so the sifting process is repeated up to k times as follows

$$h_{1k} = h_{1(k-1)} - m_{1k}, \quad (\text{eq. 46})$$

until the PMF reaches the stoppage criteria. Two criteria are typically used. The first one introduced by Huang et al. (1998) is a Cauchy convergence test. This test uses the normalised square difference between two successive siftings as

$$SD_k = \frac{\sum_{t=0}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^T h_{k-1}^2}. \quad (\text{eq. 47})$$

The squared difference SD_k has to be smaller than a predetermined value over the signal duration T . Huang and Shen (2005) propose that the value is typically between 0.2 and 0.3. However, this criterion does not ensure the first definition of the IMF on the number of extrema and zero-crossings. To correct this, Huang et al. (1999) introduced a second criterion on the agreement on the number of extrema and zero-crossings. According to this criterion, the sifting process should stop when the number of extrema and zero-crossings remain the same and are equal or differ, at most, by one for a predetermined S number of iterations. Studies on the EMD variance suggests a S number between 4 and 8 (Huang et al., 2003a).

Only one of these criteria is usually selected to stop the sifting process but secondary stoppage criteria could also be added depending of the specific need of the analysis, such as: a maximum sifting time to limit the computation time, or a period bandwidth limitation specific for every IMF to eliminate any mixed-mode IMF (Huang et al., 2003a).

Once the sifting process is stopped, the PMF becomes the first IMF designated as:

$$c_1 = h_{1k}. \quad (\text{eq. 48})$$

This first IMF, c_1 , contains the finest scale component of the signal, *i.e.* the component with the higher instantaneous frequencies. The residue, r_1 , is calculated by subtracting, c_1 , to the signal,

$$r_1 = x(t) - c_1. \quad (\text{eq. 49})$$

The sifting process is then performed again on r_1 until the second IMF, c_2 , is found. This procedure is repeated,

$$\begin{aligned} r_2 &= r_1 - c_2 \\ r_3 &= r_2 - c_3 \\ &\vdots \\ r_n &= r_{n-1} - c_n \end{aligned}, \quad (\text{eq. 50})$$

until the residue, r_n , becomes a monotonic function or a function with only one extremum, *i.e.* when no more IMFs can be extracted. This residue represents the trend of the signal. The sum of the n IMFs and the remaining trend r corresponds to the original signal, as shown in Figure 111,

$$x(t) = \sum_{j=1}^n c_j(t) + r(t). \quad (\text{eq. 51})$$

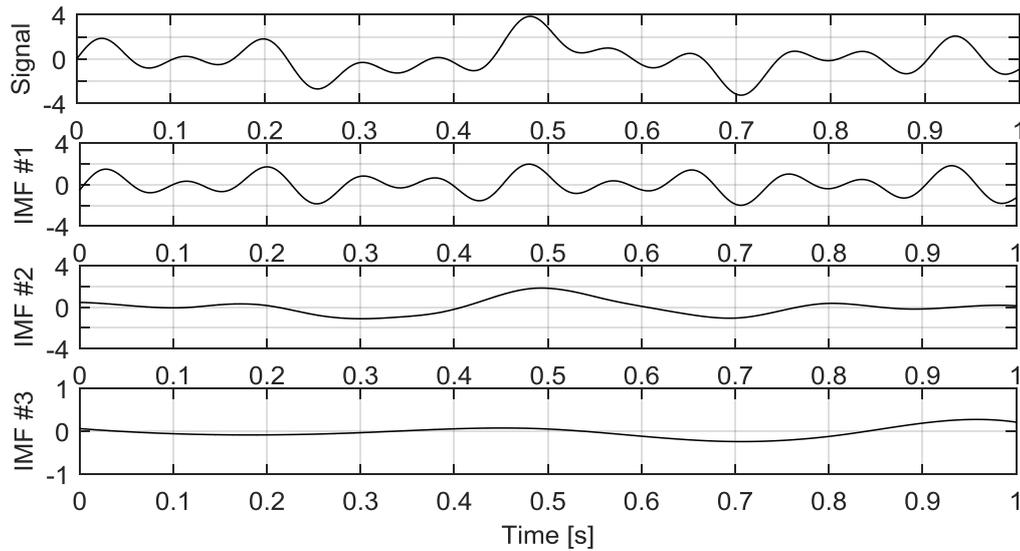


Figure 111: IMFs decomposition; the sum of the three IMFs is equal to the original signal

A.3. Time-Frequency Analysis tools

The analytic function (eq. 41) is computed on every IMF to provide the magnitude function (eq. 42) and the instantaneous frequency function (eq. 44). A common way to present these functions is the Hilbert spectrum, $H(\omega, t)$, which presents the instantaneous frequency and amplitude of every IMF as a function of time. The frequency information of the signal can also be expressed, regardless of time, as the Marginal spectrum

$$h(\omega) = \int_0^T H(\omega, t) dt. \quad (\text{eq. 52})$$

This spectrum gives the amplitude distribution of the frequencies, ω , which is different from the Fourier spectrum for a nonstationary signal. The Fourier spectrum gives the average amplitude of a persistent frequency, ω , when the Marginal spectrum gives the likelihood of a frequency component to exist. Therefore the Marginal spectrum has to be considered as a weighted nonnormalised joint amplitude-frequency-time distribution.

This joint distribution can also lead to the degree of nonstationarity, $DS(\omega)$, of the signal:

$$DS(\omega) = \frac{1}{T} \int_0^T \left(1 - \frac{H(\omega, t)}{n(\omega)} \right)^2 dt, \quad (\text{eq. 53})$$

where $n(\omega)$ is the mean Marginal spectrum,

$$n(\omega) = \frac{h(\omega)}{T}. \quad (\text{eq. 54})$$

When the signal is perfectly stationary, the Hilbert spectrum, $H(\omega, t)$, is not a function of time, so $DS(\omega)$ is zero. In that case, the Fourier spectrum makes physical sense, but otherwise it does not, because the spectral content is time variant.

A.4. Signal Detection and Noise Filtering

The HHT shows interesting proprieties when applied to broad-band noises. Flandrin et al. (2004) have shown in a numerical study on fractional Gaussian noise (*e.g.* white, pink, red, grey noises) that the EMD acts as a dyadic filter bank (or simply called filter bank). In other words, the average frequency of an IMF is twice the value of the following IMF. So the PDS of the first IMF mainly covers half of the spectrum frequency range, the second a quarter, the third an eighth and so forth (Figure 112).

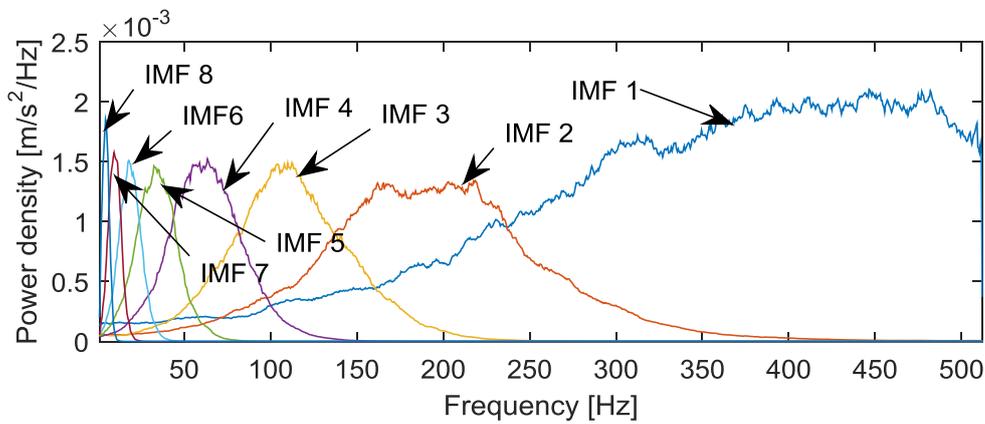


Figure 112: PDS of IMFs of white noise signal

The filter bank propriety of the EMD has an interesting application to signals affected by noise. In many applications, the measurement, $x(t)$, of a signal, $s(t)$, is spoiled by noise, $n(t)$, such as:

$$x(t) = s(t) + n(t). \quad (\text{eq. 55})$$

Several methods can be used to detect the presence of the signal and remove the noise from the measurement. However, most of them are based on linear and stationary hypotheses. Since the HHT and the EMD are not based on such hypotheses they can be useful where other methods are limited. The EMD can be used in two different ways to detect nonstationary and nonlinear signals and, ultimately, filter their noise component. The first one is a statistical significance test on the energy density and average period distribution of each IMF (Wu and Huang, 2004) and the second is a Fractional Resampling Technique (FRT) (Chen et al., 2013).

A.4.1 Energy Density/Average Period Significance Test

When applied to different white noise samples with the same sampling frequency, the EMD results in IMFs with a linear relationship between the average period, T , and energy density, E . A 1000-sample Monte Carlo simulation shows this relationship within a relatively small confidence interval (Figure 113). This relationship can be used to detect an IMF with a mode that does not belong to white noise (*i.e.* a signal) if its energy density and average period are outside the confidence interval (Wu and Huang, 2004).

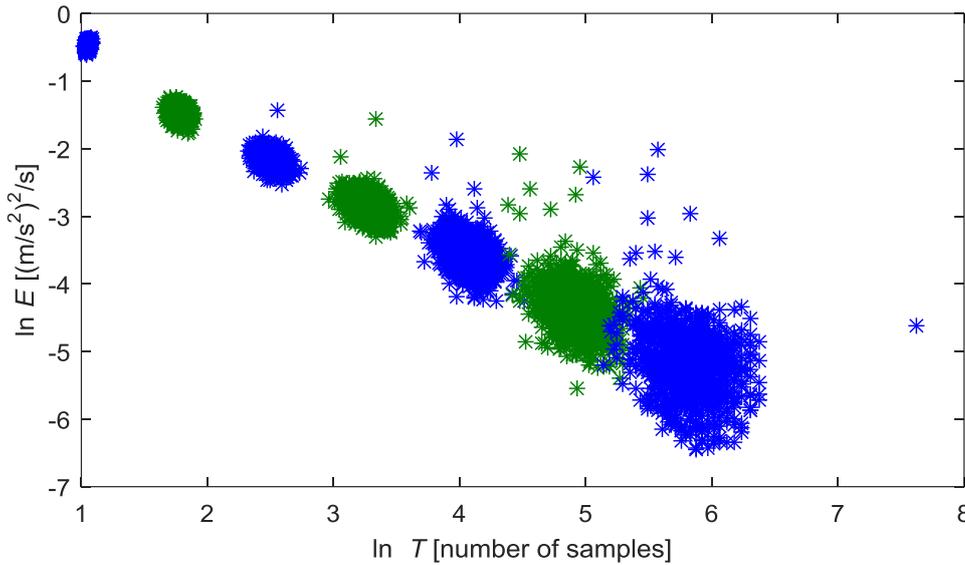


Figure 113: relationship between energy density with respect to the average period of white noise signals' IMFs. This Monte Carlo simulation represents 1000 samples. The IMFs are in ascending order starting from the left, the odd IMFs are in blue and the even in green.

A.4.2 Fractional Resampling Technique

The filter bank property of the EMD can be used as an alternate method to detect a signal buried in white noise; this is known as the Fractional Resampling Technique (FRT) (Chen et al., 2013). For fractional Gaussian noise, the average frequency of the k^{th} IMF can be calculated from the Marginal spectrum,

$$\bar{\omega}_k = \frac{\int h_k(\omega) \omega d\omega}{\int h_k(\omega) d\omega}, \quad (\text{eq. 56})$$

which depends on the sampling frequency of the signal. The average frequency of the first IMF is roughly 3/4 of the Nyquist frequency, that of the second is 3/8, the third 3/16 and so forth, because the EMD acts as a filter bank. When the noise is resampled by a factor between 1 and 2, the average frequencies gradually decrease as the Nyquist frequency decreases. At a resampling factor of 2, the average frequency of an IMF becomes the initial average frequency of the

subsequent IMF. This relation can be expressed by normalising the average frequency of the k^{th} IMF at the resampling rate M with the average frequency of the corresponding IMF at the initial resampling rate:

$$\tilde{\omega}_{Mk} = \frac{\bar{\omega}_{Mk}}{\bar{\omega}_{0k}} . \quad (\text{eq. 57})$$

As seen in Figure 114, for Gaussian noise, the average frequency of the IMFs have a linear relationship with the resampling rate. However, the same resampling behaviour is not observed on a signal superimposed with red noise^{viii} because the IMFs containing the signal retain the same average frequency over resampling. So the average frequency of every IMF will follow the resampling rate, except those containing the signal, which can be used to detect the signal.

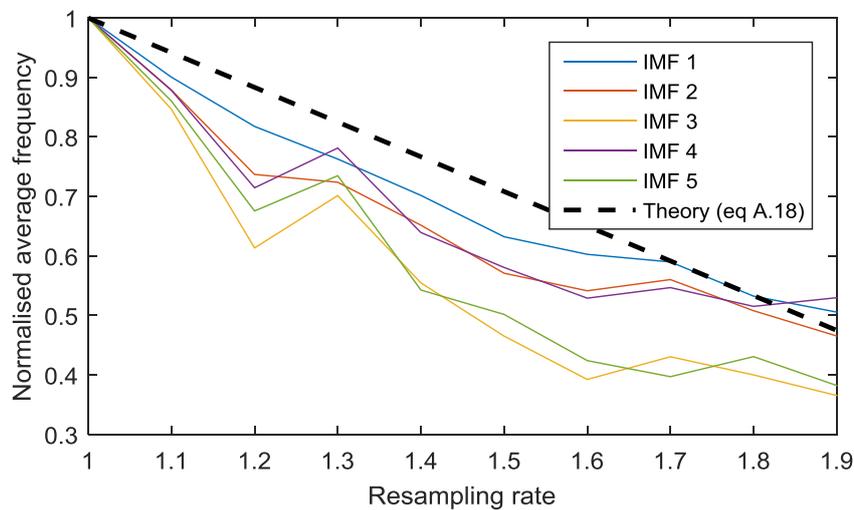


Figure 114: normalised average frequencies of IMFs as a function of the resampling rate for a red noise signal, with the dashed line representing the theoretical relationship

^{viii} Fractional Gaussian noise with PDS inversely proportional with the frequency squared

APPENDIX B: WAVELET TRANSFORM

The historical origin of the wavelet is difficult to define because it comes from several domains and was used under different names. There were at least 15 distinct roots of the theory going as far back as the 1930s (Hubbard, 1998). It is more in the 1970s and 1980s that wavelet researchers in different domains have begun to exchange information, and the technical vocabulary became more consistent. Wavelet then became a discipline of its own. It became more used and its development was accelerated. Today new wavelets can still be created but as is the Fourier analysis, the wavelet transform is at a mature state, its mathematical framework is fully proven, its computation is fast and it can be embedded into chipsets.

As described in the following sections, the wavelet transform is not strictly speaking a time-frequency analysis tool because it decomposes the signal not into frequencies, but into scales. Both are very similar with the exception that the scales do not necessarily have a physical meaning. Different techniques can be used to bring a physical meaning to the scales but these are only used in applications where frequency representation is required.

B.1. Filter Bank

There are many different ways to explain the basics of the wavelet transform. An easy way for engineers to approach it is to start with filter banks. As previously explained with the HHT, a filter bank is a series of filters that divide the signal into bandwidth sections. Essentially the first filter separates the signal into two sections; the second separates the low frequency sections into another two sections and so forth. The bandwidth proportion between the sections remains the same, *e.g.* the first is twice the frequency band of the second, the second twice the third and so forth.

The relationship between the sections' bandwidth are the same low pass, $h_0(n)$, and high pass, $h_1(n)$, filters are used throughout the filter bank analysis. The filter bandwidth changes between the iterations because the signal is first decimated after each filtering (Figure 115). In other words, at the first iteration, the original signal is filtered. Then the low and high pass sections are decimated. At the second iteration, the decimated low frequency section is filtered again

using the same filters and both resulting sections are decimated. This process is repeated until the desired scales (frequency bands) are obtained.

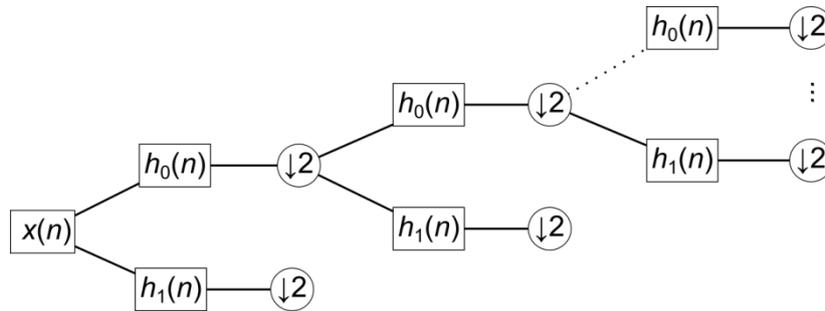


Figure 115: filter bank structure

B.2. Perfect Reconstruction

Certain filter bank structures have an interesting property. Coupled with the matching synthesis filters, the original signal can be perfectly reconstructed from the bandwidth decomposition even if the sections are decimated. Surprising as it may seem, this is also correct for the high pass filtered and decimated segmentation ($a(k)$, Figure 116). The frequency bandwidth of the signal is indeed above the Nyquist frequency, but since there are no low spectral contents in there, the synthesis filter, $f_1(n)$, can recover the original signal, $x(n)$, with a delay l .

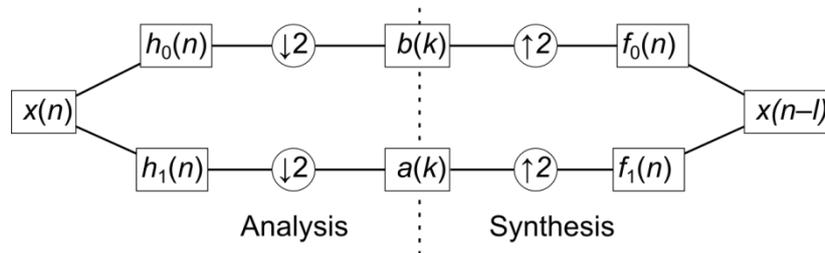


Figure 116: perfect reconstruction

The conditions needed to obtain the perfect reconstruction are twofold: the analysis/synthesis process does not create any distortion, and the aliasing caused by the down-sampling (decimation) has to be cancelled. To avoid distortion, the sum of the product in the z domain (see B.2.1) of the corresponding analysis and synthesis filters should have only one component as follows:

$$F_0(z)H_0(z) + F_1(z)H_1(z) = 2z^{-l} \quad , \quad \text{(eq. 58)}$$

where l represents a delay between the original signal and the synthesis signal.

The filtered segmentations are down-sampled which divides their Nyquist frequency by two. Because their frequency range is above Nyquist, an aliasing error is created. To perfectly

recreate the signal, this aliasing has to be cancelled by the synthesis. The condition for this aliasing cancellation can be formulated as:

$$F_0(z)H_0(-z) + F_1(z)H_0(-z) = 0 . \quad (\text{eq. 59})$$

By combining those two conditions, the relationship between analysis and synthesis filters is therefore:

$$\begin{aligned} F_0(z) &= H_1(-z) \\ F_1(z) &= -H_0(-z) \end{aligned} \quad (\text{eq. 60})$$

B.2.1 z Transform

The z transform is a discrete representation of the Fourier transform commonly used for filter design. The discrete Fourier transform of a signal $x(n)$ is:

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n) e^{-i\omega n} . \quad (\text{eq. 61})$$

To go from the frequency domain (Fourier) to the z domain, $e^{i\omega}$ is replaced by z . Therefore the z transform is:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n} . \quad (\text{eq. 62})$$

The advantage of this form of filter design is that z directly gives the filter coefficients that are used in the convolution (filtering process).

B.3. Multiresolution Synthesis

The perfect reconstruction process can be performed on a multiresolution scale (Figure 117). As the analysis progresses, the number of coefficients (a_i and b_i) is halved at every level i up to level l . To reconstruct the signal the coefficients are first upsampled and then filtered until the signal is completely reconstructed.

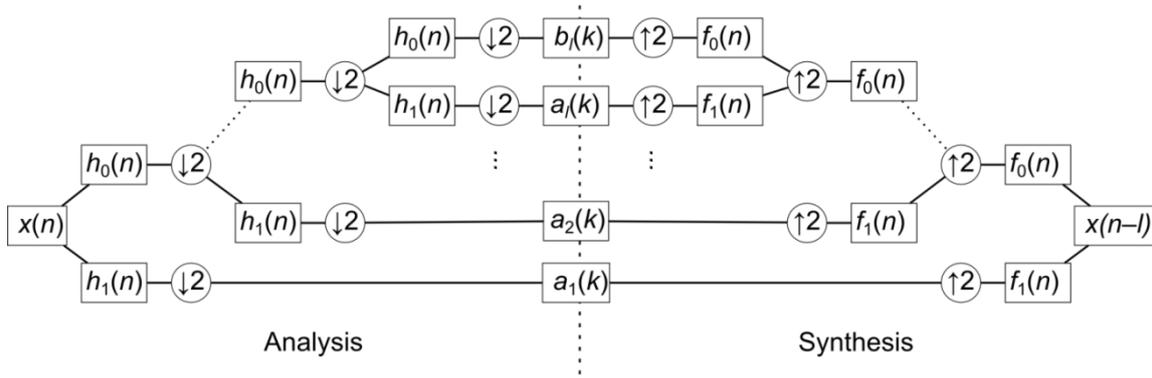


Figure 117: multiresolution synthesis

A signal can be encoded with a different resolution (scaling), such as the decimal system, when multiresolution synthesis respects these conditions:

- 1- The scaling function is orthogonal to its translations, *i.e.* the scalar product of the scaling function and any of its translations is zero.
- 2- The signal at a given resolution contains all the information of a signal at coarser resolution
- 3- The function 0 is the only object common to all the scaling function spaces V_i ,

$$\lim_{i \rightarrow \infty} V_i \cap V_i = \{0\} \tag{eq. 63}$$

- 4- Any signal can be approximated with arbitrary precision,

$$\lim_{i \rightarrow \infty} V_i = L^2(\mathbb{R}) \tag{eq. 64}$$

B.4. Wavelet Transform Formulation

Wavelet transform (in the continuous time domain) can be formulated from the filter bank analysis using the a_i and b_i coefficients, so the signal decomposition of the analysis part is the sum of the convolution:

$$x(t) = a_l(k)\phi(t-k) + \sum_{i=1}^l b_i(k)w(2^{l-i-1}t-k) . \tag{eq. 65}$$

ϕ is the scaling function or mother wavelet associated with the signal average (low pass filter) and defined such as its integral equals one:

$$\int \phi(t) dt = 1 . \tag{eq. 66}$$

The scaling function is in the Hilbert space denoted by L^2 where all functions have finite energy, such as:

$$\int |\phi(t)|^2 dt < \infty . \tag{eq. 67}$$

w is the wavelet function associated with the detail of the signal (high pass filter) and defined as its integral equals zero:

$$\int w(t)dt = 0 . \quad (\text{eq. 68})$$

The essential characteristic of the wavelet transform is that every scale comes from the scaling function and the wavelet function. This relationship is called the refinement equation, which also makes the link between the low and high pass filters of the filter bank analysis. The refinement or dilatation equation is for the scaling function (using normalised filter coefficients to signal energy constant):

$$\phi(t) = \sqrt{2} \sum h_0(k) \phi(2t-k) \quad (\text{eq. 69})$$

and for the wavelet function:

$$w(t) = \sqrt{2} \sum h_1(k) \phi(2t-k) , \quad (\text{eq. 70})$$

where the scaling function is normalised,

$$\int \phi dt = 1 . \quad (\text{eq. 71})$$

The scaling space V_i is all the combinations of $\phi(2^i t - k)$ and the wavelet space W_i is all the combinations of $w(2^i t - k)$. Considering the refinement equations means that the spaces of coarser scaling functions are a subset of the finer scaling functions. So only the empty space can be a subset of the coarser scaling space V_l and the finer scaling space V_l is a subset of L^2 ,

$$\{0\} \subset V_1 \subset V_2 \subset \dots \subset V_i \subset \dots \subset V_l \subset L^2 . \quad (\text{eq. 72})$$

Another way to represent a finer scaling space is to sum the coarser scaling and wavelet space,

$$V_i + W_i = V_{i+1} . \quad (\text{eq. 73})$$

By extension, any scale can be represented as:

$$\begin{aligned} V_1 + W_1 &= V_2 \text{ and } V_0 + W_0 = V_1 \\ V_0 + W_0 + W_1 &= V_2 \\ &\vdots \\ V_0 + \sum_{i=0}^{n-1} W_i &= V_n \end{aligned} \quad (\text{eq. 74})$$

B.5. Wavelets Creation

Those relationships can define any wavelet basis, from the simplest Haar wavelet to more recent and advanced wavelets. The wavelet creation process which starts from the scaling function can be simplified in 3 elements:

1. Scaling function is expressed in terms of low pass filter coefficients $h_0(k)$
2. The solution is defined in L^2 ,

$$\int |\phi|^2 dt = \int |\hat{\phi}|^2 d\omega < \infty \quad (\text{eq. 75})$$

3. The scaling function for N coefficients is computed from the following formula:

$$\phi(n) = \sum_{k=0}^N h_0 e^{-j\omega k} . \quad (\text{eq. 76})$$

This equation comes from the Fourier transform of the refinement equation (eq. 69),

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right), \quad (\text{eq. 77})$$

which can be extended to many scale:

$$\begin{aligned} \hat{\phi}(\omega) &= H\left(\frac{\omega}{2}\right) H\left(\frac{\omega}{4}\right) \hat{\phi}\left(\frac{\omega}{4}\right) \\ &\vdots \\ \hat{\phi}(\omega) &= \hat{\phi}(0) \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right) \end{aligned} \quad (\text{eq. 78})$$

and since the scaling function is normalised in the time domain (eq. 71),

$$\hat{\phi}(0) = 1, \quad (\text{eq. 79})$$

therefore the scaling function equation in the frequency domain is:

$$\hat{\phi}(\omega) = \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right). \quad (\text{eq. 80})$$

B.5.1 Haar Example

An interesting way to illustrate the wavelet analysis is going through the creation processes of the Haar Wavelet. The Haar wavelet is the simplest wavelet because it is created from the two normalised coefficients low pass filter,

$$h_0(n) = \begin{cases} 1/\sqrt{2} & \text{for } n = 0,1 \\ 0 & \text{anywhere else} \end{cases} \quad (\text{eq. 81})$$

The filter is expressed in the frequency domain as:

$$H_0(\omega) = \frac{1}{\sqrt{2}} + \frac{e^{-j\omega}}{\sqrt{2}} \quad (\text{eq. 82})$$

From equation (eq. 80), the scaling function is:

$$\phi(\omega) = \prod_{i=1}^{\infty} H_0\left(\frac{\omega}{2^i}\right) \quad (\text{eq. 83})$$

At $i \rightarrow \infty$, the scaling function can be expressed as:

$$\phi(\omega) = \frac{\sin(\pi\omega)}{\pi\omega} \quad (\text{eq. 84})$$

The inverse Fourier transform of this expression is:

$$\phi(t) = \begin{cases} 1 & \text{for } n = 0,1 \\ 0 & \text{anywhere else} \end{cases} \quad (\text{eq. 85})$$

The scaling function respects the refinement equation (eq. 69),

$$\phi(t) = \phi(2t) + \phi(2t-1) \quad (\text{eq. 86})$$

which is illustrated at Figure 118.

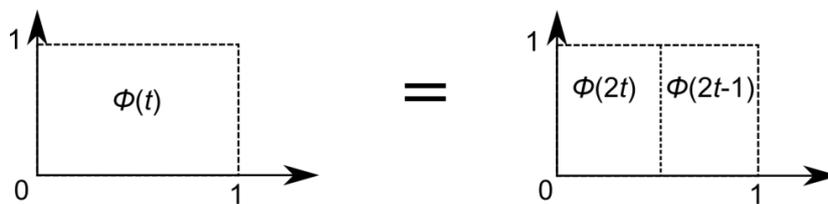


Figure 118: refinement equation of Haar scaling function

The high pass filter coefficients are calculated from equations (eq. 58) and (eq. 60),

$$\begin{aligned} 2z^{-l} &= H_1(-z)H_0(z) - H_0(-z)H_1(z) \\ 2z^{-l} &= (h_1(0) - h_1(1)z^{-1})\frac{1}{\sqrt{2}}(1+z^{-1}) - \frac{1}{\sqrt{2}}(1-z^{-1})(h_1(0) + h_1(1)z^{-1}) \quad (\text{eq. 87}) \\ \sqrt{2}z^{-l} &= (h_1(0) - h_1(1))z^{-1} \end{aligned}$$

In that case, the delay $l = 1$, so:

$$h_1(0) - h_1(1) = \sqrt{2} \tag{eq. 88}$$

Two normalised coefficients filters are used:

$$h_1(0)^2 + h_1(1)^2 = 1, \tag{eq. 89}$$

therefore,

$$\begin{aligned} h_1(0) &= \frac{1}{\sqrt{2}} \\ h_1(1) &= \frac{-1}{\sqrt{2}} \end{aligned} \tag{eq. 90}$$

These filter coefficients lead to the following wavelet function using the refinement equation (eq. 70):

$$w(t) = \phi(2t) - \phi(2t-1), \tag{eq. 91}$$

which is illustrated in Figure 119.

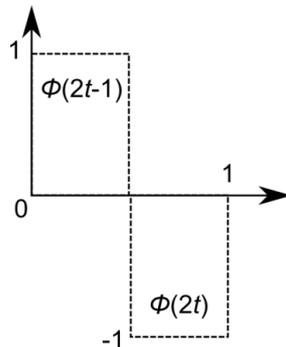


Figure 119: Haar wavelet function

As shown in Figure 120, the finer Haar scaling function, V_1 , is the sum of subsequent scaling and wavelet functions (V_0 and W_0) which is in respect with the multiresolution characteristic of wavelet transform. The area of the scaling functions remains 1 for any scales.

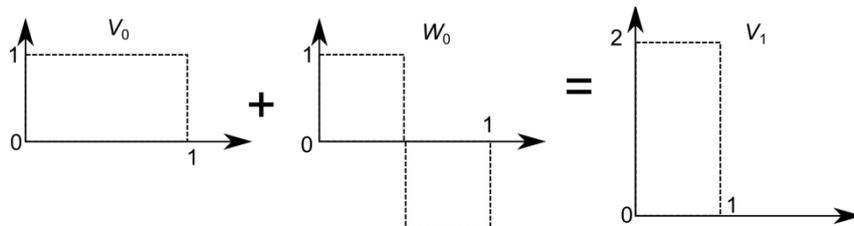


Figure 120: Haar scaling and wavelet functions multiresolution

B.6. Signal Decomposition

As stated by (eq. 65), any signal can be decomposed as a sum of scaling and wavelet functions at different scales. This is true with the Haar wavelet (Figure 121). Because only two filter coefficients are used, many scales are required to represent a smooth signal. This decomposition is equivalent to a digitalisation (binary decomposition). As the complexity and the number of filter coefficients increases, the shape of the wavelet becomes smooth and fewer scales are required to decompose a signal.

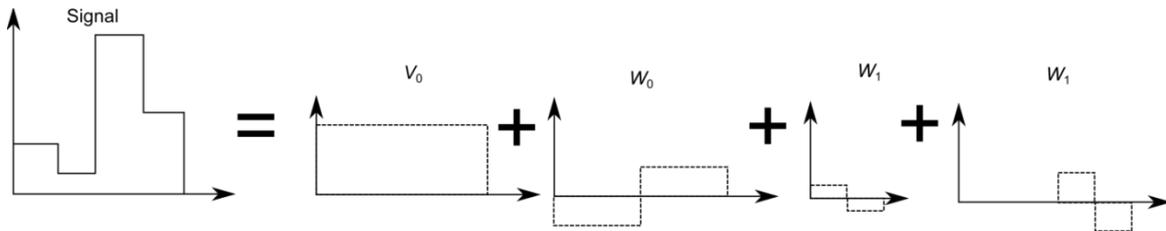


Figure 121 signal decomposition using the Haar wavelet

B.7. Wavelet Selection

There is no such thing such as a perfect wavelet. All wavelets have their benefits and disadvantages. For instance, the Haar wavelet has the shortest support (two filters coefficients) but a very coarse resolution. Therefore, the wavelet selection depends on the signal analysed. This selection is made according to three elements: the number of vanishing points, the filter length and the orthogonality.

B.7.1 Vanishing Points

The wavelet vanishing points (eq. 18) correspond to the k^{th} moments where the following integral of the function, $f(x)$, equals zero,

$$m_k = \int_{-\infty}^{\infty} f(x)x^k dx .$$

The number of vanishing points p of the wavelet function gives an indication of the precision of every scale. In other words, any degree- p polynomial can be exactly reproduced in every scale. The more vanishing points there are, the more concentrated is the signal information in the coarser scales, so fewer coefficients are needed to analysis singularities and discontinuities. More vanishing points means that the wavelet filters need longer support (more filter coefficients). For example, a Daubechies wavelet with p vanishing moments has $2p - 1$ filter coefficients.

B.7.2 Filter Length

The filter length has an effect on the processing time. This is critical for real time application, but not for RVV analysis, especially because the multiresolution makes the fast wavelet analysis faster than the FFT (Fast Fourier Transform).

The second advantage of a short filter is that it works best on short signals. Once again this advantage is worthless for RVV analysis because the signals are generally many times longer than the filter length. For instance, the Daubechies 10 wavelet only has 19 coefficients which is a long filter in the wavelet world. Compared to a RVV signal recorded at 1024 Hz, the filter length is less than 0.02 s which is very short in terms of road vehicle dynamics.

B.7.3 Orthogonality, Biorthogonality, Symmetry

Orthogonality is easily understood in two and three dimensional spaces. Two vectors are orthogonal in the two dimensional space if there is no projection possible of one vector on the other, *i.e.* they have a 90° angle between each other. The orthogonal vectors can be used as a basis in that dimension because there is no redundancy between the vectors. In other words, the sum of the energy of every component equals the total energy.

The wavelet function (analysis high pass filter) can be orthogonal with the scaling function (analysis low pass filter). Therefore, there is no redundancy between the wavelet scales. This means that the minimum number of coefficients is used to describe a signal. However, except for the Haar wavelet, orthogonality is incompatible with another wavelet property: symmetry which is important for application such as image processing. This leads to the creation of a biorthogonal wavelet where the analysis high pass filter is orthogonal with the synthesis low pass filter and the analysis low pass filter with the synthesis high pass filter.

For RVV analysis, symmetry is not necessarily needed, therefore the wavelet selection is not limited by the orthogonality type.

B.8. Signal Compression

Signal compression is nowadays the most common application of wavelet analysis. Compression is achieved using the wavelet coefficient a_i . It is assumed that within a scale, coefficients with larger values are the most important. So the coefficients below a certain threshold are discarded to compress the signal. The signal reconstruction is then made with the more valuable coefficients for every scale.

APPENDIX C: SAMPLING FREQUENCY CONDITIONING TO ENHANCE HHT PROCESSING

The Hilbert-Huang Transform (HHT) is a powerful adaptive time-frequency analysis tool. An essential step of this transform is the Empirical Mode Decomposition (EMD), which attempts to decompose the analysed signal into narrow-banded components called Intrinsic Mode Functions (IMFs). This appendix shows how the sampling frequency of signals affects the EMD. It is revealed, using both synthetic and measured data, that oversampled signals significantly increase the EMD computing time and the ability to retrieve a unique signal component in a single IMF. This appendix demonstrates how down-sampling a signal to match its spectral characteristic enhances the efficacy of the EMD.

C.1. Context

Developed in the late 1990s by Huang et al. (1998), the HHT is a fully adaptive processing time-frequency analysis method that requires very few analysis configuration parameters. Based on the Hilbert transform, the HHT is a very powerful tool to analyse nonlinear and nonstationary processes using a sum of narrow-band functions of varying instantaneous frequency and amplitude. These sought after characteristics make the HHT very attractive. Its potential has been demonstrated across many disciplines such as geophysics (Huang and Shen, 2005, Huang et al., 1999, Huang and Wu, 2008), structural health monitoring (Huang and Shen, 2005), finance (Huang et al., 2003b), image processing (Huang and Shen, 2005), structural vibration (Rouillard and Sek, 2005, Peng et al., 2005, Ayenu-Prah and Attoh-Okine, 2009) and many more.

Even if the HHT has a broad range of applications, it does not yet have a strong theoretical foundation. Therefore, guidelines on its usage are developed and published as the HHT is increasingly studied and understood. For instance there is little information as to the best sampling frequency to analyse signals even if it is known that behaviour of the EMD is affected by the signal sampling frequency (Chen et al., 2013, Flandrin et al., 2004, Mandic, 2011). Rilling and Flandrin (2006) showed that one performing EMD on signals with low sampling frequency relative to their frequency contents is subject to sampling error, *i.e.* the portion of the

signal that is lost in the sampling process. In some cases, this error for the first IMF can go up to 5 % when it has components with a frequency close to Nyquist frequency (half the sampling frequency). It is therefore recommended to oversample a signal to minimise this error, but there is no available information on the effect of oversampling on the EMD.

For many reasons, signals may be oversampled, *i.e.* recorded with a sampling frequency superior to the frequency bandwidth of interest. In shock and vibration analysis for instance, oversampled signals give better resolution around signals extrema which is useful for damping and natural frequency estimation in the time domain. However, this oversampling has a direct influence on the EMD and may require longer computing time and may also increase the mode-mixing between IMFs.

These shortcomings come from the sifting process used in the EMD. The sifting extracts the highest frequency component from the signal in the first IMF, and then the second highest component in second IMF and so on until there is only the signal trend left (see Appendix A for more detail on EMD). As Flandrin et al. (2004) have shown in a numerical study on fractional Gaussian noise (*e.g.* white, pink, red, grey noises), the EMD acts as a dyadic filter bank. In other words, the instantaneous frequency average of an IMF is twice the value of the following IMF. So the PDS of the first IMF mainly covers half of the spectrum frequency range, the second a quarter, the third an eighth and so forth (Figure 122).

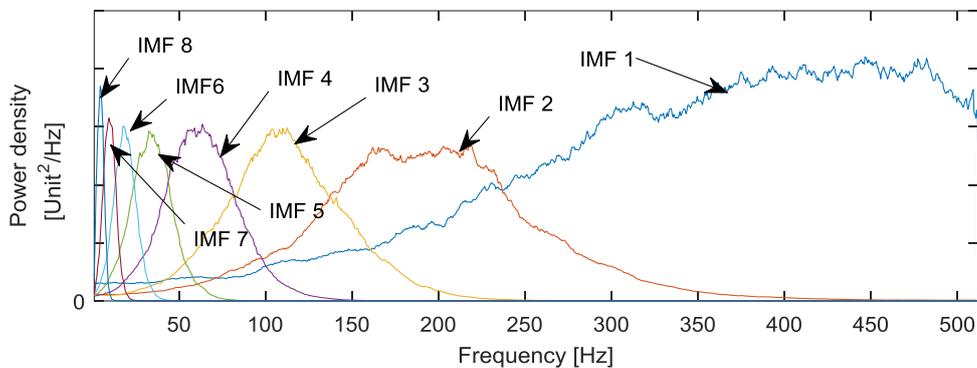


Figure 122: PDS of IMFs of white noise signal sampled at 1024 Hz (same as Figure 112)

C.2. Effect on Mode-Mixing

A limitation of the sifting process is that narrow-banded components buried in a low level of high-frequency content signal (such as measuring noise) cannot be easily extracted. This is because depending on the high-frequency content level of the signal, the first IMFs of an oversampled signal may include the narrow-banded component only, or a mix between the high-frequency content and narrow-banded signal. These different cases are demonstrated using

signals containing two harmonic components (narrow-banded) at 3.14 Hz and 10 Hz with the same amplitude. The signals have a sampling frequency of 1024 Hz and they are superimposed with white noise with an RMS level corresponding to 0.01 % (Figure 123), 0.1 % (Figure 124) and 1 % (Figure 125) of the harmonics amplitude.

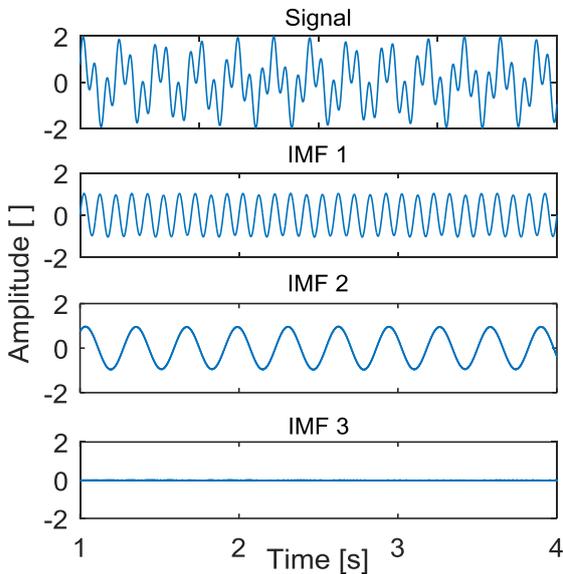


Figure 123: first three IMFs of a two harmonics signal with 0.01 % white noise with a sampling frequency of 1024 Hz

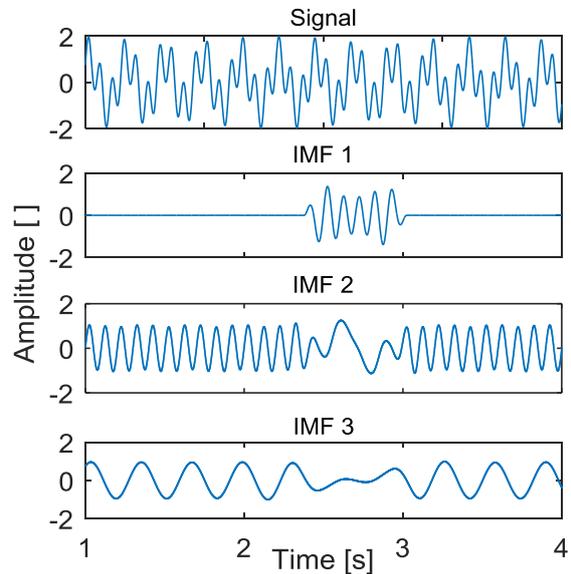


Figure 124: first three IMFs of a two harmonics signal with 0.1 % white noise with a sampling frequency of 1024 Hz

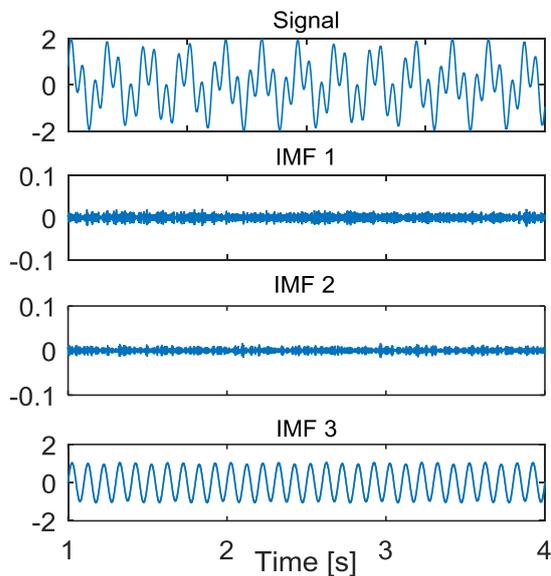


Figure 125: first three IMFs of a two harmonics signal with 1 % white noise with a sampling frequency of 1024 Hz

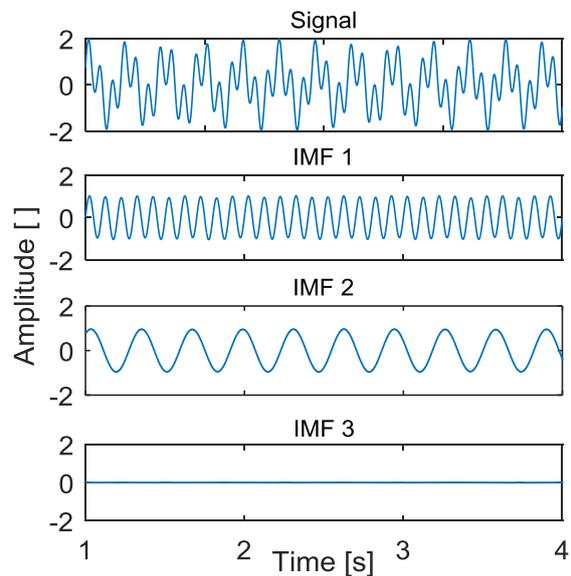


Figure 126: first three IMFs of a two harmonics signal with 0.1 % white noise with a sampling frequency of 256 Hz

The worst case in this example is the 0.1 % noise because the harmonic components do not remain in a single IMF as, in theory, they should. This phenomenon is called mode-mixing and can be prevented with different methods. Huang et al. (2003a) proposed a bandwidth limit as an EMD stoppage criterion to ensure that a single component remains in the same IMF. However this implies subjective user intervention that limits the fully adaptive nature of the EMD. Wu and Huang (2009) propose the ensemble EMD to avoid mode-mixing between IMFs to maintain its adaptiveness. This method adds sufficient white noise to the analysed signal to ensure that the IMFs bandwidths act as a filter bank by adding energy throughout the frequency bandwidth. This process is repeated several times and the IMFs are retrieved by averaging the ensemble. However, the averaging process does not guarantee that the mean IMFs respect the definition with regard to the number of extrema and zero-crossings. More detail on these mixed-mode mitigation methods are presented in section 6.1.3.

That being said, mode-mixing caused by signal oversampling can be easily avoided without using these exotic methods by simply reducing the sampling frequency to a range that closely matches the signal spectrum. For instance, mode-mixing between the IMFs of the 0.1 % white noise signal is eliminated when its sampling rate is reduced (resampled) to 256 Hz (Figure 126). The EMD process is also quicker as it does not have to deal with a large bandwidth noise.

C.3. Effect on Computing Time

Processing of oversampled signals takes more time because there are no clear components that fit the native frequency bandwidth of the first IMFs. The EMD process takes more siftings converging to narrow banded IMFs. This increase in computing speed is much more significant on larger data size. For instance, it takes 902 s for the EMD to retrieve the first 4 IMFs of a 200 s acceleration signal recorded on a road vehicle sampled at 2500 Hz using a PC with an Intel® Xeon® E5-1650 v2 (6 cores) 3.50 GHz CPU and 32 GB of RAM (Table 9). This signal is considered oversampled because its PDS drops off at about 60 dB within 300 Hz (Figure 127). This means that the sampling frequency can be reduced to 625 Hz and represents all the components bellow Nyquist (312.5 Hz) which is the highest frequency that can be represented in a signal according to the sampling theorem. It is also possible to use a low-pass filter prior to the resampling to avoid any aliasing. The EDM on the down-sampled signal is more than 100 times quicker. This is essentially caused by the reduction of EMD iterations needed to sift the first IMFs. The down-sampled signal needs about 20 times less iterations to retrieve the first four IMF than the original signal (Table 9).

Table 9: performance of the EMD for the first four IMFs of a signal with two sampling frequencies

Sampling Frequency [Hz]	EMD time (4 IMFs) [s]	IMF 1 iterations []	IMF 2 iterations []	IMF 3 iterations []	IMF 4 iterations []
2500	902	255	1563	3139	145
625	8.01	98	69	26	65

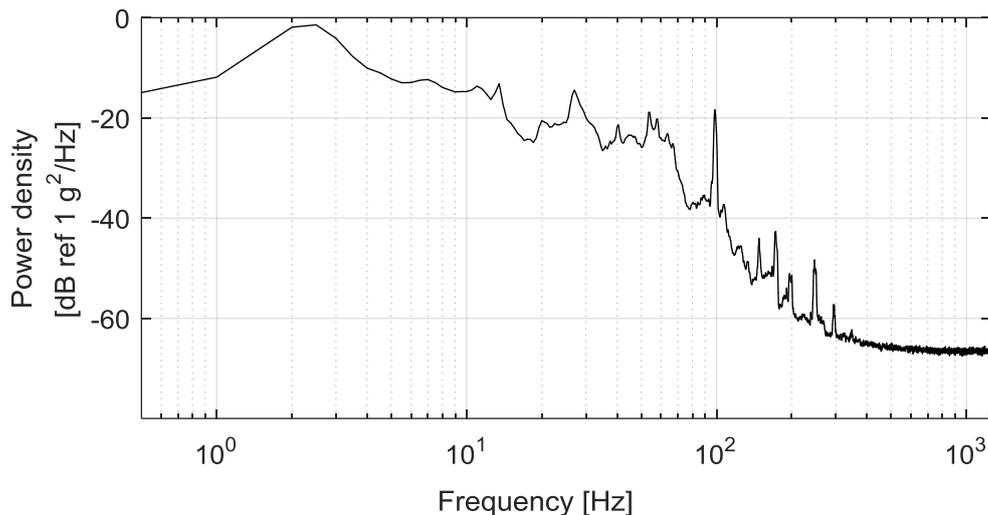


Figure 127: PDS of the acceleration recorded on a road vehicle

C.4. RVV Example

The first IMF of the down-sampled signal has a similar content to the subsequent IMF of the original signal. As shown in Figure 128, IMF 4 of the original signal is similar to IMF 1 of the down-sampled signal. The amplitude of IMFs 1 and 2 of the original signal is very low compared to IMF 4 and represents the noise in the signal. IMF 3 has a higher amplitude and its major components have a similar frequency to IMF 4, this suggests mode-mixing between them. It suggests that the IMF 4 of the original signal and IMF 1 of the down-sampled signal are the first narrow-banded component of the signal.

As shown by Rilling and Flandrin (2006), EMD of an undersampled signal may lead to a theoretical 5 % sampling error. However this theoretical error does not affect this undersampled signal. The difference between the down-sampled signal and signal reconstruction from the IMF sum is practically zero (RMS level of -300 dB). This is because its first IMF has an instantaneous frequency range mostly between 30 Hz and 200 Hz averaging at 90 Hz which is at least 1.5 times below the Nyquist frequency.

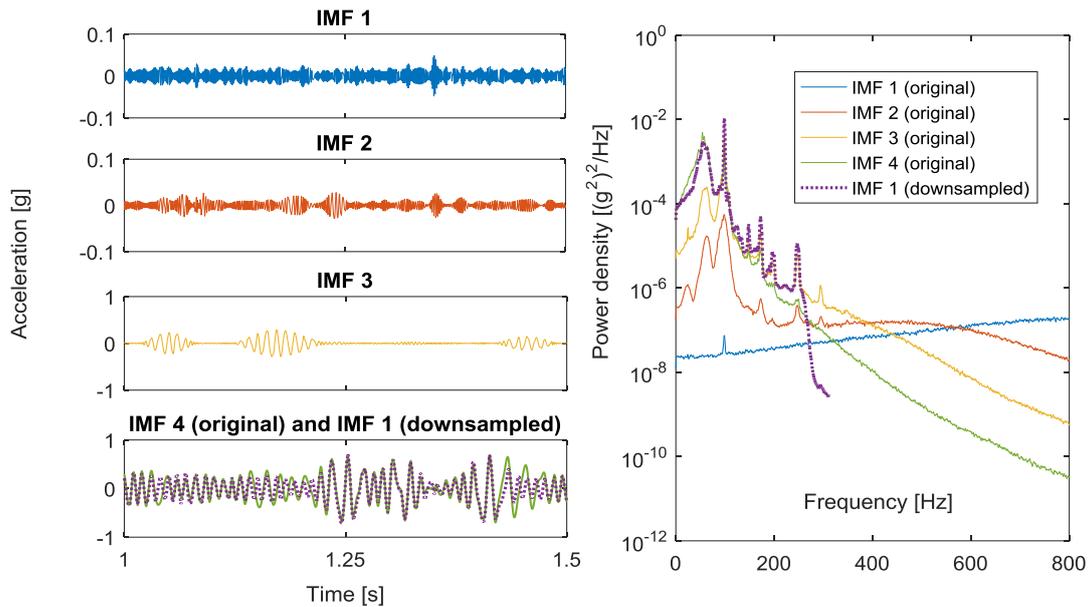


Figure 128: first four IMFs of a 200 s acceleration signal record and their PDS of a road vehicle sampled at 2500 Hz and first IMF of the same signal resampled at 625 Hz (bottom red dotted curve)

C.5. Conclusion on Sampling Frequency Conditioning

Oversampled signals have many advantages, their fine resolution gives better accuracy when analysing transient events or estimating natural frequency and damping of a structure using time domain techniques. However, oversampling adversely affects EDM and HHT analysis. Before using this time-frequency tool, it is therefore recommended to down-sample any signal to a sampling frequency that only covers the frequency bandwidth of interest with just enough range to avoid the small error that can be caused by undersampling. It was shown that resampling the signal with a Nyquist frequency 1.5 times higher than the expected highest frequency has all the benefits without creating any noticeable sampling error. This simple conditioning reduces the mode-mixing that can occur between IMFs. This also increases the EMD computing speed which is useful for real-time application.

APPENDIX D: BAYES' CRITERION

The Bayes' criterion finds the Optimal Operation Point (OOP) of a classifier by optimising the detection cost function with the Bayesian theory. For a binary detection problem, a classifier based its prediction on two hypotheses; the null hypothesis and the alternate hypothesis H_1 . There are four possible decision cases when testing these hypotheses:

1. decide H_0 when H_0 is true, called true rejection;
2. decide H_0 when H_1 is true, called miss detection;
3. decide H_1 when H_0 is true, called false detection;
4. decide H_1 when H_1 is true, called true detection.

The cost associated at each case which is respectively C_{00} , C_{01} , C_{10} and C_{11} . Given that $P(D_i, H_j)$ is the joint probability to decide D_i when H_j is true, the detection cost function is:

$$\mathfrak{R} = C_{00}P(D_0, H_0) + C_{01}P(D_0, H_1) + C_{10}P(D_1, H_0) + C_{11}P(D_1, H_1) . \quad (\text{eq. 92})$$

The equation can be simplified using Bayes' rule,

$$P(D_i, H_j) = P(D_i | H_j) P_j , \quad (\text{eq. 93})$$

where $P(D_i | H_j)$ is the conditional probability of deciding H_i when H_j is true and P_j is the *a priori* probability of H_j . For the four decision cases, the conditional probabilities are:

1. $P(D_0 | H_0)$;
2. $P(D_0 | H_1)$ also called probability of miss detection P_M ;
3. $P(D_1 | H_0)$ also called probability of false alarm P_F ;
4. $P(D_1 | H_1)$ also called probability of detection P_D .

Both hypotheses cannot coexist; hence the sum of their *a priori* probabilities equals one,

$$P_0 + P_1 = 1 . \quad (\text{eq. 94})$$

This also means that the conditional probabilities have the following relations:

$$\begin{aligned} P(D_0|H_0) &= 1 - P_F \\ P_M &= 1 - P_D \end{aligned} \quad (\text{eq. 95})$$

The detection cost function can be simplified using (eq. 93) and (eq. 95):

$$\mathfrak{R} = C_{00}(1 - P_F)P_0 + C_{01}(1 - P_D)P_1 + C_{10}P_F P_0 + C_{11}P_D P_1 \quad (\text{eq. 96})$$

The probabilities of detection and false detection are variables that depend on each other. The dependence is represented by the Receiver Operating Characteristic (ROC) curve where the classifier sensitivity represents the probability of detection P_D and the fall-out represents the probability of false detection P_F . The OOP ROC curve is therefore the point where the cost function is minimal,

$$\nabla \mathfrak{R} = \frac{\delta \mathfrak{R}}{\delta P_F} + \frac{\delta \mathfrak{R}}{\delta P_D} = 0 \quad (\text{eq. 97})$$

So the Bayes' criterion (eq. 33) defined the OOP as the point where the ROC curve derivative equals:

$$\frac{dP_D}{dP_F} = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

The case where $C_{00} = C_{11} = 0$ and $C_{10} = C_{01}$ is called MAP (Maximum *A Posteriori*) criterion which uses the ratio between the hypotheses *a priori* probabilities to find the OOP,

$$\frac{dP_D}{dP_F} = \frac{P_0}{P_1} \quad (\text{eq. 98})$$

APPENDIX E: MINIMAX CRITERION

The Bayes' criterion presented in the Appendix D is a powerful method to define the Optimal Operation Point (OOP) of a classifier. However, it cannot be used when the *a priori* probability of the classes P_0 and P_1 are unknown or could change once the classifier is designed. To overcome this issue, the OOP can be defined for the probability P_1^* where the detection cost function is maximum which means that the OOP would be really optimal for the worst case scenario. Minimising the maximal detection cost function is known as the minimax criterion.

As developed in Appendix D, the detection cost function is:

$$\mathfrak{R} = C_{00}(1 - P_F)P_0 + C_{01}(1 - P_D)P_1 + C_{10}P_F P_0 + C_{11}P_D P_1 \quad (\text{eq. 99})$$

where C_{00} , C_{01} , C_{10} and C_{11} are respectively the cost of true rejection and miss, false and true detections; P_D and P_F are the probability of true and false detection; and P_0 and P_1 are *a priori* probabilities of the null and alternate hypotheses.

The hypotheses are mutually exclusive such as,

$$P_0 = 1 - P_1 . \quad (\text{eq. 100})$$

By substituting (eq. 100) in (eq. 99), the detection cost function is expressed as:

$$\mathfrak{R} = C_{00} + P_F (C_{10} - C_{00}) + P_1 [(C_{01} - C_{00}) + P_D (C_{01} - C_{11}) - P_F (C_{10} - C_{00})] . \quad (\text{eq. 101})$$

Where maximal cost is the function derivative with respect to P_1 which corresponds to the minimax equation,

$$L(P_D, P_F) = \frac{\partial \mathfrak{R}}{\partial P_1} = (C_{01} - C_{00}) + P_D (C_{01} - C_{11}) - P_F (C_{10} - C_{00}) = 0 . \quad (\text{eq. 102})$$

The OOP of the Receiver Operating Characteristic (ROC) curve is therefore the point where the minimax equation is minimal,

$$\nabla L(P_D, P_F) = \frac{\partial L}{\partial P_D} + \frac{\partial L}{\partial P_F} = 0 , \quad (\text{eq. 103})$$

which corresponds to the point on the ROC curve where its derivative equals to (eq. 34):

$$\frac{dP_D}{dP_F} = \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})} .$$

The case where the cost for miss and false detections is the same and no cost is attributed for correct detection is called the maximum-likelihood estimator which defines the optimal operation point as where the ROC curve derivative equal 1 (eq. 35),

$$\frac{dP_D}{dP_F} = 1 .$$