Advanced signal processing for the identification and diagnosis of the condition of rotating machinery

Peeters Cédric

Vrije Universiteit Brussel & INSA Lyon

Doctoral Thesis

Advanced signal processing for the identification and diagnosis of the condition of rotating machinery

Author:
Cédric Peeters

Supervisors:
Prof. dr. ir. J. Helsen
Prof. dr. ir. J. Antoni
Prof. dr. ir. P. Guillaume

Jury:
Prof. dr. R. Randall
Prof. dr. R. Vounckx
Prof. dr. N. Martin
N. Girard
Prof. dr. ir. Q. Leclère
Prof. dr. A. R. Nejad
Prof. dr. C. Capdessus

Secretary:
Prof. dr. ir. C. Devriendt

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

November 21, 2019

N°d'ordre NNT : 2019LYSEI107

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
Laboratoire Vibrations Acoustique
En cotutelle internationale avec
Applied Mechanics, Vrije Universiteit Brussel

Ecole Doctorale N° accréditation
MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE

Spécialité / discipline de doctorat :
Mécanique

Soutenue publiquement le 21/11/2019, par :
Cédric Peeters

Advanced signal processing for the identification and diagnosis of the condition of rotating machinery

Devant le jury composé de :

Martin Nadine, Directeur de Recherche, CNRS Gibsa-lab Rapporteure
Nejad Amir, Maître de Conférences HDR, NTNU Rapporteur

Randall Robert, Professeur des Universités, UNSW Examinateur
Leclere Quentin, Maître de Conférences HDR, INSA-Lyon Examinateur
Capdessus Cécile, Maître de Conférences, Université d'Orléans Examinatrice

Antoni Jérôme, Professeur des Universités, INSA-Lyon Directeur de thèse
Helsen Jan, Maître de Conférences, VUB Directeur de thèse
Guillaume Patrick, Professeur des Universités, VUB Co-directeur de thèse

Département FEDORA – INSA Lyon - Ecoles Doctorales – Quinquennal 2016-2020

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage **secretariat@edchimie-lyon.fr** INSA : R. GOURDON	M. Stéphane DANIELE Institut de recherches sur la catalyse et l'environnement de Lyon IRCELYON-UMR 5256 Équipe CDFA 2 Avenue Albert EINSTEIN 69 626 Villeurbanne CEDEX **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **http://edeea.ec-lyon.fr** Sec. : M.C. HAVGOUDOUKIAN **ecole-doctorale.eea@ec-lyon.fr**	M. Gérard SCORLETTI École Centrale de Lyon 36 Avenue Guy DE COLLONGUE 69 134 Écully Tél : 04.72.18.60.97 Fax 04.78.43.37.17 **gerard.scorletti@ec-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 INSA : H. CHARLES **secretariat.e2m2@univ-lyon1.fr**	M. Philippe NORMAND UMR 5557 Lab. d'Ecologie Microbienne Université Claude Bernard Lyon 1 Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX **philippe.normand@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://www.ediss-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 INSA : M. LAGARDE **secretariat.ediss@univ-lyon1.fr**	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 Avenue Jean CAPELLE INSA de Lyon 69 621 Villeurbanne Tél : 04.72.68.49.09 Fax : 04.72.68.49.16 **emmanuelle.canet@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 **infomaths@univ-lyon1.fr**	M. Luca ZAMBONI Bât. Braconnier 43 Boulevard du 11 novembre 1918 69 622 Villeurbanne CEDEX Tél : 04.26.23.45.52 **zamboni@maths.univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bât. Direction **ed.materiaux@insa-lyon.fr**	M. Jean-Yves BUFFIÈRE INSA de Lyon MATEIS - Bât. Saint-Exupéry 7 Avenue Jean CAPELLE 69 621 Villeurbanne CEDEX Tél : 04.72.43.71.70 Fax : 04.72.43.85.28 **jean-yves.buffiere@insa-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bât. Direction **mega@insa-lyon.fr**	M. Jocelyn BONJOUR INSA de Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69 621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **http://ed483.univ-lyon2.fr** Sec. : Véronique GUICHARD INSA : J.Y. TOUSSAINT Tél : 04.78.69.72.76 **veronique.cervantes@univ-lyon2.fr**	M. Christian MONTES Université Lyon 2 86 Rue Pasteur 69 365 Lyon CEDEX 07 **christian.montes@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

"*All dreams can come true, only if we have the courage to chase them.*"

Walt Disney

Preface

When I started working at VUB in June of 2015, I actually had no idea I had just taken my first step onto the four-year path that would be my PhD. In fact, I agreed initially to a one-year research project that would be a collaboration between the department of Applied Mechanics at VUB and the Acoustics & Vibrations R&D department at Atlas Copco. However, after starting this alleged one-year journey, it soon became clear to me that I found the topic of the research deeply fascinating. During my master's in mechanical engineering, I already had a strong interest in courses related to either vibrations or acoustics, so this new research project fitted quite naturally into my range of interests. On top of that, I genuinely enjoyed coming to work and collaborating with the people both at VUB and Atlas Copco. Thus unsurprisingly, when my supervisor Jan Helsen proposed to extend the research project into a full-fledged PhD track, I did not really have to think twice. And so I began my four-year PhD odyssey, during which I got many opportunities to go and meet interesting people, to collaborate, to further my research, and to present my work at (inter)national events.

Just as agreeing to the one-year research project was my first step onto the PhD path, so the writing of this preface is the last step. I am now at the end of what was a captivating and thought-provoking journey, and I want to thank the people who helped or encouraged me along the way.

First off, I want to express my gratitude to the person who not only made this PhD possible, but also guided me practically every step of the way. Jan, thank you for these past four years, during which you advised me masterfully, provided me with many great opportunities to advance my career, and perhaps most important of all, believed in me and in a successful ending of this PhD. Your engagement, relentless motivation, and clear vision has always inspired me and continues to

inspire me to work hard and to look towards the future. On a personal level, I never regretted accepting the PhD since the team spirit has always been top-notch.

One of those great opportunities offered to me during my PhD, was the collaboration with my French promotor, Prof. dr. ir. Jérôme Antoni of the Laboratoire Vibrations Acoustique at INSA Lyon. This collaboration between VUB and INSA Lyon, in the form of a joint PhD, proved truly invaluable to my research and personal scientific development. When I first started out with my PhD, I spent a good deal of my time reading scientific papers about condition monitoring to study more about the topic. Then, I learned that there was a possibility to collaborate with the person whose name I saw on many of those papers. Naturally, I was initially quite nervous about how this collaboration would go. To my great fortune however, this person turned out to be very kind, patient, and helpful. Jérôme, thank you so much for not only agreeing to this collaboration, but also for helping me understand complex signal processing concepts, for making time in your very busy schedule, and in general for elevating my research level altogether.

Another person from INSA-Lyon who deserves mentioning, is Quentin Leclère. Your patience and insightful help with the multi-order probabilistic approach was very much appreciated!

I would also like to express my gratitude to the head of our Applied Mechanics department and co-promotor of this PhD, Patrick Guillaume. Together with Jan, he convinced me to start working at VUB, he advised me during my PhD, and he helped facilitate the collaboration with INSA-Lyon.

During a conference at MIT in Boston in May of 2018, I got the chance to meet the team of Prof. Steven Leeb and get a tour of their laboratory. What followed was a fruitful collaboration about vibration-based rotation speed estimation. Many thanks go out to Pete Lindahl for clarifying the ViBES method to me and illustrating it on experimental vibration data, to John Donnal for providing interesting data sets and pictures of a navy ship generator, and lastly to Steven Leeb for making it all possible.

An aspect that was invaluable for this PhD research was the collaboration with industrial partners. In this regard, I am grateful to Hans Vande Sande, Philippe Geuens, and Jeroen Firlefijn of Atlas Copco for providing me with interesting research challenges and helping me wherever necessary. Additionally, I would also like to thank Alexandros Skrimpas of MHI Vestas for the close and useful collaboration.

Thanks are also in order to my entire PhD jury for critically assessing my PhD manuscript and providing me with fruitful comments that helped augmenting the level of the manuscript.

I can undoubtedly say that one of the reasons I liked coming to VUB every day was due to my helpful, friendly, and funny colleagues. Pieter-Jan, thank you for joining our team with full conviction in what we are currently doing and for being a critical voice to discuss topics with. Timothy, thank you for working together with me and for being patient despite my slow understanding of your complicated machine learning magic tricks. Nicoletta, thank you for not thinking I was crazy when I said the word 'cepstrum' for the millionth time in a single day and for always voluntarily cleaning the coffee machine during breaks (I do hope you don't carry this burden anymore at your new job at Dana in Italy!). I also want to thank Wout, Nymfa, Christof, Tim, and Max of our neighboring office. You guys were a joy to organize the EAWE PhD seminar with, to do social events with, to go conferences with, and in general to talk to during breaks.

My stay in Lyon has also left me with happy and lasting memories. Besides being a very nice city to live in, I want to thank Gianluigi, Giorgio, Souha, Edouard & Rémi for making my time in Lyon and in the office absolutely delightful and entertaining. Thanks for the skiing, the board games in this interesting little café called "Moi j'm'en fous je triche", and the trip to les Gorges du Verdon and Bormes-Les-Mimosas!

Furthermore, thanks to Birgit, Maxime, and Jenny for helping me with all my practical or administrative problems in the past four years.

A major factor which helped keeping me sane during the PhD was squash. Therefore, thank you to my squash sparring partners in the past years and to my squash club 'De Vaart' for providing a great environment to decompress in after work by hitting a rubber ball as hard as possible against a wall.

Of course, I would not have been able to enjoy my PhD as much as I did now without the much-appreciated support of my friends from Leuven. Roeland, Daniel, Yannick, Egon, Kaylian, and Paul, thank you guys for being in my corner!

If there is one person who followed my four-year PhD journey from up close, it is my girlfriend Romina. Thank you for your patience, understanding, and support during these challenging years. You kept my life balanced and reminded me about the importance of taking a few steps back and looking at issues from a distance.

Finally, I want to thank my parents for their eternal and unconditional support of all my endeavors. I can say without a doubt that they form the pillars of any success I've ever had and that without them, I would stand nowhere.

Obviously, I am also eternally grateful to my sister for the role she played in my life. She was there every step of the way and whenever I had a problem, she would be there to listen and give me advice. Special thanks go to her husband Rob for not only teaching me how to improve my squash level, but also in general for being an awesome brother-in-law who is always willing to help wherever possible and who appreciates my amateur DJ mixes.

Lastly, I'm beyond grateful for the endless love and support of my grandparents who have stood by me from the very beginning. Your unequivocal belief in me has always motivated me to do better and never give up.

Abstract

This Ph.D. dissertation targets innovative methods for vibration-based condition monitoring of rotating machinery. Substantial benefits can be achieved from an economical and a safety point of view using condition monitoring. One of the most popular methods to gather information about the state of machine parts is through the analysis of machine vibrations. Most of these vibrations are directly linked to periodical behavior of subsystems within the machine like e.g. rotating shafts, gears, rotating electrical fields, etc. This knowledge can be exploited to enable fault-dependent processing schemes. This dissertation investigates how to implement and utilize these processing schemes and details the steps in such a procedure.

Typically, the first prerequisite for advanced analysis is the availability of the instantaneous rotation speed. This speed needs to be known since most frequency-based analysis techniques assume stationary behavior. Knowledge of the speed thus allows for compensating speed fluctuations, for example through angular resampling of the vibration signal. While there are hardware-based solutions for speed estimation using angle encoders or tachometers, this thesis investigates the potential in vibration signals for speed estimation.

After speed estimation and angular resampling, a common next step is to separate the signal into deterministic and stochastic components. The cepstrum editing procedure is examined for its efficacy and applicability. Afterwards, different filtering methods are inspected as to improve the signal-to-noise ratio of the signal content of interest. Existing methods using conventional criteria are investigated together with a novel blind filtering methodology.

The final step in the multi-step processing scheme is to search for the potential fault. Statistical indicators can be calculated on the processed time domain signal and tracked over time to check for increases. In many cases, the fault signature exhibits cyclostationary

behavior. Therefore this dissertation also examines different cyclostationary analysis techniques. Lastly, the performance of the different processing methods is validated on two experimental vibration data sets of wind turbine gearboxes.

French summary

Cette thèse porte sur des méthodes innovantes de contrôle de l'état de santé des machines tournantes par l'analyse des signaux vibratoires. En effet, la surveillance de l'état de santé des machines contribue à des améliorations substantiels des points de vue économique et de sureté. Afin d'y aboutir, l'une des manières les plus populaires est de recueillir les vibrations de la machine. La plupart de ces vibrations sont directement liées au comportement périodique des sous-systèmes de la machine tels que les arbres de rotation, engrenages, champs électriques rotationnels, etc. Cette connaissance peut être exploitée afin de concevoir une méthodologie adaptée à chaque type de défaut. Cette thèse s'intéresse aux étapes de la mise en œuvre de cette méthodologie. En règle générale, la première condition préalable à l'analyse avancée de l'information récoltée est la disponibilité de la vitesse intantanée de rotation. Cette vitesse doit être connue car la plupart des techniques du traitement du signal sont adaptées aux conditions de fonctionnement stationnaires. Ainsi, la connaissance de la vitesse permetra de compenser les fluctuations de vitesse, par exemple par le ré-échantillonnage angulaire du signal de vibration. Malgré l'existence d'outils de mesure permettant l'estimation de la vitesse tels que les codeurs et les tachymètres, cette thèse étudie le potentiel d'estimer la vitesse instantanée de rotation à partir des signaux vibratoires. Après l'estimation de la vitesse et le ré-échantillonnage angulaire, une étape suivante courante consiste à séparer le signal en composantes déterministes et stochastiques. Dans ce sens, l'efficacité et l'applicabilité de la procédure d'édition du cepstre sont analysées. Ensuite, différentes méthodes de filtrage sont appliquées au signal résiduel afin d'améliorer le rapport signal sur bruit. Pour cette fin, les méthodes existantes utilisant des critères conventionnels sont étudiées en parallèlle avec une nouvelle méthodologie aveugle de filtrage. La dernière étape du processus de traitement consiste à diagnostiquer le

défaut potentiel. Ainsi, des indicateurs statistiques sont calculés sur le signal obtenu après traitement et suivis dans le temps pour vérifier leurs variations. Dans de nombreux cas, la signature du défaut présente un comportement cyclostationaire. Par conséquent, cette thèse examine également différentes techniques d'analyse de la cyclostationarité. Enfin, les performances des différentes méthodes de traitement sont validées sur deux ensembles de données expérimentales de vibrations issues de boîtes de vitesses d'éoliennes.

Dutch summary

Deze dissertatie bespreekt innovatieve methoden om de gezondheidstoestand van roterende machines te monitoren door trillingssignalen te analyseren. Het bewaken van de gezondheidstoestand van de machines draagt bij tot aanzienlijke verbeteringen vanuit economisch en veiligheidsoogpunt. Een van de meest populaire opties voor het monitoren van machines is om de trillingen van de machine te verzamelen. De meeste van deze trillingen zijn direct gerelateerd aan het periodieke gedrag van machinesubsystemen zoals roterende assen, tandwielen, roterende elektrische Velden, etc. Deze kennis kan worden gebruikt om een methodologie te ontwerpen die is aangepast aan elk type defect. Dit proefschrift richt zich op de verschillende stadia in de implementatie van zulk een methodologie. Over het algemeen is de beschikbaarheid van de rotatiesnelheid een eerste vereiste voor een geavanceerde analyse van de verzamelde trillingsinformatie. Deze rotatiesnelheid dient gekend te zijn omdat de meeste signaalverwerkingstechnieken zijn aangepast aan stationaire werkingsomstandigheden. Kennis van de rotatiesnelheid maakt het aldus mogelijk om snelheidsschommelingen te compenseren, bijvoorbeeld door het herbemonsteren van het trillingssignaal in het hoekdomein in plaats van het tijdsdomein. Ondanks het bestaan van meetinstrumenten voor het meten van de rotatiesnelheid zoals encoders en tachometers, bestudeert dit proefschrift het potentieel voor het estimeren van de instantane rotatiesnelheid met trillingssignalen. Na het schatten van de rotatiesnelheid en de herbemonstering, is een gebruikelijke volgende stap het splitsen van het signaal in deterministische en stochastische componenten. De efficiëntie en toepasbaarheid van de methodes gebaseerd op het gebruik en aanpassen van het cepstrum worden hiervoor onderzocht. Vervolgens worden verschillende filtermethoden toegepast op het residueel signaal om de signaal-ruisverhouding te verbeteren. De dissertatie bekijkt naast bestaande methoden die conventionele

criteria gebruiken ook een nieuwe zogenaamde blinde filtermethode. De laatste stap in het analyse proces is om het potentieel aanwezige defect te diagnosticeren. Hiervoor worden statistische indicatoren berekend op het residuele signaal. Deze indicatoren worden in de tijd opgevolgd om zo mogelijke variaties en anomalieën te detecteren. Vaak vertoont het trillingspatroon van het defect cyclostationair gedrag. Daarom onderzoekt dit proefschrift ook verschillende technieken voor het analyseren van cyclostationariteit. Ten slotte wordt de effectiviteit van de verschillende analysemethoden gevalideerd op twee data sets van experimentele trillingsgegevens gemeten op tandwielkasten voor windturbines.

Contents

List of Figures

List of Tables

List of Abbreviations

CM	**C**ondition **M**onitoring
O&M	**O**perations & **M**aintenance
PCA	**P**rincipal **C**omponent **A**nalysis
SCADA	**S**upervisory **C**ontrol **A**nd **D**ata **A**cquisition
ANN	**A**rtificial **N**eural **N**etwork
SVM	**S**upport **V**ector **M**achine
RNN	**R**ecurrent **N**eural **N**etwork
NBM	**N**ormal **B**ehavior **M**odeling
DFIG	**D**ouble **F**ed **I**nduction **G**enerator
SHM	**S**tructural **H**ealth **M**onitoring
AE	**A**coustic **E**mission
PEC	**P**ower **E**lectronic **C**onverter
WECS	**W**ind **E**nergy **C**onversion **S**ystem
IGBT	**I**nsulated **G**ate **B**ipolar **T**ransistor
OEM	**O**riginal **E**quipment **M**anufacturer
IIoT	**I**ndustrial **I**nternet **o**f **T**hings
OPC	**O**pen **P**latform **C**ommunications
SQL	**S**tructured **Q**uery **L**anguage
CAPEX	**Cap**ital **Ex**penditures
OPEX	**Op**erating **Ex**penditures
IAS	**I**nstantaneous **A**ngular **S**peed
DRS	**D**iscrete **R**andom **S**eparation
SANC	**S**elf **A**daptive **N**oise **C**ancellation

LP	**L**inear **P**rediction
TSA	**T**ime **S**ynchronous **A**veraging
OMA	**O**perational **M**odal **A**nalysis
IRF	**I**mpulse **R**esponse **F**unction
TFR	**T**ime **F**requency **R**epresentation
WT	**W**ind **T**urbine
STFT	**S**hort **T**ime **F**ourier **T**ransform
MHD	**M**ulti **H**armonic **D**emodulation
MOPA	**M**ulti **O**rder **P**robabilistic **A**pproach
MED	**M**inimum **E**ntropy **D**econvolution
ACEP	**A**utomated **C**epstrum **E**diting **P**rocedure
IMF	**I**ntrinsic **M**ode **F**unction
IF	**I**nstantaneous **F**requency
TKEO	**T**eager **K**aiser **E**nergy **O**perator
EKF	**E**xtended **K**alman **F**ilter
MLE	**M**aximum **L**ikelihood **E**stimation
FDEO	**F**requency **D**omain **E**nergy **O**perator
TQR	**T**ime **Q**uefrency **R**epresentation
STCT	**S**hort **T**ime **C**epstrum **T**ransform
VKF	**V**old **K**alman **F**ilter
SHD	**S**ingle **H**armonic **D**emodulation
ISHD	**I**terative **S**ingle **H**armonic **D**emodulation
SNR	**S**ignal-to-**N**oise **R**atio
REB	**R**olling **E**lement **B**earings
CPW	**C**epstrum Pre-**W**hitening
NREL	**N**ational **R**enewable **E**nergy **L**aboratory
SS	**S**pectral **S**ubtraction
FFT	**F**ast **F**ourier **T**ransform
SE	**S**quared **E**nvelope

SES	**S**quared **E**nvelope **S**pectrum
AWGN	**A**dditive **W**hite **G**aussian **N**oise
SDOF	**S**ingle **D**egree **o**f **F**reedom
BSF	**B**all **S**pin **F**requency
FTF	**F**undamental **T**rain **F**requency
BPFO	**B**all **P**ass **F**requency **O**uter-race
BPFI	**B**all **P**ass **F**requency **I**nner-race
MDOF	**M**ultiple **D**egree **O**f **F**reedom
CRB	**C**ylindrical **R**oller **B**earing
TRB	**T**apered **R**oller **B**earing
HSS	**H**igh **S**peed **S**haft
ISS	**I**ntermediate **S**peed **S**haft
LSS	**L**ow **S**peed **S**haft
MOMEDA	**M**ultipoint **O**ptimal **M**inimum **E**ntropy **D**econvolution **A**djusted
MKCD	**M**aximum **C**orrelated **K**urtosis **D**econvolution
ICS2	**I**ndicator **o**f **S**econd-order **C**yclostationarity
AR	**A**uto**R**egressive
RPE	**R**elative **P**rediction **E**rror
LPC	**L**inear **P**rediction **C**oding
RQ	**R**ayleigh **Q**uotient
VSDFT	**V**elocity **S**ynchronous **D**iscrete **F**ourier **T**ransform
SIMO	**S**ingle **I**nput **M**ultiple **O**utput
CSC	**C**yclic **S**pectral **C**oherence
EES	**E**nhanced **E**nvelope **S**pectrum
DTF	**D**ynamometer **T**est **F**acility
GMF	**G**ear **M**esh **F**requency

List of Symbols

A	amplitude
f	frequency
C_c	complex cepstrum
C_r	real cepstrum
c_{LP}	long-pass liftered real cepstrum
B_w	bandwidth
K_w	number of windows
$N_{overlap}$	number of overlapping points
$\mathfrak{F}$	Fourier transform
$\mathfrak{F}^{-1}$	inverse Fourier transform
T	fundamental period
$H(s)$	transfer function in frequency domain
$\boldsymbol{h}$	impulse response
$\boldsymbol{F}$	Fourier matrix
I_e	negentropy in the time domain
I_E	negentropy in the spectral domain
D_{KL}	Kullback-Leibler divergence (or relative entropy)
E_i	the i^{th} octile
D_r	roller diameter
D_p	pitch diameter
N_r	number of rollers
$\mathbb{E}\left\{.\right\}$	Expected value operator

$\mathbb{V}.$	Variance operator
α	harmonic order
ϵ_x	squared envelope of x
ϕ	constant phase
θ	instantaneous angle of rotation
σ	standard deviation
Ψ_c	Teager-Kaiser Energy Operator
ω	angular frequency
γ	acceleration of the instantaneous rotation speed
σ_e	prediction error of squared envelope

Chapter 1

Introduction & Outline

1.1 Condition Monitoring

The process of tracking the health of machinery is commonly known as condition monitoring or *CM*. Typically, it involves recording data, analyzing this data, and then inspecting the resulting indicators for potential significant changes that could be symptomatic of a defect. Incorporating condition monitoring in the Operations and Maintenance (O & M) scheme opens the door for predictive maintenance. This means that maintenance can be planned which allows scheduling on beforehand what actions need to be performed and when exactly. Ideally, degradation is detected early on such that only minor preventative actions need to be undertaken and that more grave repercussions are avoided. This is actually the main reason why condition monitoring is used, since the machine operator does not want a reduced machine lifespan. Condition monitoring is most often employed on rotating machinery (e.g. pumps, internal combustion engines, compressors, presses, gearboxes, generators, electric motors, ...), but can also be utilized for static utilities and machinery (e.g. cables, pipes, converters). It should not be confused with structural health monitoring, which is a term reserved for the monitoring of engineering structures.

Predictive maintenance is a term that has gained a lot of attention over the past few years, but it is not all hype. In total it is estimated that the predictive maintenance market is currently worth over 3 billion US dollars in 2019 and will continue to grow with a projected value

of 10.7 billion US dollars by 2024 [1]. The company PricewaterhouseCoopers (PwC) points out in their predictive maintenance market survey of 2018 [2] that a majority of companies is very interested in improving and expanding their predictive maintenance activities. They defined four maturity levels where visual inspections represent level 1, instrument inspections and continuous condition monitoring levels 2 and 3, and big data analytics with condition-based decision-making represent level 4. They refer to this last level as Predictive Maintenace 4.0 or PdM 4.0. To give a quick idea about the integration and development of condition monitoring and predictive maintenance in the industry, some key findings of their survey are presented:

- 11% of companies are already combining condition monitoring with big data to drive decision-making, i.e. PdM 4.0 (shown in Fig. 1.1).
- 60% of companies have either intentions or already concrete plans to use big data together with condition monitoring to drive decision-making and improve uptime (shown in Fig. 1.2).
- Companies that have no plans for PdM 4.0, attribute this most often to the lack of budget or to being irrelevant for their business case.
- 95% of companies that use PdM 4.0, confirm that it is responsible for improving one or more key maintenance value drivers.
- 60% of companies see an asset uptime increase of on average 9%.

Based on these numbers, it is clear that condition monitoring and predictive maintenance is here to stay and will continue growing in the foreseeable future. In the next section some common types of condition monitoring are discussed to give the reader a broad overview of what is available and the current practices.

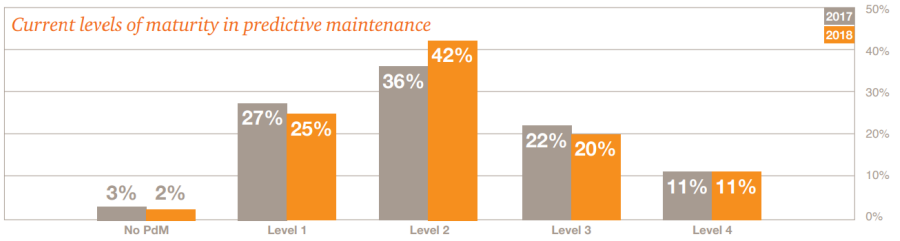

FIGURE 1.1: Distribution of predictive maintenance maturity levels according to a PwC 2018 market survey, as reproduced from [2].

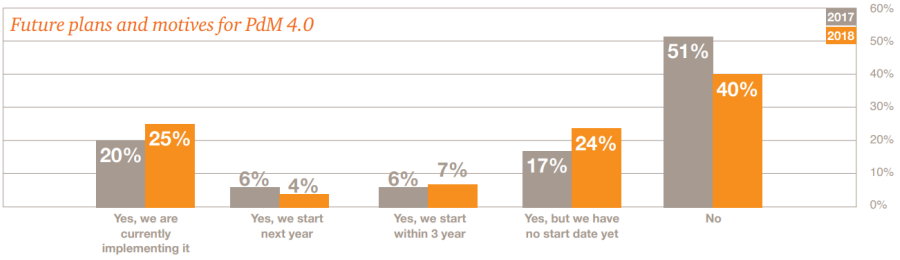

FIGURE 1.2: Distribution of future plans of predictive maintenance according to a PwC 2018 market survey, as reproduced from [2].

1.1.1 Types of condition monitoring

Condition monitoring is an umbrella term that spans various different ways of tracking the health state of a machine. A short overview of the most commonly used types of condition monitoring is provided. It should be noted that the research described in this dissertation concentrates solely on *vibration-based* monitoring and that often the focus lies on wind energy applications. The reason for the focus on vibration-based monitoring is specified later in the dissertation in section 1.1.1.7.

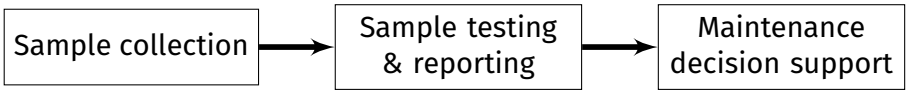

FIGURE 1.3: Lubrication condition monitoring scheme [3].

1.1.1.1 Lubricants

Lubration-based condition monitoring is often used as an early warning system to indicate incipient deterioration. Typically, viscous oil or grease is used as lubricant in machines and it is analyzed for potential contamination and property changes. A conventional lubrication scheme consists of three steps [3] (also shown in Fig. 1.3:

1. lubricant sampling, including sample scheduling and collecting
2. lubricant sample testing and reporting
3. maintenance decision support by analyzing and interpreting the results

The sampling of the lubricant is nowadays often still executed in an off-line manner. This means that a sample of the lubricant is taken out of the machine and then analyzed in a laboratory. The downside of this approach is that there can be significant delays before a change in the lubricant contamination or properties is detected. Therefore, there is a rising interest in continuous on-line sampling where samples are drawn and examined while the machine is operational.

Some parameters of lubricants that are commonly investigated are shown in Table 1.1. These parameters are then used as input for further analysis. Examples of processing methods of the lubrication data include thresholding, statistical analysis, and spectroscopic analysis. The lubrication data and its processed derivatives can then be analyzed using statistical approaches (e.g. trending, regression,...), artificial intelligence, model-based approaches (e.g. simulation and experimental models), or hybrid methodologies.

Type	Parameters
Additives	Boron, Barium, Calcium, Magnesium, Molybdenum, Phosphorus, Sodium, Silicon, Zinc [4, 5]
Contamination	Vanadium, Water, Sodium, Boron, Coolant, Soot/carbon, Potassium, Silicon [6–8]
Wear elements	Chromium, Titanium, Iron, Copper, Lead, Tin, Aluminum, Nickel, Vanadium, Silver [9–12]
Physical & chemical	Total Base Number (TBN), Total Acid Number (TAN), Viscosity @ 400C, Viscosity @ 1000C, Flash point [13–15]

TABLE 1.1: Summary of lubricant parameters that are commonly investigated (based on [3]).

1.1.1.2 Thermals

Temperature is probably one of the most commonly tracked parameters of a machine to assess its health state [16]. This is primarily due to the fact that temperature abnormalities can be caused by many different factors (e.g. faulty operation control, damaged materials, malfunctioning electrical connections,etc.).

One way to utilize temperature is to simply track temperature values measured by e.g. a thermocouple and then use these values as input to a statistical or machine learning approach. Both approaches try to detect anomalous behavior either using thresholds based on historical experience and regression techniques or using expected behavior modeling (e.g. K-means clustering, Bayesian Ridge Regression, etc. [17]).

Another way to monitor the thermals of a machine is through the use of infrared thermography. It allows to measure the temperature in real time and in a non-contact manner. Thermography has enjoyed a steady increase in interest and usage in condition monitoring applications. Examples of various thermography applications include

the monitoring of structures [18, 19], chemical processes [20], electrical equipment [21–23], wood and paper inspection [24, 25], manufacturing processes [26–28], material fatigue [29–31], machinery [32–34], aerospace [35], plastic and tensile deformation [36–39], nutrition [40], electronics [41, 42], nuclear power [43], etc. Additionally, thanks to the ever-improving infrared camera technology, thermography is developing into a more economical, robust and precise technique.

The main principle of thermography is based on the infrared radiation an object emits if it is above 0 K or -273°C. Typically this infrared radiation is observed by an infrared camera whose output can then be examined in the form of thermal imagery. Figure 1.4 displays a conventional thermography setup to monitor a mechanical system, i.e. an impeller. Infrared imagery usually has a low signal-to-noise ratio, thus necessitating the need to enhance the measured images. Some image analysis methods that are employed to process the infrared images include principal component analysis (PCA) [44], texture-based analysis [45], global and fuzzy thresholding [46, 47], morphological segmentation [16, 48, 49], edge detection filters [16], etc. The goal of most of these techniques is to detect potential hot-spots and quantify them in the form of condition indicators.

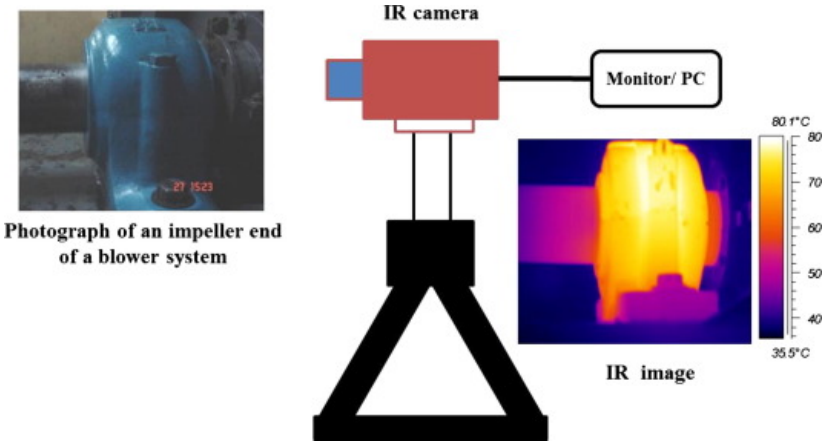

FIGURE 1.4: Overview of a common setup used for infrared thermography experiments, reproduced from [16].

1.1.1.3 SCADA

Supervisory Control and Data Acquisition (SCADA) is a control system structure that serves to provide high-level process information. In practice, this involves the use of computers (e.g. PID and PLC controllers) and data communication networks. SCADA systems have become one of the most common forms of industrial control systems. A good example is its usage in wind turbine performance monitoring. A SCADA system is installed in larger wind turbines almost by default, thus providing valuable information about the operating behavior of the turbine. It should be noted that when SCADA data is mentioned or referred to in this dissertation, it explicitly refers to SCADA data usage in a wind turbine context.

SCADA data in the wind turbine case is typically either sampled at a 10 minute or a 1 second interval. From a condition monitoring point of view, this data can serve as a low-cost potential solution since no extra sensors are required. The entire list of parameters tracked in the SCADA data is typically quite extensive, but an overview of basic SCADA parameters is given in Table 1.2. Normally the extent and quality of the SCADA data depends on the turbine manufacturer. Other possible usages besides condition monitoring include power curve analysis [50] and modeling with e.g. k-nearest-neighbour [51], spare part demand forecasting [52], and load monitoring [53].

In the literature, many researchers investigated the potential of SCADA data for usage in condition monitoring of different turbine components. While wind turbine SCADA systems were not developed for dedicated condition monitoring purposes, interest in using it for such practices increased the moment maintenance became a priority. Tautz et al. [54] categorize the different SCADA-based monitoring methods into five classes:

i Trending

ii Clustering

iii Normal behavior modeling

iv Damage modeling

TABLE 1.2: Overview of some basic SCADA parameters (based on [54–64]).

Category	**SCADA parameter**
Environmental	wind speed, wind direction, ambient temperature, nacelle temperature
Electrical	active power output, power factor, reactive power, generator voltages, generator phase current, voltage frequency
Control variables	pitch angle, yaw angle, rotor shaft speed, fan speed/status, generator speed, cooling pump status, number of yaw movements, set pitch angle/deviation, number of starts/stops, operational status code
Temperatures	gearbox bearing, gearbox lubricant oil, generator winding, generator bearing, main bearing, rotor shaft, generator shaft, generator slip ring, inverter phase, converter cooling water, transformer phase, hub controller, top controller, converter, controller, grid busbar

v Assessment of alarms and expert systems

i. The challenge of the first monitoring class lies within the interpretation of the trends. Due to the high variability of the operating regime of modern wind turbines, changes in the SCADA parameters do not automatically correspond to the presence of a defect. A very straightforward approach is to monitor the SCADA parameters over a long period of time and use statistical thresholds for alarming.

ii. Clustering deals with the problem of the visual interpretation of the trends. When large numbers of wind turbines need to be monitored efficiently, it becomes imperative to have an automatic manner to classify the turbines as 'healthy' or 'faulty'. Kusiak et al. analyzed

drivetrain and tower accelerations using SCADA data by means of a modified k-means clustering [65] conditioned on the wind speed. Kim et al. [66] and Catmull [60] applied self-organizing maps instead of k-means to build such clusters. Despite these advancements, the interpretation of the clustering results is often still perceived to be difficult [54].

iii. Normal behavior modeling (NBM) employs the same idea of anomaly detection as the previous techniques, but focuses more on the empirical modeling of the measurements. The residual error between the modeled and observed parameter then serves as a health indicator. A basic example of NBM involves the use of linear and polynomial models. A linear autoregressive model with exogenous inputs (ARX) was used by Garlick et al. [67] to detect generator bearing failures from the bearing temperature. Higher-order polynomial full signal reconstruction (FSRC) models of drivetrain temperatures were developed by Wilkinson et al. [63] to detect gearbox and generator failures.

Other approaches typically involve more machine learning-based techniques such as artificial neural networks (ANN) or support vector machines (SVM). These approaches allow for modeling of non-linear relationships between measurements from training data sets. Without going into too much detail of how a neural network functions, the network architecture normally consists of one input layer, a varying number of hidden layers and an ouput layer. The layers can consist of varying numbers of neurons whose inputs are the outputs of the previous layer or the input parameters. A neuron has a non-linear transfer function that merges the inputs and an activation function that determines whether or not output is produced. The training phase of the network then boils down to optimizing the input weights. Standard examples of networks are feed-forward and recurrent neural networks (RNN). Schlechtingen et al. [68] compared a simple linear model approach to a neural network approach in a study performed on 14 months of SCADA data from ten 2 MW offshore wind turbines. The generator bearing temperature was modeled using the nacelle temperature, generator stator temperature, produced power and generator speed. In this study the ANN outperformed the linear model for bearing fault detection due to the strong seasonal influences on the measurements.

iv. Damage modeling tries to incorporate more physical insight into the modeling step. Instead of training empirical normal behavior models, the measurements are interpreted with physical models to improve the accuracy. Gray et al. [69] developed a damage model using physical failure modes of interest to estimate the failure probability. A general scheme for a physics-based monitoring approach was proposed by Breteler et al. [70] and is shown in Fig. 1.5. A theoretical model of a wind turbine gearbox with Double-Fed Induction Generator (DFIG) was built by Qiu et al. [71]. This model utilized temperature trending and the underlying thermodynamic principles to track the degradation of a generator ventilation fault.

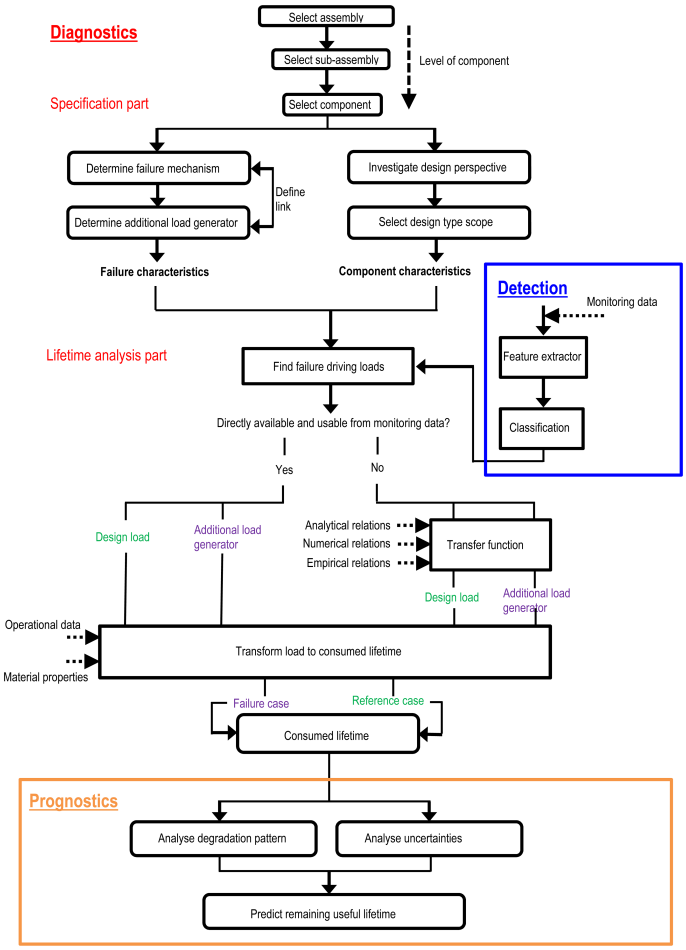

FIGURE 1.5: Framework for a physics-of-failure-based methodology, reproduced from [70].

v. The last class, assessment of alarms and expert systems, looks at the outputs of the SCADA control alarms or the NBM output alarms. A typical example of this class is the analysis of the status codes of the wind turbine. Status code processing approaches typically investigate the possibility of extracting useful, actionable information about the health of the turbine from these status codes and there exist many

different ways to do this. For example, Chen et al. [72] used a probabilistic approach with Bayesian networks to track down root causes for failures such as a pitch fault. Qiu et al. [73] also utilized Bayes' theorem and compared the extracted patterns using a Venn diagram of which an example can be seen in Fig. 1.6. Other approaches often involve machine learning such as Kusiak et al. [74] who used neural network ensembles to predict status codes and their severity to detect a diverter malfunctioning.

Lastly, a significant amount of research examines the use of "expert systems to interpret the status codes or model outputs. Often these systems are based on using fuzzy logic to determine a diagnosis for anomalies. Example research works that are based on or employ fuzzy logic are given in [58, 62, 64, 75–77].

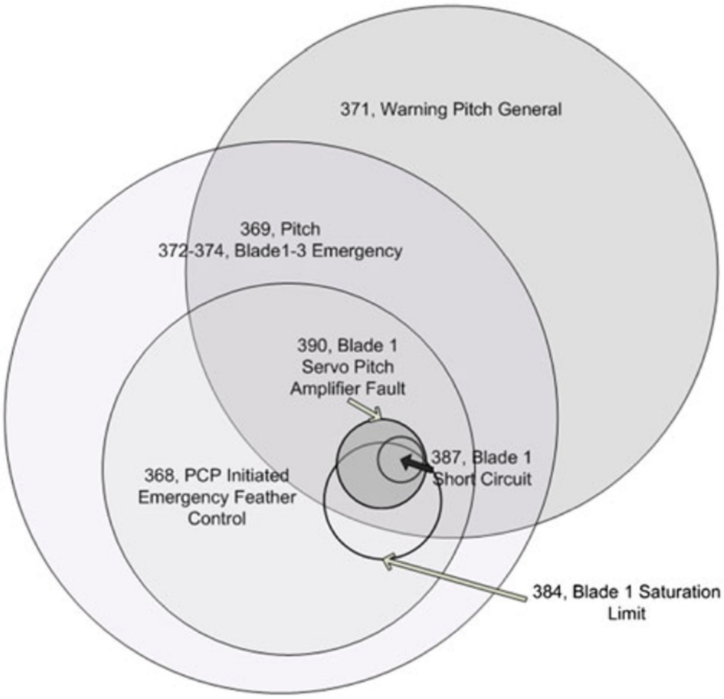

FIGURE 1.6: Example probability-based alarm Venn-diagram for a pitch malfunctioning, reproduced from [73]. The size of the circles is proportional to the conditional probability of two overlapping alarms. The intersections represent the alarms occurring concurrently. In this example, alarm 387 was determined to be the origin and key trigger for subsequent alarms.

1.1.1.4 Computer vision

The ability to assess the condition of a machine, process or structure visually has been a research field of interest for quite some years by now. This has mainly to do with the unique benefits of vision-based monitoring. It offers a possible non-contact, non-destructive, long distance, high-precision, large-range, multi-target solution. Because the visual observations can be made from a distance, it circumvents potential issues such as electromagnetic interference close to the object of interest. The technology behind the vision hardware has also

continuously improved over the past decades which pushed the advancements in this field ever further. A vision-based measurement system is essentially similar in setup as a thermography setup since both involve image acquisition and analysis. This means it generally also consists of a device such as a digital camera, computer and image analysis software.

A very common example for vision-based monitoring is structural health monitoring (SHM) of large structures, e.g. bridges. An example of a typical setup used to measure structural displacements at certain designated locations is presented in Fig. 1.7. The displacements of a structure can be tracked using targets attached to the structure. There are many ways to achieve accurate tracking of these targets, e.g. digital image correlation [78], mean-shift tracking [79], CamShift tracking [80], interferometric phase-shift image matching [81], Lucas-Kanade template tracking [82], robust object search [83], etc. Apart from just measuring the displacement amplitudes, these measurements can also be used for other purposes such as structural strain and stress estimation [84, 85], system identification of the structural dynamics (i.e. natural frequencies, modal damping ratios, mode shapes) [86–91], and crack inspection and characterization [92–96].

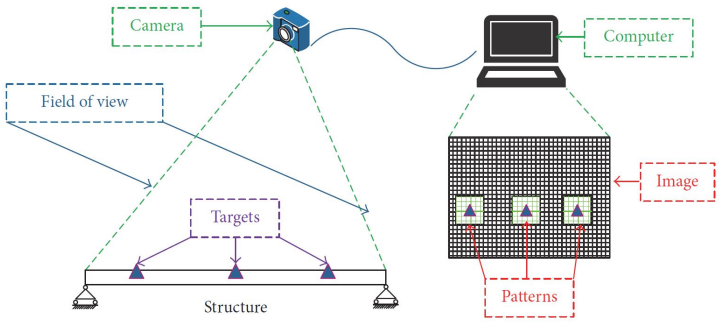

FIGURE 1.7: Example of a general 2D vision-based displacement measurement setup, reproduced from [97].

Besides SHM, other possible examples include the condition monitoring of tools used in manufacturing processes [98], surface roughness monitoring of workpieces [99], or rotation speed monitoring of shafts [100].

1.1.1.5 Acoustic emission & Ultrasound

Acoustic emission (AE) can be described as the radiation of elastic waves in solid structures that result from an internal structural change in the material. This usually occurs as a consequence of plastic deformation or crack initiation. Local sources of elastic waves lead to small surface displacements which can be measured. A common frequency range for AE waves is within the 1 kHz to 1 MHz band. The most popular uses for AE include source localization [101–103], material characterization [104, 105], and health monitoring [106]. Next to non-destructive testing, AE can also be used for process monitoring and in high-pressure fluid situations.

Applications of AE on machinery can include both high and low-speed machinery and in the past decade there has been a considerable amount of research conducted in the field of AE-based machine monitoring [107–112]. Health monitoring of rolling element bearings is a much investigated topic in AE [110, 113–116]. Of particular interest are the papers investigating the potential for AE to detect faults in bearings rotating at extremely low speeds, e.g. below 100 rpm. Vibration analysis has traditionally had a difficult time in reliably detecting these low-speed bearing faults since the fault signatures are badly excited due to the low acceleration levels and thus get buried in the background noise. Several papers investigate the potential of AE to overcome this issue. Jamaludin et al. [117] analyzed AE signals of rolling bearing rotating at 1.12 rpm. Wu et al. [118] examined the AE signature of a large wheel rotating at 0.32 rpm. Many other examples can be found in literature on this topic [119–122]. However, despite all these efforts, low speed damage monitoring is still an ongoing research topic in AE with potentially significant improvements to be made.

1.1.1.6 Electrical signatures

Current and voltage signals can be used for the condition monitoring of electrical machinery. This analysis of electrical currents or voltages to monitor the health of an electrical motor, generator, or transformer is not a new concept. Already in the early 1970s the use of currents was proposed to monitor the motors in nuclear power plants [**13**]. Faults that are often encountered in electrical machinery can generally be categorized as follows [**3**, **13**, **23**]:

- Static or dynamic air-gap irregularities
- Stator faults such as opening or shorting of a stator phase winding coil
- Broken rotor bar
- Cracks in the rotor end-rings
- Faulty connection of the stator windings
- Eccentricity in the shaft
- Bearing failures

Bonaldi et al. [**11**] and Verucchi et al. [**24**] claim that about 40% of defects can be attributed to bearing failures, 30 to 40% to stator faults, 10% to rotor faults, with the remaining 10% being a variety of other defects. A common defect in motors or generators is the presence of air-gap eccentricity. This occurs when the distance between the rotor and stator is not uniform. Figure 1.8 displays the three potential air-gap scenarios. Static eccentricity means that the position of the minimal radial air-gap is fixed, while dynamic eccentricity means that the position follows the rotation of the rotor. Static eccentricity typically leads to sidebands of the grid frequency in the current spectrum, while dynamic eccentricity further modulates the signal with the dynamic rotation frequency of the airgap [**11**]. Many current signature analysis approaches employ frequency analysis to detect faults early on. This approach is also becoming more and more popular in the wind energy sector.

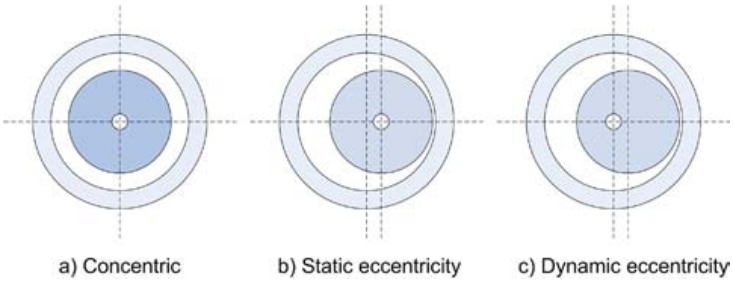

FIGURE 1.8: Example of the three different types of air-gap eccentricity, reproduced from [123].

Due to the rapid growth of the rated power of wind turbines and projects in harsher environments (e.g offshore wind farms), the reliability of the electrical components is seen as a critical issue for wind turbine operation. Power Electronic Converters (PECs) are responsible for 25% of the total failures and 14% of the total downtime in wind turbines [124]. All modern multi-megaWatt wind turbines operate at a variable speed thanks to their PEC, which is either partially power rated (with a DFIG) or fully power rated (normally with a synchronous generator). Nowadays, DFIGs are the most used generator type in high-power variable speed drives in wind energy conversion systems (WECSs) [125]. A diagram of a DFIG topology can be seen in Fig. 1.9. The most common faults in a PEC is a short- or open-circuit of the insulated-gate bipolar transistors (IGBTs) [126]. Normally, they are caused by thermomechanical and electrical stress or by misfiring pulses from the gate driver unit. IGBT short-circuits are destructive in nature and readily detected by hardware protection circuits, leading to the immediate shutdown of the wind turbine. Open-circuit (OC) faults can be caused by the failure of one or more IGBTs, or by the failure of a complete phase either on the rotor or grid side. These faults can remain undetected for a long period time because of the much less severe consequences, increasing consequently the thermal, electrical and mechanical stresses on the wind turbine components. Therefore, their early detection is important for reducing unfavorable loading of the wind turbine drivetrain and thus avoiding secondary faults.

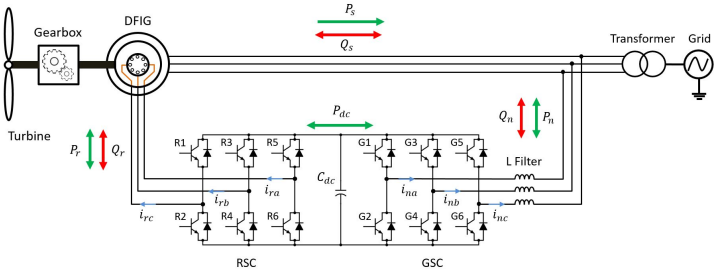

FIGURE 1.9: DFIG wind turbine topology [127].

An example of the impact of an open-circuit fault on the currents of a generator with a synchronous speed of 1500 rpm is shown in Fig. 1.10. Normally these currents should purely sinusoidal. These currents can be measured using sensors like clamp probes, current transformers, or resistors. Different algorithms can then be utilized to process these currents and track the changes in their characteristics. Apart from conventional frequency analysis, other commonly used approaches include Park's vector analysis [128–131] and artificial intelligence techniques [132–134].

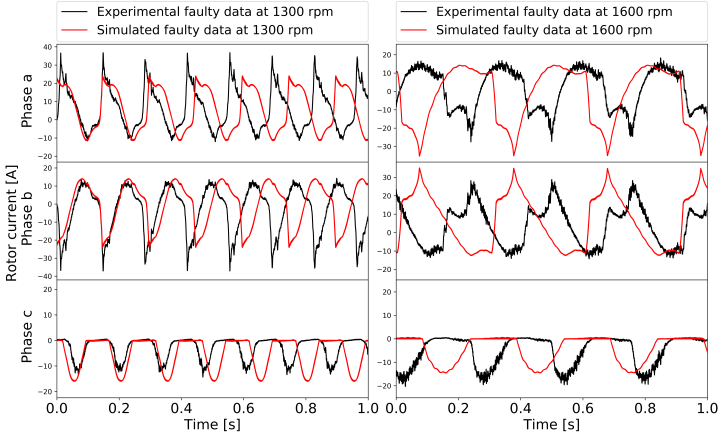

FIGURE 1.10: Example of simulated and experimental data for each phase of rotor currents in a faulty open-switch regime. Left: sub-synchronous speed at 1300 rpm, right: super-synchronous speed at 1600 rpm [127].

1.1.1.7 Vibrations

When components of rotating machinery need to be monitored, vibration analysis is probably the most prevalent approach. This has in large part to do with the fact that it not only allows for analysis of the high-frequency content generally associated with gears and bearings, but is also fairly straightforward to install, proven to be effective, and normally allows to pinpoint faults. These aspects cannot be said for some of the approaches described earlier. For example, the analysis of the lubricants, thermals and SCADA data of a machine normally happens at a low or very low sample rate ($\leq$1 Hz). They also track parameters that can show a significant delay in response to an emerging fault, whilst vibrations react immediately to faults. Computer vision then is not readily applicable inside a gearbox. Acoustic emission is a potential alternative to vibration analysis since it also allows for meaningful high-frequency processing of the machine to detect gear or bearing fault characteristic frequencies. However, as mentioned in Section 1.1.1.5, based on literature, acoustic emission analysis needs additional research to really establish itself as being as effective as vibration analysis. Often the sample rates required for AE are also very high, making a practical implementation and roll-out difficult.

The history of vibration analysis goes back to the late 1970s [**41**]. In particular the aerospace and offshore oil industry were early adopters of vibration-based condition monitoring [135]. In the early days, there was already a large focus on structural health monitoring but the detection of bearing faults quickly gained interest of researchers as well. This is evidenced in Fig. 1.11 and 1.12. These two figures display the occurrence percentage of the bi-grams "structural monitoring,"bearing monitoring, "gear monitoring, and "vibration monitoring in the corpus of English books indexed by Google, from 1900 to 2000 [136]. The offshore oil industry was mainly interested in detecting imminent failure of their expensive drilling and pump equipment [**51**] and according to Farrar et al. [**41**] the most mature and successful application of vibration-based condition monitoring has probably been in the field of rotating machinery [137]. While vibration analysis is employed for both structural health monitoring and machine health monitoring, this dissertation focuses on the latter and more in particular on rotating

machinery.

FIGURE 1.11: Evolution of the occurrence of the bi-grams "structural monitoring,"bearing monitoring, and "gear monitoring in the corpus of English books indexed by Google, from 1900 to 2000 [136].

FIGURE 1.12: Evolution of the occurrence of the bi-gram "vibration monitoring in the corpus of English books indexed by Google, from 1900 to 2000 [136].

1.1.2 Industry 4.0 and Industrial Internet of Things

The wide implementation of the internet all over the world has made continuous data-connectivity possible for a wide range of systems: machines, vehicles, ... Figure 1.13 shows that we are now in the fourth phase of industrial development, namely the one of cyber-physical

systems. A cyber-physical system is a mechanism that is controlled or monitored by computer-based algorithms, tightly integrated with the Internet and its users [138]. Original Equipment Manufacturers (OEMs) and end-users of machines have taken this new opportunity to instrument their machines for condition monitoring purposes. Today, common practice is to perform condition assessment intermittently on the different machines in a process chain. Typically, an engineer goes around in the factory and acquires measurements manually using a hand-held device. Acquisition is done from time to time for those machines that are important but non-critical. Today, permanent monitoring systems only target plant critical machines. Those monitoring systems take measurements automatically at intermittent moments in time. These measurements are then processed automatically on the monitoring device or sent to a central location for interpretation. However, these practices are changing. With the rise of Industry 4.0 and correspondingly the Industrial Internet of Things (IIoT), more and more instrumentation will be available in machines, either directly mounted by the OEM or installed during lifetime by owner-operators. Particularly, the IIoT context is pushing for sensors that are directly connected to the internet. As such, measurements can be transferred continuously to a cloud data-center. These can be raw measurements or data preprocessed at the edge.

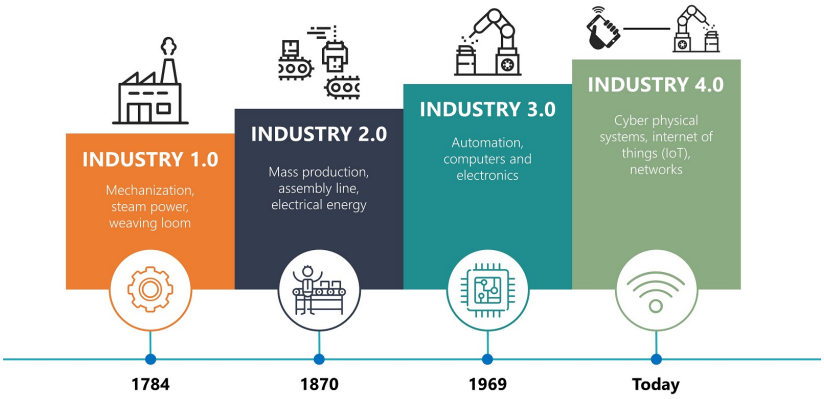

FIGURE 1.13: Overview of the four phases of industrial development, as reproduced from [139].

These large amounts of sensor data can be used for condition monitoring purposes. In contrast to current practices using intermittent data, it will be possible to collect continuous streams of data. Such streams have both a finer granularity and longer signal length. These aspects allow to get a more detailed analysis and get a better grasp on the normal system characteristics and behavior changes a healthy machine experiences. This is particularly interesting for modern machines designed for operation in a wide speed and load range. For these machines, the range of normal operation is widening and thus also to difficulty for anomaly detection to distinguish between normal and abnormal behavior.

Nonetheless, only increasing the amount of data brings no added value. Processing data needs to result in directly actionable information for machine operators. As such integrated condition monitoring techniques are necessary to automatically process the data and combine the health indicators into conclusions about system condition. Long-term data allows long-term trending and anomaly detection on features extracted using signal processing techniques or from machine learning models. To get good insights in system health it is suggested to use a combination of the methods mentioned above and

fuse the results into a limited number of health scores.

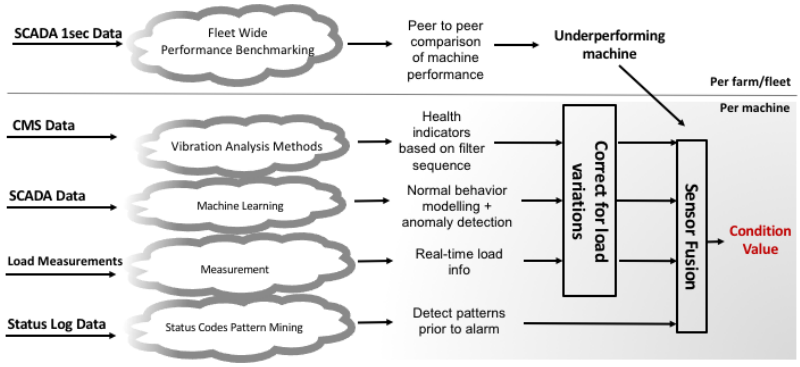

FIGURE 1.14: Approach envisioned at VUB to monitor wind turbine drivetrains in the IIoT context.

Several challenges however need to be overcome in order to be able to fully exploit the opportunities granted by Industry 4.0 and continuous data. First, a balance needs to be found between data transfer to a cloud data-center and local processing on the acquisition device. Second, given that continuous streaming results in large data quantities, there is a need for automated condition monitoring approaches able to autonomously process these large amounts of monitoring data. Third, the IIoT trend makes that a wider variety of sensor data types is available for analysis. Each of these data sources requires a dedicated approach. Moreover, the resulting multitude of indicators from all these analyses need to be fused into a joint conclusion about system health by applying sensor fusion approaches. An example of how we at VUB envision a monitoring approach for wind turbines in the current Industry 4.0 climate is shown in Fig. 1.14.

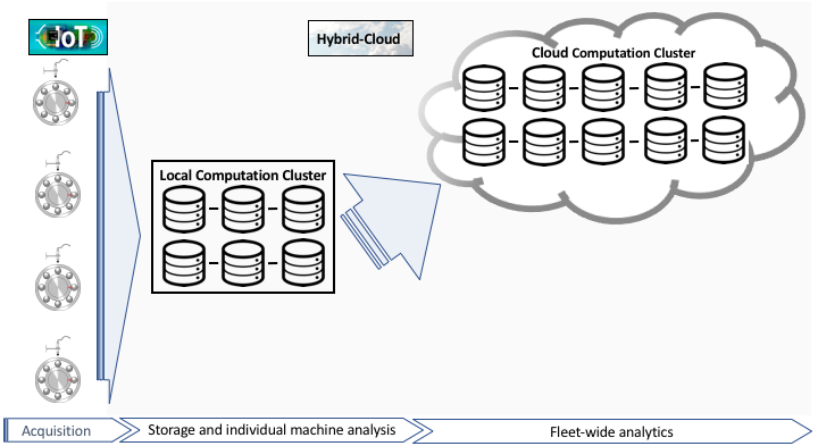

FIGURE 1.15: Overview of cloud approach used at VUB

An important aspect within this IIoT monitoring approach is data storage and cloud processing. In order to be able to integrate the different monitoring approaches, the data needs to be made available in a central data-storage location. Figure 1.15 illustrates the approach that we at VUB are currently using. Sensor data is being collected by a number of sensors. These can be embedded SCADA system sensors or additionally mounted vibration sensors. Sensor data is typically collected by a data-acquisition system and then made available to external data-clients by means of standardized Open Platform Communications (OPC) servers. Typically, this sensor data is stored at a local centralized data-cluster. Recently, there has been a trend to move from traditional historian and SQL databases towards big data storage approaches such as Hadoop or no-SQL databases [140].

Another application that has recently gained significant traction thanks to the IIoT progress enabling it, is the analysis of a whole fleet of machines or assets, so-called fleet monitoring. To get a global overview of asset performance, fleet-wide operational monitoring is done to compare performance of systems that are alike. Different monitoring parameters can be used.

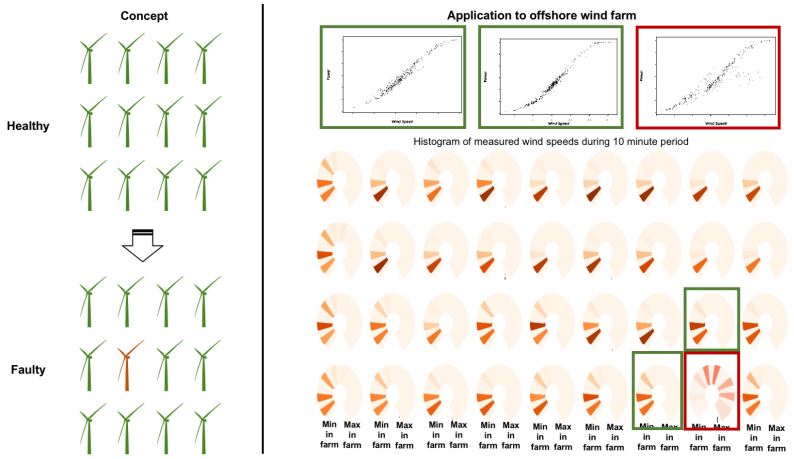

FIGURE 1.16: Fleet-wide benchmarking example of underperformance

Figure 1.16 illustrates some results of a fleet-wide analysis done on a wind turbine farm to detect underperformance. This underperformance is detected by means of turbine-to-turbine comparison of turbines that are closely spaced and where it can be assumed that they are subjected to similar wind loading conditions. The red turbine was detected to exhibit underperforming behavior. The power curves for three neighboring turbines are shown (two green and one red box). These power curves were constructed using the operational data of one day. Based on the comparison of the three power curves (two green and one red) of the neighboring turbines, it can be seen that the red turbine shows significantly different behavior than the green turbines. Particularly for full load conditions the red turbine is underperforming, since a significant amount of operational points (dots in the power curve) is below the expected power curve.

This underperformance could be due to many factors. Further failure mode analysis unveiled that the controller was behaving strangely due to unreliable wind speed measurements. In Fig. 1.16, the histogram of the wind speed is shown for each turbine in the farm. For most turbines, the wind speeds measured by the anemometers on the nacelle of the turbine show consistent results. For the red turbine on the other

hand the wind speed measurements showed high variability. In other words, the turbine controller assumed the wind speeds to be higher than was the case in reality. In this case the turbine was using a control scheme tailored for higher wind speeds (pitch control), which resulted in the turbine underperformance. Fleet-wide comparison was able to unveil this in a straightforward way and can be expanded to other monitoring approaches.

1.2 Research gap

The main objective of this dissertation is the development and validation of new ways to employ vibration signals for the extraction of condition information with a focus on the potential for automation. In particular the different steps involved in going from a raw measured vibration signal to condition indicator are investigated separately. Every step is also illustrated on both simulated and experimental data sets to be as comprehensive as possible with regards to results and performance. During the research for this dissertation, some limitations and/or shortcomings of certain steps were identified in typically used vibration processing schemes. These opportunities for improvement were thus further investigated and also form the basis for the chapters of this document. First, it was observed that vibration-based speed estimation techniques benefit from utilizing the information contained within multiple harmonics and thus this idea was employed to derive a new methodology in chapter 2. Second, removal of deterministic signal content through the manual editing of the cepstrum produced promising results and therefore its automation potential is further investigated in chapter 3. Third, the majority of processing schemes involves some form of filtering step. Looking at the available solutions for blind vibration filtering in literature, it was noticed that many such techniques rely on time-domain statistics such as impulsiveness to recover a fault signal. Knowing that the application domain of cyclostationarity in condition monitoring has greatly expanded in the last decade, chapter 4 looks into a new way of estimating blind filters while making use of the cyclostationary properties of a signal. Lastly, while there is a vast amount of literature available that focuses on the development of individual techniques, there is still relatively little

research that details the combination of multiple advanced processing techniques on measurement data. Therefore, the dissertation is concluded with an in-depth investigation of measurement data using multiple processing steps.

1.3 Research scope

1.3.1 Wind energy

Wind energy is a rapidly growing source of (renewable) energy in Europe. In 2018, about 11.7 GW of new wind energy was installed in Europe [141]. It is nowadays the second largest form of power generation capacity in the EU with 189 GW, just behind natural gas (See Figs. 1.17 and 1.18), and it accounts for about 19% of Europe's total installed power generation capacity. It is therefore safe to say that wind energy is not going anywhere anytime soon and that investments in its progress will continue for quite some time still. The latter is also confirmed by Fig. 1.19 where it can be seen that most EU countries are investing heavily in expanding their wind energy capacity. Alongside this expanding capacity, the interest in cutting the Operating Expenses (OPEX) of wind energy also increases.

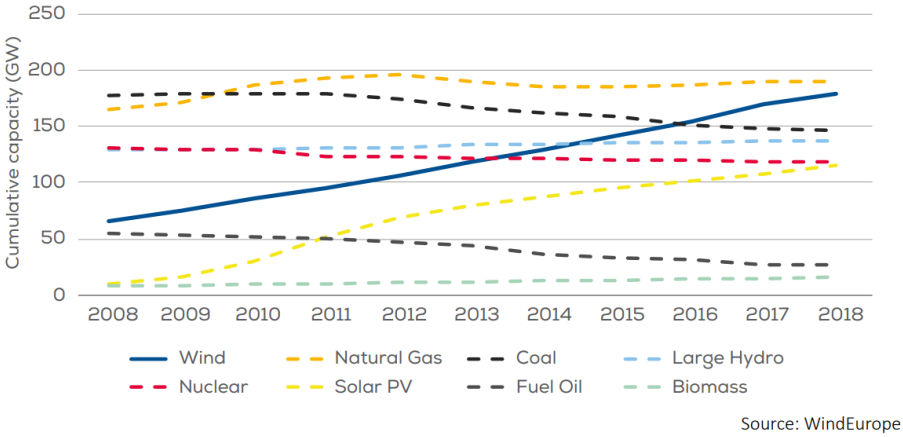

FIGURE 1.17: Total power generation capacity in the European Union from 2008 to 2018, as reproduced from [141].

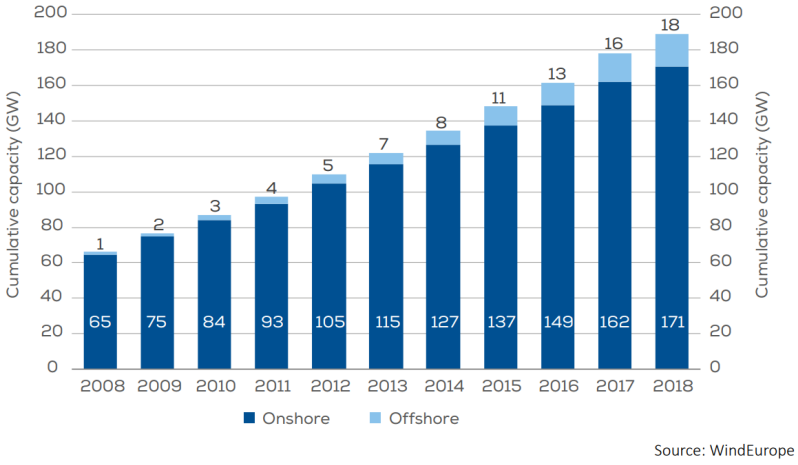

FIGURE 1.18: Cumulative onshore and offshore installations in Europe from 2008 to 2018, as reproduced from [141].

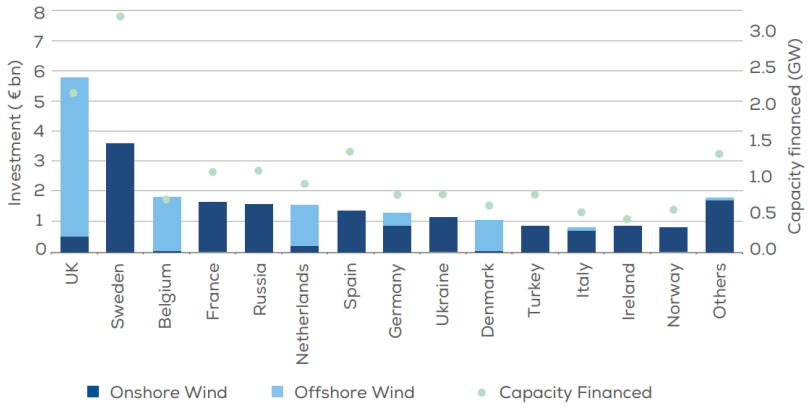

FIGURE 1.19: New financing in 2018 by country (expressed in billion € and GW), as reproduced from [141].

Wind turbines are exposed to strong dynamic excitation events such as varying wind speeds, electricity grid events, or sea waves (for offshore wind turbines) [142]. The production cost of electricity by wind turbines is strongly influenced by the reliability of the wind turbine systems [143]. The downtime and repair costs significantly contribute to the economic impact of faults. In particular gearboxes in wind turbines exhibit substantial downtime in case of failure [144] and bearings turn out to be the most critical component in wind turbine gearboxes [145, 146]. In general, rolling element bearings are one of the most used components in wind turbine drivetrains.

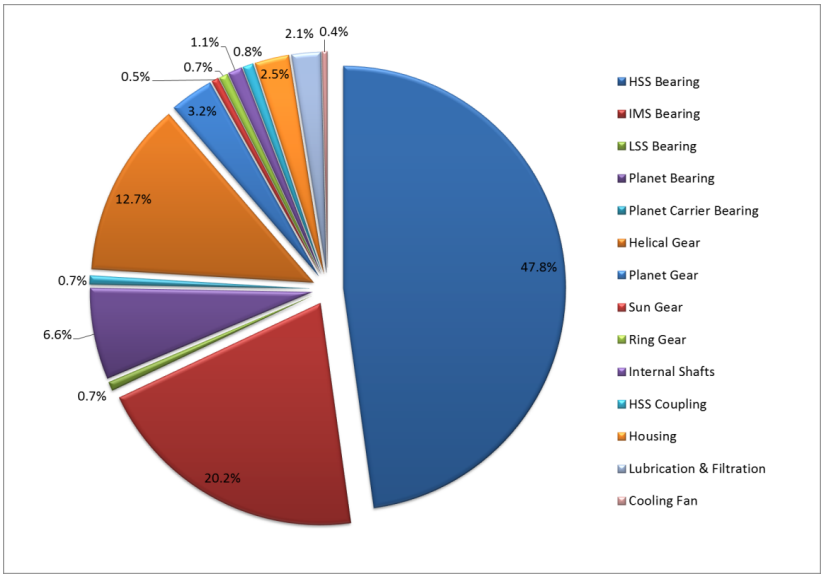

FIGURE 1.20: Diagram with failure statistics based on the NREL Gearbox Reliability Database, as reproduced from [147].

According to the statistics provided by the NREL gearbox reliability database (shown in Fig. 1.20), more than half (76%) of wind turbine gearbox failures are caused by bearings, with gear faults being the second major cause for failure (17.1%) [147]. It is estimated that overall more than 90% of all rotating machines [148] contain rolling element bearings. Unfortunately, they are susceptible to a multitude of premature deficiencies and less than 10% of rolling element bearings reach their expected basic L10 life, the life at which ten percent of the bearings can be expected to have failed due to normal fatigue failure for that particular application. These observations imply a need for an improved comprehensive condition-based maintenance program. However, there are still some hurdles to be overcome before such an exhaustive program becomes fully feasible. This dissertation regularly uses the example of wind energy to illustrate the performance of methods on because the wind energy industry is currently one of the most interested parties in improving their reliability through condition monitoring and because wind energy applications form an interesting

research challenge due to the varying load and speed.

1.3.2 Other industries

Beside the wind energy application, there are many other industries that are looking at condition monitoring and predictive maintenance to improve the reliability of their machines and reduce operational costs. Common examples of other industries where condition monitoring is employed are the Hydrocarbon Processing industry (Oil & gas / petrochemical), (hydroelectric) power plants, food processing, transportation (planes, helicopters, trains, cars), mining, farming equipment, etc. All these applications utilize rotating machinery in some way and typically the used equipment is expensive to replace and brings about a significant downtime cost when they halt production. These two reasons are typically one of the main reasons to install condition monitoring tools. In the foreseeable future however, it is very likely that condition monitoring solutions will not only be limited to big, expensive machinery, but will also expand its usage to cheaper industrial equipment and potentially even to everyday household appliances.

1.4 Research challenges

Analyzing vibrations in search of information about a machine's condition is a multi-faceted problem and has been approached by researchers in many different ways. Retrieving diagnostic information for a general case is not that simple though since the processing procedure necessary to go from a raw vibration signal to a possible diagnosis is not straightforward. Lately, there has been an ever-increasing amount of attention towards vibration-based condition monitoring of machines, and correspondingly there has been an increase in contributions towards tackling the problem of designing such a processing scheme. While some researchers focus on just a single step in this usually multi-step process, others focus on the entire methodology in itself.

While there are many analysis procedures out there tackling the analysis issue in different ways, the scheme utilized in this dissertation consists out of five steps:

1. Instantaneous rotation speed estimation & compensation of its fluctuation
2. Deterministic and stochastic signal content separation
3. Filter design to increase the fault signal-to-noise ratio
4. Residual signal analysis
5. Indicator calculation and "*smart*" tracking

Ideally, an end-user of this processing approach would use a black box procedure with the raw vibration signals as input and diagnostic information about the machine as output. Attempts have already been made in the past to incorporate several techniques into one method chain in order to offer a more elaborate and/or automated analysis of vibration measurements. Randall & Antoni proposed a semi-automated procedure for bearing diagnostics using a combination of order tracking, discrete component removal, minimum entropy deconvolution, spectral kurtosis and envelope analysis [149]. A great deal of researchers also try combining several machine learning methods (e.g. artificial neural networks, support vector machines, genetic algorithms, etc.) in order to improve the robustness of automated fault detection [150]. In theory, a lot of method combinations are conceivable and deciding the optimal one will remain an open question for quite a while most likely. The proposed five-step processing scheme is introduced further in the following sections and each step serves as a standalone chapter in which it is illustrated in more depth.

1.4.1 Rotation speed estimation

Instantaneous rotation speed estimation has become a key part of many condition monitoring procedures for rotating machinery. The ability to track the rotational speed of a system is a critical requirement for the majority of vibration-based condition monitoring methods. Information about the speed enables compensating for potential

speed variations that would otherwise impair conventional frequency-based methods. The problem of instantaneous speed estimation based on the vibration signals themselves is one that has received a significant amount of attention in recent years. Installing encoders or tachometers has become a lot less attractive due to the potential cost savings that can be obtained by simply utilizing an accelerometer instead. However, trying to find a speed estimation method that fits a certain application "best" is not so straightforward if one inspects the available literature. It turns out that there are many articles that present slight variations or extensions to already existing techniques.

The importance of an accurate speed estimation scheme is underlined by the existence of a special issue on Instantaneous Angular Speed processing and angular applications in the journal of Mechanical Systems & Signal Processing [151] and of a conference dedicated solely to condition monitoring in non-stationary operations, coined CMMNO [152]. There is also quite a large body of literature on instantaneous frequency estimation in the general field of signal processing. However, the estimation of instantaneous speed from vibrations specifically addresses a more difficult problem, which often jeopardizes the use of these general signal processing techniques. Typical problems that can occur for the case of instantaneous frequency estimation from vibration signals are:

- Extremely low signal-to-noise ratio
- Strongly colored noise
- Harmonic interference (a special case of the former point)
- Amplification by resonances (e.g. band-pass filtering in signal processing terms)
- Non-persistent and fading harmonics
- Presence of multiple harmonics
- ...

Currently, the main benefit of using a direct speed measurement approach, i.e. an angle encoder, as compared to an indirect approach,

i.e. vibration-based speed estimation, is still the accuracy. When using an indirect method such as vibration-based speed estimation, the analysis is often confronted with several interfering sources due to the nature of the vibration measurements, e.g. resonances, non-speed related harmonics, noise, etc. These effects can have a detrimental effect on the accuracy as compared to angle encoders. On the other hand, vibration sensors are often cheaper and easier to install than angle encoders. Sometimes it can be unfeasible to instrument the shaft of interest with a standard encoder when the shaft is too large or not accessible. Also the data acquisition system needs to be capable of reading in and processing the encoder data stream and often requires dedicated measurement hardware. The capability to simultaneously measure the vibrations for condition monitoring purposes and estimate the speed based on those vibrations allows to circumvent the encoder-related practical inconveniences and costs.

This dissertation and more in particular Chapter 2 focuses on estimating the IAS based on the information contained within the vibration signal. While it is possible to utilize the IAS itself as a form of indicator for the condition of the machine, this typically involves installing a high-resolution encoder [153–155] in order to reach the required accuracy of the IAS to be able to detect small faults. Analyzing signals in the angular domain is one of the main reasons why most condition monitoring procedures include speed estimation. Chapter 2 discusses the state-of-the-art methods in vibration-based speed estimation and compares the performance of some of these methods on experimental data sets. Additionally, this chapter discusses a novel speed estimation method that shows very promising results with regards to accuracy and robustness.

1.4.2 Discrete-random signal separation

A large entity of condition monitoring research investigates the separation of signal mixtures into sub-components originating from different sources. An important step in this respect is the separation of stochastic signal content from deterministic content. The former is considered to entail signals such as measurement or process noise from pumps, turbines, but also cyclostationary signals such as rolling

element bearing faults. The latter often consists of the gear or shaft harmonics. This separation step is considered to play a significant role in the proper detection of mechanical faults.

Bearing faults are typically detected by analysis of the measured signal's envelope spectrum [156–158]. Unfortunately its presence in the envelope spectrum is often masked by high-energy deterministic components [159]. Separating the bearing faults from the masking signal content assumes that a bearing fault signal is stochastic due to random jitter on the fundamental period of the fault frequency [160]. This corresponds to the random slip of the rolling elements and a bearing signal can be considered to be second-order (quasi-) cyclostationary. These effects cause a smearing of the bearing frequencies in the amplitude spectrum, while the deterministic signals manifest themselves as discrete peaks [161]. This property forms the basis for many stochastic-deterministic separation methods. Example methods that were proposed in the past are the discrete/random separation (DRS) technique [162] (see Fig. 1.21 for an example), self-adaptive noise cancellation (SANC) [163], linear prediction filtering (LP) [164], and the (generalized) time synchronous average (TSA) [165, 166]. Another method that has proven its worth in this regard is the cepstrum editing approach [167].

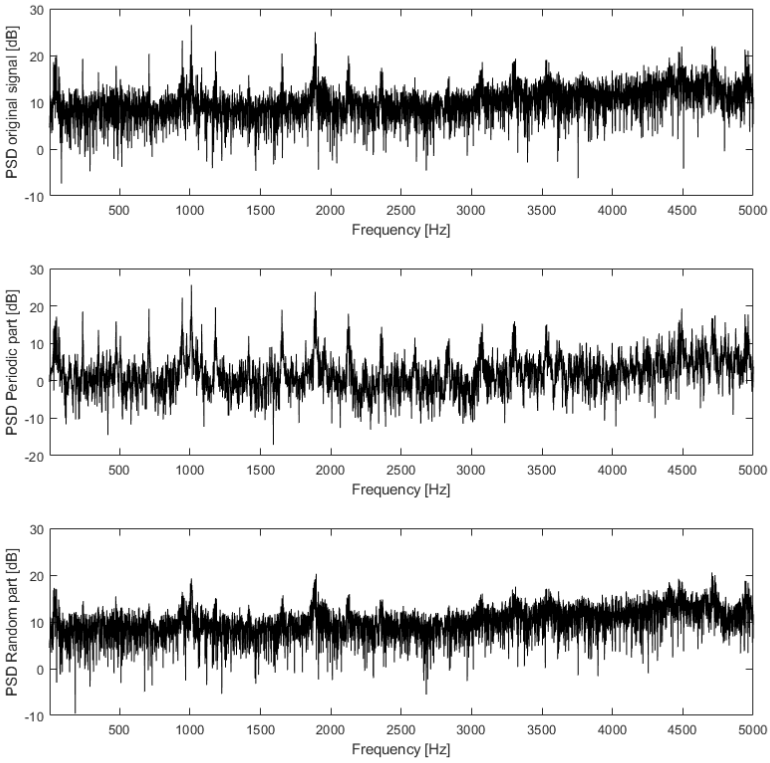

FIGURE 1.21: Example of signal separation using DRS [162] on an experimental gearbox vibration signal (data from [168]), (Top) Original signal spectrum (Middle) Deterministic signal spectrum (Bottom) Stochastic signal spectrum.

Cepstrum analysis is a signal processing tool with already quite a lengthy history. Bogert et al. [169] introduced the cepstrum first in 1963 as "the power spectrum of the logarithmic power spectrum with the purpose of detecting echoes in seismic signals. Since then the description of the cepstrum has evolved and other definitions for the cepstrum have been identified afterward, like the "complex cepstrum by Oppenheim and Schafer [170, 171]. Various discussions were held

about the proper definitions and properties of the cepstrum [172, 173] and different forms were described like the differential cepstrum [174] and the mean differential cepstrum [175], which have their use mainly in operational modal analysis (OMA). Recently however the cepstrum has come into the limelight for its use in the field of vibration-based condition monitoring. It turns out that the cepstrum can be used in an efficient and effective manner as well to separate deterministic from stochastic content [167]. This separation property of the cepstrum is further investigated in Chapter 3.

1.4.3 Signal filtering

Vibration signals normally contain a mixture of different signal components originating from many different sources. On top of that, the source signals get distorted by the transmission paths from the vibration sources to the transducer that actually measures the resulting vibration. The latter effect has a considerable influence on the measured signature of impulsive signals, e.g. coming from impacts due to local defects such as spalls or pitting in gears or bearings. Also, if the impulse response function (IRF) is shorter than the spacing between the impulses (which can be the case for high-speed machines), it can prove difficult to find the repetition frequency of the impulses, e.g. through envelope analysis.

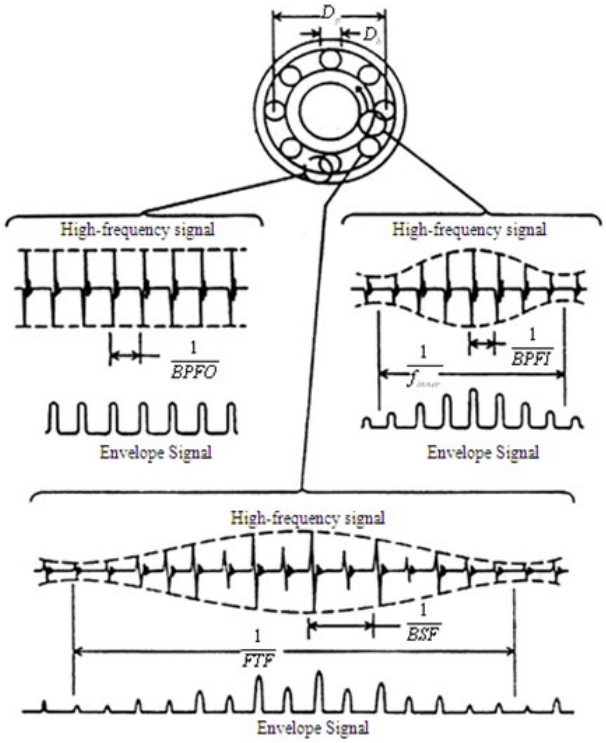

FIGURE 1.22: Typical impulse train signatures for the four different bearing fault frequencies (FTF, BSF, BPFO & BPFI), as reproduced from [176]).

Therefore, in order to increase the signal-to-noise ratio of the fault signature, an additional filtering step is often introduced after the signal separation. This filtering step often revolves around finding a filter that optimizes a certain criterion that characterizes the fault signature of interest. In the case of impulsive events such as a bearing fault, the source signal is typically assumed to be an impulse train at the repetition frequency of the bearing characteristic fault frequency, as shown in Fig. 1.22. The basic idea employed to find this source signal is then to find an inverse filter that nullifies the transmission path by maximizing the impulsiveness or kurtosis of the filtered signal. Besides maximizing the kurtosis, there are many possible optimization criteria that can be used to find a filter. Chapter 4 discusses this broad topic of signal

filtering approaches.

1.4.4 Fault detection & diagnosis

Probably the only step that is used in every single condition monitoring procedure out there is the step of fault detection. This is easily explained by the fact that every measured signal has to be processed in some way to determine whether or not the signal characteristics have changed as compared to the past. Only after this step of fault detection, the next one of diagnosis can be taken. The concept of fault detection is quite wide-ranging and there exist many potential ways of identifying discrepancies in the signal characteristics.

The two most commonly used approaches are tracking the statistics of the time domain waveform and the frequency spectrum of the signal. In the early days of condition monitoring, it used to be standard practice to just calculate the vibration energy levels and compare them to standardized criteria for vibration severity [177, 178]. These criteria were originally developed using subjective empirical observations however, often even observed by touching the machine with a finger and assessing the degree of vibration severity [176]. It was quickly realized that a more accurate means of assessing machine condition could be achieved through the use of frequency analysis. Initially, the use of difference spectra and one-to-one comparisons was popular for tracking the condition. Gradually, the number of techniques used in the frequency analysis also expanded and methods such as time-frequency diagrams and phase analysis were employed to detect defects [179, 180]. Nowadays, there are many potential ways to obtain a time-frequency representation (TFR) of a signal, typically divided into linear and quadratic or bilinear categories. Example TFR techniques include the short-time Fourier transform (STFT), the wavelet transform (WT), their synchrosqueezed variants [181, 182], the Wigner-Ville Distribution [183], the Choi-Williams distribution [184], the Zhao–Atlas–Marks time-frequency distribution (also known as the cone-shape distribution) [185], etc. In general though, it is important to note that each one of the TFRs still has to respect the uncertainty principle, also called the Gabor limit, stating that the product of the standard deviation in time and frequency is limited. It is possible to obtain a higher

resolution, but at the cost of different signal components interfering with each other [186].

Besides time-frequency analysis, also the analysis of the phase of signals became a staple in many condition monitoring schemes. It turns out that many defects in rotating machinery (e.g. related to the combustion or pressure in cylinders and pumps) induce perturbations in the torsional vibration of shafts. Frequency demodulation is therefore used on shaft encoder signals to detect these perturbations.

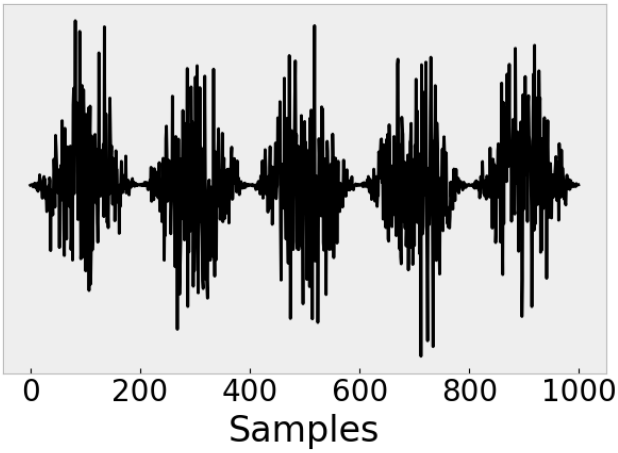

FIGURE 1.23: Example of a second order cyclostationary signal (i.e. amplitude-modulated white noise).

In recent years, a relatively new concept to condition monitoring has received a lot of attention in research, namely the concept of cyclostationarity. While Chapter 3 focuses on the separation of deterministic and stochastic separation of signals, it is possible to also discriminate cyclostationary signals from other signal content. Bearing faults for example typically introduce second-order cyclostationary content in the signal, meaning their second order statistics are periodic (i.e. they have a periodic autocorrelation). An example of a second-order cyclostationary signal is shown in Fig. 5.3. Chapter 5

discusses the different ways signals can be processed using the mentioned statistical, spectral, and cyclostationary approaches and how the resulting information is translated into condition indicators.

1.4.5 Towards automation

Thanks to the decreasing cost in sensors and data acquisition systems and the ever-increasing computational power, it has become feasible to do extensive measurement campaigns where vibrations are acquired on multiple channels and machines simultaneously. The difficulty however is in the fact that all this acquired data needs to be processed in a meaningful and efficient manner. Nowadays it is not practical or possible anymore to have an expert dedicated to only looking at analysis results and condition indicators continuously. This excess of data necessitates profound automation of the analysis procedure and the interpretation of the analysis results. On top of that, experimental vibration data coming from a machine that operates in a highly dynamic context (e.g. a wind turbine) normally shows a large dependency on its operating conditions. This dependency can invalidate standard condition monitoring approaches such as statistical thresholding of the indicators. Therefore, the calculated condition indicators also need to be compensated for this dependency such that the expected behavior for every operating regime can be learned and utilized for anomaly detection. Chapter 5 thus also goes into detail about how the analysis results can be summarized into a form that is workable for a human mind and independent of operating conditions.

1.5 Original contributions

This dissertation discusses many of the state-of-the-art methods commonly used in vibration-based condition monitoring, but this section details the original and novel contributions of this dissertation.

1.5.1 Contributions related to rotation speed estimation

A major part of the PhD research was dedicated to investigating instantaneous speed estimation methodologies. While initially the research started with implementing and comparing existing speed estimation techniques, the research afterwards progressed towards the development of an entirely novel method for rotation speed estimation, namely the **M**ulti-**H**armonic **D**emodulation (**MHD**) method. This technique uses the concept of single harmonic phase demodulation and extends it to allow for simultaneous demodulation of multiple harmonics present in the signal. Considerable time and effort was also spent on improving the reliability of the speed estimation results of this method by developing a rigorous manner in which to weight the different harmonics based on their signal-to-noise ratio. This was achieved through a maximum-likelihood estimation approach. Based on the assumption that most rotating machines introduce many well-excited and speed-related harmonics into the signal, this method has the potential to often be the most accurate rotation speed estimation method available. This can mainly be attributed to the usage of the information contained within multiple harmonics instead of just one single harmonic.

A special notice should also be given to the **M**ulti-**O**rder **P**robabilistic **A**pproach (**MOPA**). This method, originally developed by Leclère et al. [187] in 2016, was not only the subject of continued collaboration between INSA Lyon and VUB, but also the subject of further investigation into how the method performs on experimental wind turbine data and how it could be improved.

Lastly, the acquired findings and experience in speed estimation are discussed in a detailed comparison of several advanced speed estimation techniques on experimental data sets. This examination of different methods was mainly done due to the lack of an extensive and quantifiable review of the methods available in literature. Therefore this last contribution is intended to be useful to both newcomers and more experienced researchers in the field of vibration-based rotation speed estimation.

1.5.2 Contributions related to discrete-random signal separation

A strong focus at the start of the PhD research also went to the further development of an automated discrete-random separation methodology. To achieve the goal of signal separation into deterministic and stochastic content, the cepstrum editing procedure, originally proposed by Randall and Sawalhi [167] in 2011, is automated and its performance is investigated comprehensively on simulated and experimental data.

1.5.3 Contributions related to signal filtering

While many different filtering approaches were investigated during the course of the PhD, the main novelty is the development of a new blind filtering framework from a cyclostationary perspective. In the past, blind filters such as Minimum Entropy Deconvolution (MED) [188] and its derivatives [189] used statistics of the time waveform to find an optimal frequency filter. In contrast, the methodology proposed in this dissertation employs the assumption that most faults (especially bearing faults) introduce second order cyclostationary content in the signal and uses that knowledge to derive filters that maximize in a blind manner this cyclostationarity. Especially the usage of an iterative filter updating procedure with the generalized Rayleigh quotient is highlighted since it allows for a quick and effective computation of the filter and a very flexible adaptation to many other optimization criteria.

1.5.4 Contributions related to fault detection, diagnosis, and data interpretation

The main contribution here is the development of a fully automated and scalable monitoring approach using the developed advanced processing methods. The contribution in itself can be considered to be more a collection of different techniques than just a single technique. Due to the vast amount of measurement data that Industry 4.0 brings with it, it has become paramount nowadays to have an efficient means to analyze all that data. This dissertation details how to process and

interpret all those measurements using a hybrid approach consisting of physics-based signal processing and data-driven models.

1.6 Outline of the dissertation

Chapter 1 gives a general introduction to the different types of condition monitoring and describes the benefits and drawbacks of each. Special emphasis is placed on the emergence of Industry 4.0 and the application of condition monitoring to wind energy and other industries. Lastly, the different research challenges and objectives are highlighted.

Chapter 2 discusses the state-of-the-art in vibration-based instantaneous angular speed estimation and proposes a novel method for an accurate multi-harmonic phase demodulation procedure. This new method is fully derived theoretically and afterwards validated on both simulated and experimental data. Additionally, a thorough comparison is made of several different advanced speed estimation methods on three different experimental data sets.

Chapter 3 details the state-of-the art in signal separation techniques such as time synchronous averaging and investigates the performance of the automated cepstrum editing procedure (ACEP) on both simulated and experimental data.

Chapter 4 elaborates on the topic of indicator tracking and using these indicators for filtering approaches. Also a new methodology is defined for deriving blind filtering approaches using complex optimization criteria based on the cyclostationary properties of the signal. This new framework is illustrated for different indicators based on predictability or sparsity of the envelope signal or envelope spectrum respectively. Validation is again done on both simulated and experimental data.

Chapter 5 provides an overview of common methods for fault detection and diagnosis such as spectral and envelope analysis. It also elaborates on the cyclostationarity concept. Finally, it proposes a hybrid physics-based and data-driven monitoring approach.

Chapter 6 illustrates and validates the processing methods discussed in this dissertation on experimental data sets. Emphasis is put on the proper combination of the different tools and on the validation of their performance. Two case studies showcase the developed methods on vibration data measured on wind turbine gearbox housings.

Chapter 7 completes the dissertation with general conclusions about the described state-of-the-art methods and the newly proposed ones. Ideas for future research challenges are discussed together with suggestions for improvements of the proposed techniques.

Chapter 2

Rotation speed estimation

2.1 Introduction

Accurate knowledge of the rotation speed of a machine is necessary to gain insight in the mechanical signatures of the machine components and it is of vital importance to the majority of leading-edge vibration-based condition monitoring techniques. Many rotating machines indeed do not operate at a constant speed but operate at varying speed regimes. Such non-stationary conditions require an efficient and robust way of estimating the instantaneous angular speed (IAS) in order to not invalidate techniques such as order analysis and angular resampling. While in the past this information would typically be measured using an angle encoder on one of the rotating shafts [190–192], recent research efforts concentrate primarily on extracting the IAS directly from the vibration signal due to the potential cost savings or installation problems.

In industry currently, it is still common practice to use rotary encoders to estimate the speed due to the reliability and the accuracy. A whole range of rotary encoder types exists. Incremental encoders are rotary feedback speed and position transducers that produce coded output, typically in the form of a pulse train with a frequency proportional to the rotation speed. Absolute encoders do not produce pulse trains, but generate binary code that allow estimating the shaft position accurately. Typically, absolute encoders operate using an optical method, such as light sources and photo-sensors in combination with a disk made of regions with different transparency. A downside of this approach is that the photosensors must be mounted close to

the rotor disc, and a small tolerance is required at the relative sensor positions [193]. If the working distance of the photosensors can be increased from the actuators, it leads to more flexible applications and this is where vision-based solutions come in.

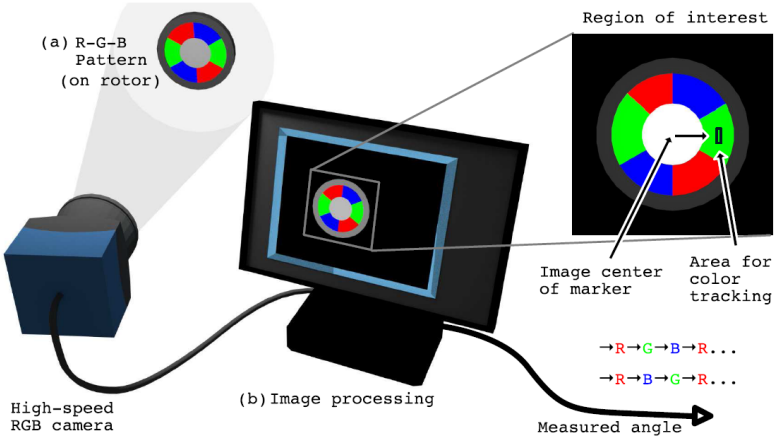

FIGURE 2.1: Example concept of a visual encoder where an RGB pattern is used to track the rotation speed through image processing, as reproduced from [193].

Computer-vision algorithms have been used in the past for visual tracking of an object [194], examples are techniques such as SIFT [195] and SURF [196]. Vision-based methods are applicable for objects positioned relatively far from the observer and thus do not have the same structural limitation as standard optical encoders. Nowadays, there are quite a few publications available that investigate rotation speed measurement using vision-based methods [197–201]. Each paper suggests different approaches in tracking the rotation speed and angle, but most of them are based on tracking high-contrast markers or speckle patterns (as shown in Fig. 2.1). Typically this involves using expensive camera equipment capable of filming at very high frame rates. Recently, Zhong et al. [100] showed that artificial fringe patterns and linear array sensors can be employed to estimate with high

precision the instantaneous rotational speed. Following the Industry 4.0 trend, there is an increasing focus on cost efficiency also in this research field. Wang et al. [202, 203] show that vision-based rotation speed estimation is feasible using a low-cost imaging device. Their methodology is based on tracking the degree of similarity, quantified by a two-dimensional correlation, between consecutive images. The periodicity of the low-resolution similarity can then be exploited using the Chirp-Z Transform. While these low-cost, vision-based encoders do not achieve the same accuracy yet as the standard optical encoders, this deficit will certainly decrease in the future when imaging sensors become even cheaper and algorithms more optimized.

While it is possible to install an encoder or tachometer to measure the IAS directly on the machine, this imposes an additional cost for the machine operator since a technician needs to dedicate time and resources to the instrumentation. Because of this increased cost, there has been a significant amount of research done on the estimation of the instantaneous angular speed from the vibration signal itself. Even though there have been many publications in the past decade detailing various “encoder-less” or “tacholess” speed estimation methods, most of them can be categorized into a limited number of groups. Therefore, the following section gives a brief summary of the state-of-the-art in vibration-based speed estimation.

2.2 Overview of state-of-the-art

Tracking in a TFR A large number of contributions focus on extracting speed information by tracking harmonics in a time-frequency representation (TFR) of the vibration signal with varying degrees of complexity. Typical examples of such TFRs are the short-time Fourier transform (STFT), the wavelet transform, or a Wigner-Ville distribution. Urbanek et al. [204] use a simple maximum tracking in the spectrogram in a comparison with phase-based demodulation methods. Furthermore, multiple research papers investigate the possibility to improve

the STFT for instantaneous frequency tracking by increasing the resolution when necessary. Kwok et al. [205] implement an adaptive short-time Fourier transform that chooses the optimal window parameters based on an entropy concentration measure of the STFT. Cheung et al. [206] use two kernel functions of different supports to obtain a wideband and a narrowband spectrogram. To preserve the localization characteristics, they implement a combined spectrogram using the geometric mean of the corresponding STFT amplitudes. Peng et al. [207] use a chirplet transform with a polynomial kernel to extend the standard chirplet transform to non-linear IAS estimation with improved frequency resolution. Sekhar et al. [208] investigate the effect of interpolation on the polynomial Wigner-Ville distribution and IAS estimation. An alternative to the STFT that is often used when more flexibility regarding frequency and time resolution is required, is the wavelet transform. Gryllias et al. [209] use complex shifted Morlet wavelets to find the optimal shift and bandwidth from which to determine the instantaneous speed. Current signals can also use this type of techniques to determine the speed since they are very similar to vibration signals. Aller et al [210] use an analytic wavelet transform on the stator current signal of an AC machine combined with a simple ridge tracking algorithm.

A fairly recent development is the synchrosqueezing transform. In 2009, Daubechies et al. [181] proposed this new technique in the context of audio analysis. Generally, it can be considered as a special type of reassignment method, similar to what had been done before for the STFT and other conventional TFRs [211, 212]. The purpose of this type of reassignment techniques is to improve the concentration of signal components in the time-frequency plane making them more suitable for visual analysis and for methods such as ridge extraction. Correspondingly, there has been a surge in papers making use of the advantages of this new technique [213–215]. Xi et al. [215] propose a frequency-shift synchrosqueezing algorithm to generate the TFR that afterward gets used as input for the Viterbi algorithm to find the IAS.

Shi et al. [213] employ a step-wise demodulation transform in combination with STFT-based synchrosqueezing to determine the IAS. However, there is still strong concern about whether or not synchrosqueezing actually increases the reliability of speed tracking on noisy multi-component vibration signals [216–218] compared to an STFT or Wavelet transform.

Other papers focus more on improving the tracking in a TFR rather than improving the TFR itself. Barrett et al.[219] use Hidden Markov Models to incorporate probability in the evolution of the instantaneous frequency. Schmidt et al. [220] incorporate a priori probabilistic knowledge about the instantaneous frequency of the system to increase the robustness of the maxima tracking in the STFT. Quite a few papers focus on implementing a robust ridge detection scheme to track the IAS [221, 222]. The TFR used for ridge detection varies based on the application and sometimes also based on the preference of the authors, but most of them take into account some a priori knowledge about the physical system under investigation. Wang et al. [222] propose the non-linear squeezing time-frequency transform in combination with ridge detection to estimate the IF. Iatsenko et al. [182] use an improved dynamic path optimization method to efficiently estimate the candidate path that best represents the IAS. They also investigate the applicability and performance of their method on different TFRs such as the STFT, the wavelet transform, and their synchrosqueezed variants.

Some techniques try to utilize more than just one single harmonic present in the signal and its TFR. These methods try to make full use of all the mechanical events linked to the speed, e.g. all synchronous gear and shaft harmonics in a gearbox. Often in complex rotating machinery, there are different operating regimes, potentially leading to other harmonics to be excited or to significant amplitude differences of the tracked harmonics. Therefore, the idea of using multiple harmonics makes sense for the purpose of increasing robustness. For example Zimroz et al [223] divide the STFT into different frequency sub-bands belonging to different harmonic orders and detect the instantaneous meshing frequency for each of the orders by modeling

the noise levels and using a threshold based on the spectral kurtosis. Afterward they perform an averaging of the different normalized instantaneous meshing frequencies to obtain a mean estimate of the speed. A fairly recent development is the multi-order probabilistic approach [187, 224]. This approach does not require a priori knowledge about the exact harmonic related to a certain periodic mechanical event but views the STFT as a probability density function (pdf) map of the speed. These multi-harmonic methods perform particularly well when the harmonic structure of the signal is well known and excited.

Phase demodulation Another major group of speed estimation methods makes use of band-pass filtering and phase demodulation based on the analytic signal. Usually one speed-related harmonic is selected based on its signal-to-noise ratio (SNR) and then used for phase demodulation after band-pass filtering around that harmonic [204, 225–227]. There are also multiple different ways to obtain the demodulated phase. Bonnardot et al. [225] introduced the standard demodulation approach as described above in order to angularly resample the vibration signal and they indicate some considerations and limitations about the technique. Boudraa et al. [228] use the Teager Energy Operator in combination with Empirical Mode Decomposition to obtain instantaneous frequency (IF) estimates for every Intrinsic Mode Function (IMF). This approach does rely on the assumption that each IMF corresponds to a band-pass filtered, IF-related harmonic. A unique approach related to demodulation was proposed by Randall et al. [229] based on a new interpretation of the Teager Kaiser Energy Operator (TKEO). It avoids issues one might get with unwrapping the phase in the standard phase demodulation approach and is based on utilizing amplitude demodulation in the form of the squared envelope of the band-pass filtered signal.

Other approaches Lastly, there are a number of lesser used techniques for speed estimation. Cardona et al. [230] use a square-root cubature Kalman filter to estimate the speed and for order tracking of the

signal. They estimate the number of orders necessary for the Kalman filter based on the number of high amplitude harmonics. Scala et al. [231] use an extended Kalman filter non-stationary sinusoid tracker (EKF-NST) that also allows for both frequency estimation and order tracking. The downside of most model-based approaches is that they require a lot of input parameter tweaking which is often difficult to automate. Another unique approach is based on the scale transform. Combet et al. [232] use a short-time scale transform to estimate the instantaneous speed relative fluctuation based on the varying time-scale factor along the vibration signal.

2.2.1 Research gap

Based on the available literature, it is clear that there are a lot of possible variations on the base methods out there, resulting in often very similar speed estimation techniques. This chapter starts off by proposing a promising new method belonging to the group of demodulation -based methods and validating it on simulated data in section 2.3. Afterward, the advantages, disadvantages, and performance of the described different types of techniques are assessed in detail. Since it is impossible to implement and test all of the different method variants, a selection is made from the different groups mentioned in section 2.2 to provide a qualitative comparison. The techniques are examined on three experimental data sets originating from three different machines, namely a wind turbine gearbox, an aircraft engine and the generator of a ship.

2.3 Multi-harmonic demodulation

Speed estimation through phase demodulation has proven itself already multiple times in the past to be a fairly reliable way to obtain accurate instantaneous angular speed estimations [204, 225–228]. It only requires a single well-excited and speed-related harmonic order with preferably no crossing orders and it is also very easy to implement. Despite its popularity, the possibility to employ more than one harmonic simultaneously in the demodulation has not yet been investigated.

The method proposed here is based on the idea of extending phase demodulation of just a single harmonic to multiple harmonics. The name of the technique is therefore coined as the **M**ulti-**H**armonic **D**emodulation (**MHD**) method. The technique does require an a-priori rough estimate of the speed and is partly inspired by Ref.[226].

The main concept of the technique relies on processing the harmonics with a combination of complex demodulation, band-pass filtering, and phase averaging.

2.3.1 Theory & derivation

To find an expression of the instantaneous angular speed (IAS) that takes into account multiple harmonics, we initially start by looking at an analytic signal consisting of one harmonic k and no noise:

$$x_k(n) = A_k(n)e^{j(\alpha_k\theta(n)+\phi_k)} \tag{2.1}$$

with n being the sample number, k the harmonic number, K the total number of harmonics, A the amplitude, α the harmonic order, ϕ the constant phase, θ the instantaneous angle of rotation. The derivative $\dot{x}_k(n)$ (the notation means a sample of the time derivative) is related to the narrow-band signal $x_k(n)$ as follows:

$$\dot{x}_k(n) = \dot{A}_k(n)e^{j(\alpha_k\theta(n)+\phi_k)} + A_k(n)(e^{j(\alpha_k\theta(n)+\phi_k)}j\alpha_k\dot{\theta}(n)) \tag{2.2}$$

We can rewrite the relationship between $x_k(n)$ and $\dot{x}_k(n)$ as follows, starting from Eq. 2.2:

$$\dot{x}_k(n) = \dot{A}_k(n)e^{j(\alpha_k\theta(n)+\phi_k)} + A_k(n)(e^{j(\alpha_k\theta(n)+\phi_k)}j\alpha_k\dot{\theta}(n)) \tag{2.3}$$

$$= (\frac{\dot{A}_k(n)}{A_k(n)} + j\alpha_k\dot{\theta}(n))A_k(n)e^{j(\alpha_k\theta(n)+\phi_k)} \tag{2.4}$$

$$= (\frac{\dot{A}_k(n)}{A_k(n)} + j\alpha_k\dot{\theta}(n))x_k(n) \tag{2.5}$$

This linear relationship between $x_k(n)$ and $\dot{x}_k(n)$ can be written as follows:

$$\dot{x}_k(n) = \beta_k(n)x_k(n) \tag{2.6}$$

where $\beta_k(n)$ is equal to:

$$\beta_k(n) = \frac{\dot{A}_k(n)}{A_k(n)} + j\alpha_k\dot{\theta}(n) \tag{2.7}$$

The following is then true:

$$\Im\{\beta_k(n)\} = \alpha_k\dot{\theta}(n) \tag{2.8}$$

and also:

$$\dot{\theta}(n) = \Im\{\frac{\dot{x}_k(n)}{\alpha_k(n)x_k(n)}\} \tag{2.9}$$

This is a well-known identity and used in the past to circumvent phase unwrapping [233]. Typically, the measured signal is not a mono-component signal, so the harmonic $x_k(n)$ is obtained approximately through band-pass filtering.

With noise If we introduce process noise (e.g. due to imperfect sampling process and numerical differentiation process) $\nu_k(n)$ to Eq. 2.6, the formulation resembles a state space model where the state and

measurement equation are respectively given by:

$$y_k(n) = x_k(n) + \epsilon_k(n) \tag{2.10}$$
$$\dot{x}_k(n) = \beta_k(n)x_k(n) + \nu_k(n) \tag{2.11}$$

where $x_k(n)$ is the source signal and $\epsilon_k(n)$ is the measurement noise. In reality, $x_k(n)$ and $\dot{x}_k(n)$ are both "experimental" variables, like the input and output of a system. Thus, $\dot{x}_k(n)$ is no longer given by the mathematical derivative of Eq. 2.1 here. To avoid confusion, we change the notation for $\dot{x}_k(n)$ to $z_k(n)$:

$$z_k(n) = \beta_k(n)x_k(n) + \nu_k(n) \tag{2.12}$$

While process noise is an idealization of the problem; "measurement" noise (e.g. due to background noise from other vibration sources) will also be present in practice, even if narrow-band filtering around the harmonics will reduce it. The effect is to bias downward the estimators. However, as is stated later on more clearly, this bias is assumed to more desirable as compared to having an unbounded variance.

Considering only process noise, the least-square error estimator of $\dot{\theta}(n)$ using all the filtered narrow-band harmonic signals to minimize the square of $\nu_k(n)$, is given by:

$$\hat{\dot{\theta}}(n) = \underset{\dot{\theta}(n)}{\operatorname{argmin}} \sum_{k}^{K} ||z_k(n) - j\alpha_k\dot{\theta}(n)x_k(n)||^2 \tag{2.13}$$

Calculating the derivative and setting it to zero, gives the following:

$$\frac{\delta\hat{\dot{\theta}}(n)}{\delta\dot{\theta}(n)} = \frac{\delta}{\delta\dot{\theta}(n)}\Big(\sum_{k}^{K}||z_k(n) - j\alpha_k\dot{\theta}(n)x_k(n)||^2\Big) = 0 \quad (2.14)$$

$$\sum_{k}^{K}\frac{\delta}{\delta\dot{\theta}(n)}\Big(\Big[z_k(n) - j\alpha_k\dot{\theta}(n)x_k(n)\Big]\Big[z_k(n) - j\alpha_k\dot{\theta}(n)x_k(n)\Big]^*\Big) = 0 \quad (2.15)$$

$$\begin{aligned}\sum_{k}^{K}\frac{\delta}{\delta\dot{\theta}(n)}\big[&z_k(n)\dot{x}_k^*(n) + j\alpha_k\dot{\theta}(n)x_k^*(n)z_k(n)\\ &-j\alpha_k\dot{\theta}(n)x_k(n)\dot{x}_k^*(n) + \alpha_k^2\dot{\theta}^2(n)|x_k(n)|^2\big] = 0\end{aligned} \quad (2.16)$$

$$\sum_{k}^{K} j\alpha_k\Big[x_k^*(n)z_k(n) - x_k(n)\dot{x}_k^*(n)\Big] + \sum_{k}^{K}2\alpha_k^2|x_k(n)|^2\dot{\theta}(n) = 0 \quad (2.17)$$

$$\hat{\dot{\theta}}(n) = \frac{\sum_k^K j\alpha_k\Big[x_k(n)\dot{x}_k^*(n) - x_k^*(n)z_k(n)\Big]}{\sum_k^K 2\alpha_k^2|x_k(n)|^2} \quad (2.18)$$

$$\hat{\dot{\theta}}(n) = \frac{\sum_k^K j\Big[x_k(n)\dot{x}_k^*(n) - x_k^*(n)z_k(n)\Big]}{\sum_k^K 2\alpha_k|x_k(n)|^2} \quad (2.19)$$

and rewriting the term in the nominator as follows:

$$\begin{aligned}x_k(n)\dot{x}_k^*(n) - x_k^*(n)z_k(n) = &[\frac{\dot{A}_k(n)}{A_k(n)} - j\alpha_k\dot{\theta}(n)]x_k^*(n)x_k(n) -\\ &[\frac{\dot{A}_k(n)}{A_k(n)} + j\alpha_k\dot{\theta}(n)]x_k(n)x_k^*(n)\end{aligned} \quad (2.20)$$

$$= -2j\alpha_k\dot{\theta}(n)x_k(n)x_k^*(n) \quad (2.21)$$

$$= -2j\Im\{z_k(n)x_k^*(n)\} \quad (2.22)$$

gives after substitution into Eq. 2.19 :

$$\hat{\dot{\theta}}(n) = \frac{\sum_k^K j(-2j)\Im\{z_k(n)x_k^*(n)\}}{\sum_k^K 2\alpha_k|x_k(n)|^2} \quad (2.23)$$

$$\rightarrow \hat{\dot{\theta}}(n) = \frac{\Im\{\sum_k^K z_k(n)x_k^*(n)\}}{\sum_k^K \alpha_k|x_k(n)|^2} \quad (2.24)$$

The advantage of this formulation is to embody several harmonics in the demodulation and therefore the denominator of Eq. 2.24 has a low probability of approaching zero, thus practically guaranteeing a finite variance. This reduces the need to do an a-posteriori smoothing of the resulting speed profile (which is often necessary for standard single harmonic demodulation). Eq. 2.24 has an unbounded variance for the case where only a single harmonic is demodulated (it then boils down to Eq. 2.9), since zero values (or close to zero values- are allowed for $|x_k(n)|$. This means that the ratio can get arbitrarily large, making its mean value unbounded. This unbounded variance makes it a bad estimator with unbounded mean square error. Equation 2.24 has a bounded variance for multiple harmonics due to the positive denominator, but at the cost of potential bias. In general, this is probably more desirable since the mean square error is bounded in this case.

For a single harmonic, a band-pass filter is typically defined based on the minimum and maximum expected frequency of that harmonic. Due to overlapping harmonic content, this approach limits the maximum bandwidth of such a band-pass filter and thus the maximum amount of speed fluctuation. This can be mitigated by using a windowed band-pass filtering approach. For the multi-harmonic demodulation method as proposed above, this overlap issue is avoided in another way. In practice, the complex harmonic signals $x_k(n)$ can be obtained by complex demodulation using a rough speed estimation. This way the harmonics get shifted towards zero frequency (DC) and one can employ a simple and more narrow low-pass filter for every harmonic. The required a-priori rough speed can be obtained in multiple ways that require little input, e.g. using maximum tracking in the spectrogram. Using the rough speed for the complex demodulation does mean that when calculating $\dot{\theta}$ as specified in Eq.2.24, the result is the deviation of the instantaneous frequency of the estimated speed compared to the rough speed.

Another reason to do complex demodulation plus low-pass filtering prior to using the formula in Eq.2.24 is because of the differentiation involved. Differentiating filters usually increase high-frequency noise as can be seen from their transfer function. Close to zero frequency they have a linear response and the influence of the multiplication with the frequency in the spectral domain is reduced when

combined with low-pass filtering. An example of the transfer function magnitude responses for four differentiation schemes are shown in Fig.2.2. In the shown applications a simple first order difference is employed since it has the lowest computation time and the produced results ended up being not significantly influenced by the type of differentiation.

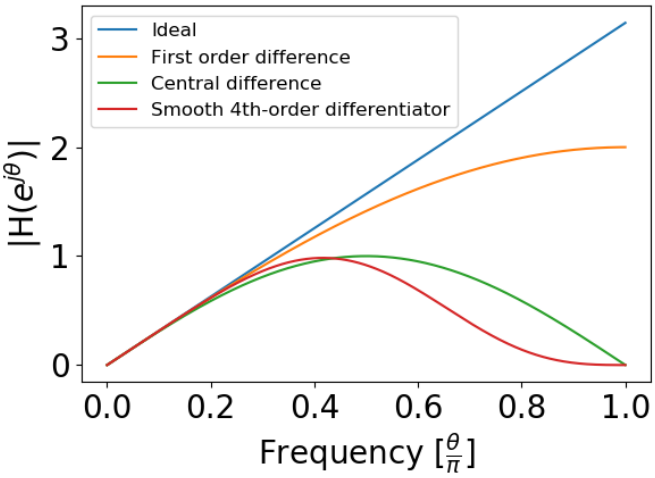

FIGURE 2.2: Magnitude responses of four differentiators.

2.3.1.1 Maximum likelihood estimation of harmonic weights

Looking at Eq. 2.24, we can see that every narrow-band harmonic included in the estimation will be uniformly weighted. However, in real signals it is rarely the case that every harmonic is equally well-excited nor do we expect that every harmonic is well-excited all the time. It is entirely possible that the filtered signal $x_k(n)$ consists entirely of noise and has no meaningful phase information. Also some harmonics can vary in signal-to-noise ratio over time. Therefore it makes sense to try to find a time-dependent weight for each harmonic to take into account when calculating the IAS such that low SNR harmonics are weighted less than high SNR harmonics. The employed approach derives the maximum likelihood estimates of these weights.

After complex demodulation and filtering, we have a set of signals and their derivatives from the measurements that can be formulated as the following set:

- $\{[x_1(n), z_1(n)], [x_2(n), z_2(n)], ..., [x_K(n), z_K(n)]\}$

We again start from the linear relationship between $x_k(n)$ and $z_k(n)$ (as already defined in Eq. 2.12:

$$z_k(n) = \beta_k(n)x_k(n) + \nu_k(n) \tag{2.25}$$

where $\beta_k(n)$ is the regression coefficient and $\nu_k(n)$ the process noise.

We then need to make three assumptions:

- Assumption 1: the process noise is a univariate complex Gaussian distribution with zero bias and unknown variance $\sim N(0, \sigma_k^2(n))$
- Assumption 2: the statistics of the process noise are independent of time and the number of the harmonic k
- Assumption 3: the process noises on the different harmonics are assumed to be uncorrelated

The second assumption is wrong since $\nu_k(n)$ does depend on time and the harmonic number k, but it allows us to use a simple Gaussian distribution.

We now write the probability density function of the process noise (a univariate complex Gaussian distribution):

$$P(x_1, x_2, ..., x_k|\dot{\theta}(n)) = \prod_k^K \frac{1}{\pi\sigma_k^2(n)} e^{\frac{-|z_k(n)-\beta_k(n)x_k(n)|^2}{\sigma_k^2(n)}} \tag{2.26}$$

This equation can be used as the likelihood function, i.e. the density function of $z_k, x)$ conditioned to β_k. The idea behind Eq. 2.26 is to use the asymptotic minimum variance property of the likelihood estimator to find weights that reduce the estimation variance of $\dot{\theta}$ as much as possible.

For a univariate complex Gaussian distribution there is no square root in the PDF nor a factor 2. Thus rewriting the PDF gives us the

likelihood function:

$$P(x_1, x_2, ..., x_k|\dot{\theta}(n)) = \frac{e^{\sum_{k=1}^{K} \frac{-|z_k(n)-\beta_k(n)x_k(n)|^2}{\sigma_k^2(n)}}}{\pi^{KN} \prod_k^K \sigma_k^2(n)} = L(\dot{\theta}(n)) \tag{2.27}$$

Now we need to find $\dot{\theta}(n)$ that maximizes $L(\dot{\theta}(n))$, which translates into minimizing (due to the minus) the log-likelihood.

$$\begin{aligned} \hat{\dot{\theta}}(n)_{ML} &= \text{argmax}(L(\dot{\theta}(n))) \\ &= \text{argmin}(-ln(L(\dot{\theta}(n)))) \\ &= \text{argmin}(\sum_{k=1}^{K} \frac{|z_k(n) - \beta_k(n)x_k(n)|^2}{\sigma_k^2(n)} + constant) \end{aligned} \tag{2.28}$$

The constant does not depend on $\dot{\theta}(n)$ so this can be ignored for the minimization.

To find a local minimum of the log-likelihood function, we take the derivative and set it equal to zero:

$$\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)} = 0 \tag{2.29}$$

We need to take into account that the likelihood function is actually a function of $x_k(\dot{\theta}(n))$, this means that for $L(x_k(\dot{\theta}(n)))$:

- $L \in \mathbb{R}$
- $x_k \in \mathbb{C}$
- $\dot{\theta}(n) \in \mathbb{R}$

We need to derive the real and imaginary part separately, which gives for the real part:

$$\frac{\delta(\Re\{-ln(L(\dot{\theta}(n)))\})}{\delta\dot{\theta}(n)} = \Re\{\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)}\} \tag{2.30}$$

We can rewrite Eq.2.30 as follows:

$$\Re\{\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)}\} = \frac{1}{2}\left(\left(\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)}\right) + \left(\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)}\right)^*\right) \tag{2.31}$$

using these identities:

$$\begin{aligned} z &= a + jb \\ \Re\{z\} &= \frac{z + z^*}{2} \\ \Im\{z\} &= \frac{z - z^*}{2} \\ |z| &= zz^* \end{aligned} \tag{2.32}$$

where z is just a complex number here.

If we write out the first term within the brackets of Eq. 2.31, this gives:

$$\begin{aligned} \frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)} = \sum_{k=1}^{K} &\frac{[z_k(n) - \beta_k(n)x_k(n)]^*(-j\alpha_k x_k(n))}{\sigma_k^2(n)} + \\ &\frac{[z_k(n) - \beta_k(n)x_k(n)](j\alpha_k x_k^*(n))}{\sigma_k^2(n)} \end{aligned} \tag{2.33}$$

The second term is simply the conjugate of Eq. 2.33:

$$\begin{aligned} \left(\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)}\right)^* = \sum_{k=1}^{K} &\frac{[z_k(n) - \beta_k(n)x_k(n)](j\alpha_k x_k^*(n))}{\sigma_k^2(n)} + \\ &\frac{[z_k(n) - \beta_k(n)x_k(n)]^*(-j\alpha_k x_k(n))}{\sigma_k^2(n)} \end{aligned} \tag{2.34}$$

The real part now becomes:

$$\begin{aligned} \Re\{\frac{\delta(-ln(L(\dot{\theta}(n))))}{\delta\dot{\theta}(n)}\} = \sum_{k=1}^{K} &\frac{[z_k(n) - \beta_k(n)x_k(n)]^*(-j\alpha_k x_k(n))}{\sigma_k^2(n)} + \\ &\frac{[z_k(n) - \beta_k(n)x_k(n)](j\alpha_k x_k^*(n))}{\sigma_k^2(n)} \end{aligned} \tag{2.35}$$

Doing the same for the imaginary part using $\Im\{z\} = \frac{z-z^*}{2}$, we can see that the first (Eq. 2.33) and second term (Eq. 2.34) cancel each other out. Thus, Eq.2.29 becomes:

$$\sum_{k=1}^{K} \frac{[z_k(n) - \beta_k(n)x_k(n)]x_k^*(n)j\alpha_k}{\sigma_k^2(n)} + \frac{[z_k(n) - \beta_k(n)x_k(n)]^* x_k(n)(-j)\alpha_k}{\sigma_k^2(n)} = 0 \tag{2.36}$$

In Eq. 2.36 we can recognize again the structure of $\Re\{z\} = \frac{z+z^*}{2}$, so we rewrite it as:

$$\Re\{\sum_{k=1}^{K} \frac{[z_k(n) - \beta_k(n)x_k(n)]x_k^*(n)j\alpha_k}{\sigma_k^2(n)}\} = 0 \tag{2.37}$$

To move closer to the expression in Eq.2.24 for $\hat{\dot{\theta}}(n)$, we rewrite Eq. 2.37 with the imaginary part instead, using following identity:

$$\begin{aligned} z &= a + jb \\ \Re\{jz\} = \Re\{j(a+jb)\} &= -b = -\Im\{z\} \end{aligned} \tag{2.38}$$

This results in:

$$\Im\{\sum_{k=1}^{K} \frac{[z_k(n) - \beta_k(n)x_k(n)]x_k^*(n)\alpha_k}{\sigma_k^2(n)}\} = 0 \tag{2.39}$$

Splitting the equation up, gives:

$$\Im\{\sum_{k=1}^{K} \frac{z_k(n)x_k^*(n)\alpha_k}{\sigma_k^2(n)}\} = \Im\{\sum_{k=1}^{K} \frac{\beta_k(n)|x_k(n)|^2\alpha_k}{\sigma_k^2(n)}\} \tag{2.40}$$

Now recall that the imaginary part of $\beta_k(n)$ is $\alpha_k\dot{\theta}(n)$ and that the other product factors are real, making the second part equal to:

$$\Im\{\sum_{k=1}^{K} \frac{\beta_k(n)|x_k(n)|^2\alpha_k}{\sigma_k^2(n)}\} = \dot{\theta}(n)\sum_{k=1}^{K} \frac{|x_k(n)|^2\alpha_k^2}{\sigma_k^2(n)} \tag{2.41}$$

This then gives us the maximum likelihood estimate for $\dot{\theta}(n)$:

$$\hat{\dot{\theta}}(n)_{ML} = \frac{\Im\{\sum_{k=1}^{K} \frac{z_k(n)x_k^*(n)\alpha_k}{\sigma_k^2(n)}\}}{\sum_{k=1}^{K} \frac{|x_k(n)|^2\alpha_k^2}{\sigma_k^2(n)}} \tag{2.42}$$

The optimal weights for each harmonic are therefore defined as:

$$w_k(n) = \frac{\alpha_k}{\sigma_k^2(n)} \tag{2.43}$$

As can be seen from Eq. 2.43, this does require knowledge of the variance of the noise $\sigma_k^2(n)$.

Noise variance derivation Consequently, we need to derive the maximum likelihood estimate for the noise variance. The noise variance is assumed to be constant over a short time interval $[n-I, n+I]$. The pdf for the noise variance is then given by:

$$P(x_k|\sigma_k^2) = \prod_{i=n-I}^{n+I} \frac{1}{\pi\sigma_k^2} e^{\frac{-|z_k(i)-\beta_k(i)x_k(i)|^2}{\sigma_k^2}} \tag{2.44}$$

$$\hat{\sigma}_{k\,ML}^2 = \text{argmax}(L(\sigma_k^2)) \tag{2.45}$$

with the likelihood in this case being:

$$L(\sigma_k^2) = \frac{1}{\pi^{(2I+1)}(\sigma_k^2)^{(2I+1)}} e^{\sum_{i=n-I}^{n+I} \frac{-|z_k(i)-\beta_k(i)x_k(i)|^2}{\sigma_k^2}} \tag{2.46}$$

and the log-likelihood:

$$\ln(L(\sigma_k^2)) = (2I+1)\ln(\pi) + (2I+1)\ln(\sigma_k^2) + \frac{1}{\sigma_k^2}\sum_{i=n-I}^{n+I} -|z_k(i)-\beta_k(i)x_k(i)|^2 \tag{2.47}$$

We again derive the maximum log-likelihood:

$$\frac{\delta(-\ln(L(\sigma_k^2)))}{\delta\sigma_k^2} = 0 \tag{2.48}$$

$$-(2I+1)\frac{1}{\sigma_k^2} + \frac{1}{\sigma_k^4}\sum_{i=n-I}^{n+I}|z_k(i) - \beta_k(i)x_k(i)|^2 = 0 \tag{2.49}$$

$$\sigma_k^2 = \frac{1}{2I+1}\sum_{i=n-I}^{n+I}|z_k(i) - \beta_k(i)x_k(i)|^2 \tag{2.50}$$

To simplify further we can assume that $\frac{\dot{A}_k(n)}{A_k(n)}$ is slow and small compared to the instantaneous frequency:

$$\hat{\sigma}_{k\,ML}^2 \approx \frac{1}{2I+1}\sum_{i=n-I}^{n+I}|z_k(i) - j\alpha_k\dot{\theta}(i)x_k(i)|^2 \tag{2.51}$$

Since we ignored the $\frac{\dot{A}_k(n)}{A_k(n)}$ term, this estimate for $\hat{\sigma}_k^2$ will be an overestimation. The idea for the weights is that a high variance of the noise will lead to a low weight. The procedure will iteratively switch between estimating $\hat{\hat{\theta}}(n)_{ML}$ and $\hat{\sigma}_{k\,ML}^2$. For a first estimate of $\hat{\theta}(n)_{ML}$ we can assume σ_k^2 to be equal to 1 or k or even $e^{\gamma k}$, depending on the observable evolution of the signal-to-noise ratio for higher harmonics.

2.3.2 Simulation assessment

In order to highlight the influence of the maximum likelihood weighting, a simple example is chosen where a signal with a normalized sample rate of 1 Hz and length of 20000 samples is simulated that consists of the following:

- 5 sinusoidal harmonics with a base rotation speed of 0.03 Hz, but importantly the harmonic numbers are chosen to be *odd*, meaning that there are no 2^{nd}, 4^{th},6^{th}, 8^{th}, nor 10^{th} harmonic present in the signal.
- The 5th harmonic however is chosen to be *asynchronous* with the speed. This means the speed-synchronous signal content only consists of the 1^{st}, 3^{rd}, 7^{th}, and 9^{th} harmonic.
- Additionally, there is also 1 very lightly damped resonance in the middle of the 7th harmonic, which is detrimental for the phase demodulation of that harmonic.
- The 5 harmonics are all amplitude modulated
- Additive white Gaussian noise is added with an SNR of -5 dB

The spectrogram of the resulting simulated signal is shown in Fig. 2.3. It is easily observed that the 7^{th} harmonic around 0.2 Hz is masked by the lightly damped resonance and the 5^{th} harmonic around 0.15 Hz is asynchronous. In total the MHD is run for 20 iterations with and without MLE weighting of the harmonics to see the difference in performance.

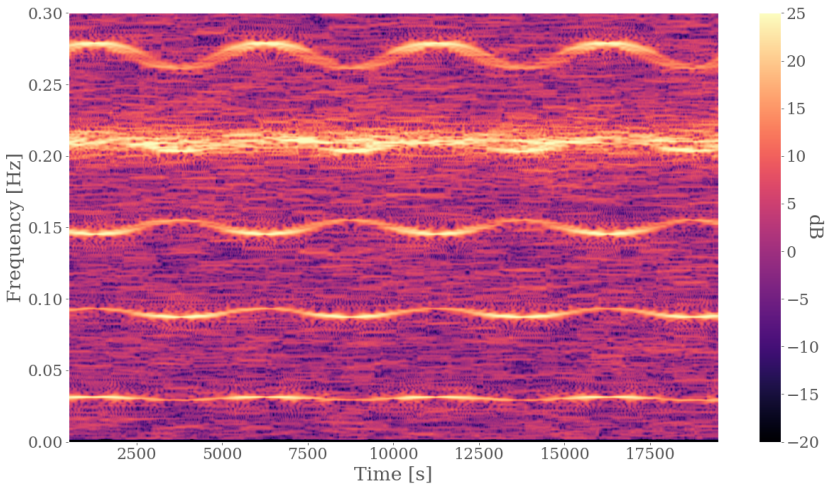

FIGURE 2.3: Spectrogram of the simulated signal.

Figure 2.4 displays the resulting speed estimation. It is clearly observed that in this case it is imperative to use harmonic weighting in order to have a proper selection of the right harmonics for the estimation. Of course, a similar result could be obtained by manually selecting only the 1^{st}, 3^{rd}, and 9^{th} harmonic for the MHD method without MLE. However, this example illustrates the automation potential of the full method.

To further illustrate the functioning of the method with MLE, the time-dependent weighting is shown in Fig. 2.5 for the odd harmonics in red with the actual amplitude modulation of each harmonic in black. The figure indicates that the 1^{st}, 3^{rd}, and 9^{th} harmonic follow rather well the amplitude modulation pattern and have a significantly higher value constantly as compared to the 5^{th} and 7^{rd} harmonic (y-axes are exactly the same scale).

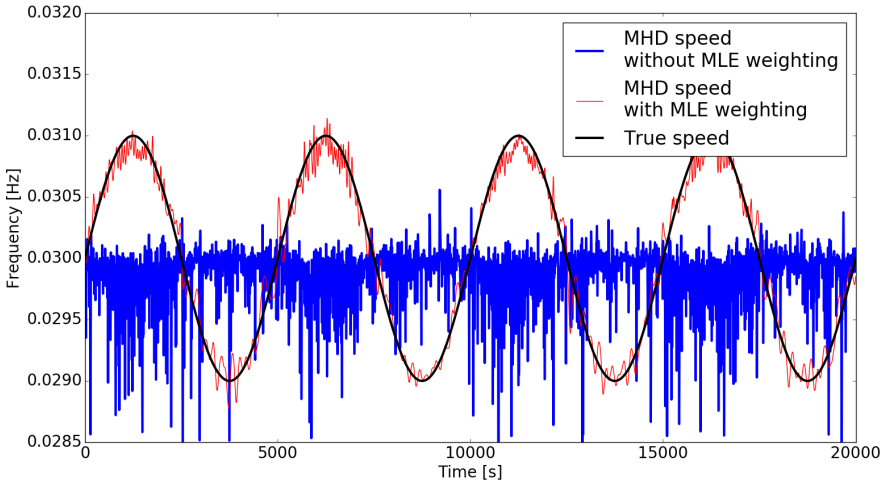

FIGURE 2.4: Resulting speed profiles of the estimation with and without maximum likelihood estimation of the weights as compared to the true speed.

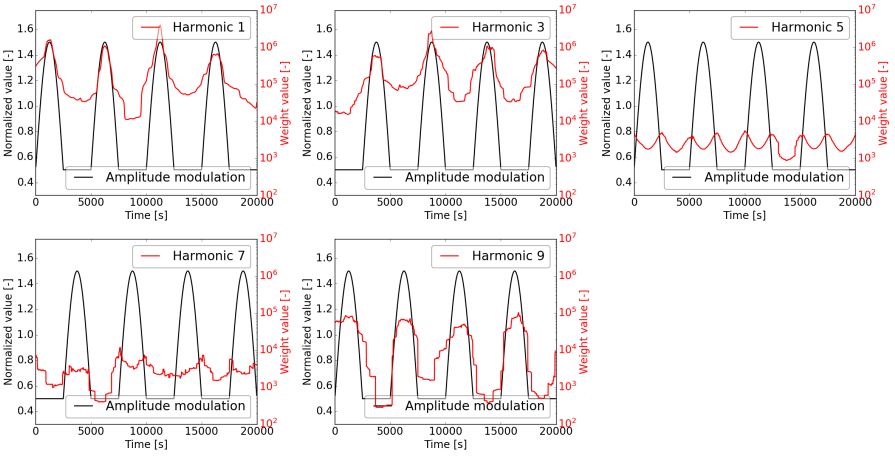

FIGURE 2.5: The ML-estimates in time compared to the actual amplitude variation of the excited harmonics.

Lastly, to get an idea about the convergence of the weights, the time-averaged weights are displayed in Fig. 2.6 for every iteration

and every harmonic. Apart from the dominance of the weights of the 1st, 3rd, and 9th harmonic, it can be noticed that after approximately 5 iterations the weights have practically already converged to their final values.

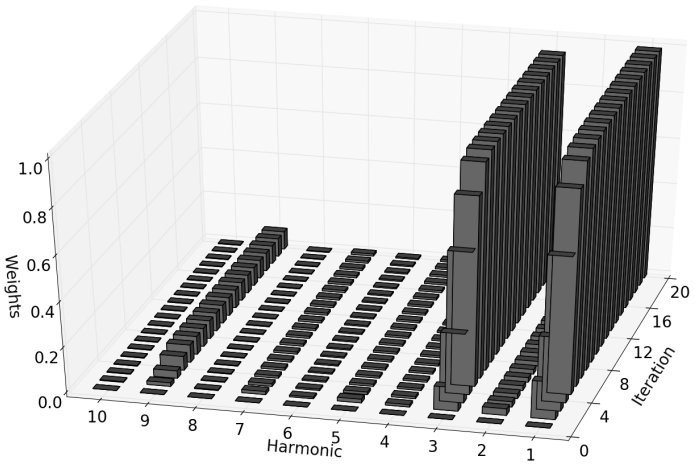

FIGURE 2.6: Evolution of the MLE harmonic weights per iteration and per harmonic.

The next step is now to validate the method on experimental data, which is done in section 2.5. First however, the theory behind some other techniques is detailed such that an in-depth comparison can be made of all techniques together on the same data sets.

2.4 Other IAS estimation methods

The other methods that are compared in this chapter with regards to their speed estimation performance, are in essence either based on demodulation or on a time-frequency representation of the signal, since these represent the majority of the techniques out there. This means that no purely model-based or scale transform approaches are investigated. In total seven other methods besides the MHD method are assessed:

1. Phase demodulation
2. Iterative phase demodulation
3. Demodulation based on the Teager-Kaiser Energy Operator
4. Multi-order probabilistic approach
5. ViBES method
6. Cepstrum-based multi-order approach
7. Maximum tracking in combination with a Vold-Kalman filter

A short background summary of every method is provided to explain some of the key details concerning the methods. For more detailed information, interested readers are referred to the origin articles of each method, with the exception of the ViBES method which has not been published yet. Keep in mind that some of the techniques presented here can also be used in different combinations with other techniques, but investigating all possible combinations would be unfeasible.

2.4.1 Phase demodulation based on the analytic signal

Perhaps the most used approach for vibration-based instantaneous angular speed estimation is based on phase demodulation of a shaft-speed related harmonic that exhibits a high signal-to-noise ratio. This method is fairly straightforward and is mainly based on using an (ideal) band-pass filter around a well-separated, high SNR harmonic of the rotation speed. After defining the optimal lower and upper cutoff frequencies for the band-pass filter, the harmonic is isolated from the complex spectrum. Next, the analytic signal $x_{analytic}(t)$ can be obtained by inverse Fourier transforming the complex band-pass filtered spectrum to the time domain without its negative frequencies. This $x_{analytic}(t)$ is then ideally a mono-component signal with a high SNR and can be written in its exponential form:

$$x_{analytic}(t) = A(t)e^{j\phi(t)} \; with \; A(t) \geq 0 \tag{2.52}$$

Thus, the instantaneous phase is given by the imaginary part of the logarithm of the analytic signal:

$$\phi(t) = \Im(log(x_{analytic}(t))) \tag{2.53}$$

Finally, the instantaneous frequency is estimated based on the variation of the unwrapped phase:

$$f_{inst}(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{2.54}$$

The instantaneous angular speed can then be obtained by multiplying $f_{inst}(t)$ with the correct kinematic ratio. Figure 2.7 shows a flowchart of the full approach that clearly does not require a lot of steps and is easy to implement.

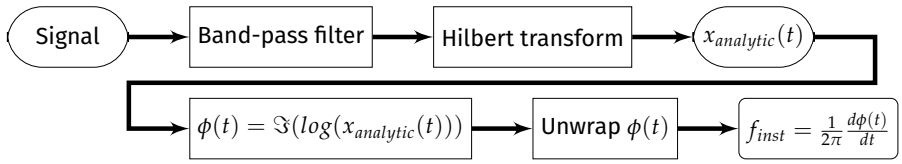

FIGURE 2.7: Overview of a standard phase demodulation approach.

While this approach can deliver very accurate IAS estimations in cases where such a single, constantly present, and dominant harmonic is existing, there are ample cases where this approach has its limitations regarding applicability. In complex rotating systems, the deterministic components are not necessarily all harmonically related, causing crossing orders and skewing the extracted instantaneous phase. Additionally, such systems often operate in strongly varying conditions which can cause harmonics to fade into the noise. Another limitation of this technique is the boundaries for the possible speed fluctuation. When a harmonic at a higher frequency is chosen, this often accompanies a reduced possible relative bandwidth size for the band-pass filter due to overlapping of lower and higher order harmonics. Take for example the straightforward case where there are only harmonics of the shaft speed present in the signal at the average frequency f_{shaft}, so no sidebands related to the meshing of gears.

The maximum speed fluctuation $\Delta\dot{\theta}_{max}$ that can be present in the signal to allow for band-pass filtering the nth harmonic so that it does not overlap with its lower order harmonic at $(n-1)$ and higher order harmonic at $(n+1)$ is:

$$\Delta\dot{\theta}_{max} = \frac{f_{shaft}}{2n+1} \tag{2.55}$$

with f_{shaft} being the average of the minimum and maximum shaft speed. This means that in order to define a band-pass filter on the second harmonic $n = 2$ so that it does not overlap with the third harmonic, the speed fluctuation of the signal cannot be greater than $\frac{f_{shaft}}{5}$.

This simple example shows that the possible speed fluctuation quickly gets limited by the maximum bandwidth that does not cause overlap with other harmonics. A small adaptation of the method is therefore made for long signals with strong speed fluctuations. For such signals, it is impossible to incorporate the full signal in one demodulation step since the required filter bandwidth would be too large and encompass multiple harmonic orders. Thus the demodulation is done consecutively on windowed sections of the signal. In this dissertation a rectangular window with a standard overlap of 50% is chosen for all phase demodulation methods. The obtained speed profiles of overlapping sections are averaged together using a weighted average based on a Hanning window with its center in the middle of the window. This is done to reduce the influence of the end effects associated with using an ideal FFT band-pass filter.

2.4.2 Iterative phase demodulation

The previous paragraph touches upon the fact that it is not always possible to define a band-pass filter on the full signal when the speed fluctuation is too high. Therefore, a possible remedy [234] is to use a lower order harmonic first. This way an initial speed estimate can be obtained which can afterward be used for angular resampling of the signal. After angular resampling, the speed fluctuation still present in the signal is due to the estimation error in the initial speed estimate, but it should have greatly decreased. Thus, another phase demodulation step can be performed using a higher order harmonic now with a

band-pass filter that has a smaller bandwidth thanks to the decreased speed fluctuation. Figure 2.8 shows a simplified diagram of this iterative demodulation approach.

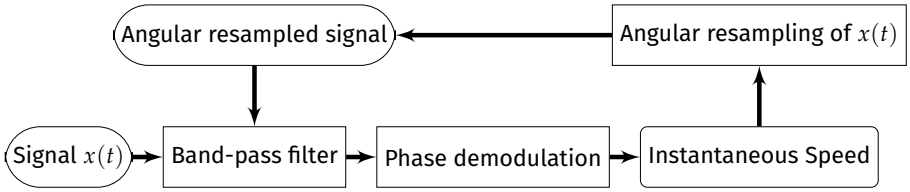

FIGURE 2.8: Overview of the iterative demodulation approach.

2.4.3 Teager-Kaiser Energy Operator

An interesting new way of looking at the Teager Kaiser Energy Operator (TKEO) was found by Randall et al. [229]. Instead of using the TKEO for tracking the energy in speech signals, they show that it is possible to estimate the speed directly from the TKEO formulation.

The TKEO operator Ψ_c is defined in continuous form as follows:

$$\Psi_c(x(t)) = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \tag{2.56}$$

If the TKEO is applied to a monocomponent signal such as an amplitude and frequency modulated sine (which in practice can be achieved through band-pass filtering), then the signal $x(t)$ can be written as:

$$x(t) = A(t)sin(\phi(t)) \; with \; \phi(t) = \int_t \omega(t)dt \tag{2.57}$$

with $A(t)$ and $\omega(t) =$ the time-dependent amplitude and frequency respectively. If the variation of these two quantities is slow, then:

$$\dot{x}(t) \approx \omega(t)A(t)cos(\phi(t)) \tag{2.58}$$

$$\ddot{x}(t) \approx -[\omega(t)]^2 A(t)sin(\phi(t)) \tag{2.59}$$

The TKEO then becomes:

$$\Psi_c(x(t)) = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \tag{2.60}$$
$$\approx [\omega(t)]^2[A(t)]^2(cos^2\phi(t) + sin^2\phi(t)) \tag{2.61}$$
$$\approx [\omega(t)]^2[A(t)]^2 \tag{2.62}$$

The squared envelope $se_x(t)$ of $x(t)$ is defined as:

$$se_x(t) = x^2(t) + \hat{x}^2(t) \tag{2.63}$$

with $\hat{x}(t)$ the hilbert transform of $x(t)$. It then follows that:

$$\Psi_c(x(t)) = [\omega(t)]^2[A(t)]^2 = [\omega(t)]^2 se_x(t) = se_{\dot{x}}(t) \tag{2.64}$$

$$\omega(t) = \sqrt{\frac{se_{\dot{x}}(t)}{se_x(t)}} \tag{2.65}$$

As Randall & Smith point out in their paper [229], it is quite straightforward to implement this technique. A zero phase-shift ideal FFT filter can be used to band-pass filter the signal. The squared envelope $se_x(t)$ is obtained by simply inverse Fourier transforming the filtered band back to the time domain without the negative frequencies and squaring the amplitude. The squared envelope of the derivative $se_{\dot{x}}(t)$ is obtained in exactly the same way but an additional multiplication with $j\omega$ is done in the spectral domain concurrently with the band-pass filtering. Multiplying with $j\omega$ is equivalent to applying an ideal differentiation filter on the signal. It avoids partially the typical differentiation issue related to noise amplification of higher frequencies (due to the differentiation filter gain that increases linearly with frequency) thanks to the limited bandwidth of the band-pass filter. The use of ideal Fourier filtering does create some end effects related to the Gibbs phenomenon, but these can be alleviated if one can transform a slightly larger signal part. The start and the end of the signal can then simply be discarded in further analysis. Since the entire procedure is mostly carried out in the frequency domain, Randall & Smith propose to rename the technique to "Frequency Domain Energy Operator" or **FDEO** in a more recent paper [235]. In the same paper, Randall & Smith also point out that a closely related and potential alternative

for the FDEO can be found using the derivative of the instantaneous phase of the analytic signal [233]:

$$\omega(t) = \dot{\phi}(t) = \frac{x(t)\dot{\hat{x}}(t)}{A_x^2(t)} = \Im\Big[\frac{\dot{x}_a(t)}{x_a(t)}\Big] \tag{2.66}$$

where $\dot{x}_a(t)$ is the time derivative of $x_a(t)$, which again can be obtained by multiplying the spectrum by $j\omega$ over a select frequency band. This equation is not further investigated in this chapter.

2.4.4 Multi-order Probabilistic approach

The multi-order probabilistic approach (MOPA) belongs to the group of techniques that try to utilize more than just one harmonic order in the signal for the speed estimation. The general idea behind MOPA as proposed by Leclère et al. [187] is based on regarding the instantaneous spectrum, which can be obtained through a short time Fourier transform (STFT), of the vibration signal as a probability density function (pdf) of the IAS Ω. Consequently, if the spectrum has a high amplitude at frequency f, there is a high probability that the shaft frequency is equal to f/H_i with H_i being the excitation order or, for the cases described below, the gear ratios. The STFT can be calculated for a signal $x(n)$ as follows:

$$STFT_x(m,k) = \sum_{n=0}^{N-1} x(n)w(n-m)e^{-2\pi jk\frac{n}{N}} \tag{2.67}$$

with N the window length, n the sample index, m the window index, and k the frequency index. It is important to define a range for the IAS in which the user expects the IAS to reside. This range has a lower bound Ω_{min} and an upper bound Ω_{max}. The following pdf can then be constructed:

$$\begin{cases} p(\Omega|H_i) = \frac{1}{\xi_i}A(H_i\omega) & \text{for } \Omega_{min} < \omega < \Omega_{max} \\ p(\Omega|H_i) = 0 & \text{for } \omega < \Omega_{min} \mid \omega > \Omega_{max} \end{cases} \tag{2.68}$$

with $A(f)$ a whitened version of the vibration signal's spectrum and ξ_i a normalization factor to make sure the pdf has unit area:

$$\xi_i = \int_{\Omega_{min}}^{\Omega_{max}} A(H_i\omega)d\omega. \tag{2.69}$$

The purpose of the whitening is essentially to reduce the influence of resonances on the generated pdf, since it is undesirable to give a too high probability to a certain part of the spectrum only due to the increased amplitudes because of a resonance. The used whitening technique should be chosen based on the application.

To improve the IAS estimation and utilize more of the information potential of the spectrum, one has to include more than just one pdf based on one gear ratio or meshing order. Afterwards these different pdfs can be combined together in one pdf by multiplication. Equation 2.68 does not take into account the possibility that a part of the spectrum for a certain harmonic H_i can exceed the Nyquist frequency. In this case the pdf is made uniform above f_{max}/H_i:

$$\begin{cases} p(\Omega|H_i) = \frac{1}{\xi_i} A(H_i\omega) & \text{for } \Omega_{min} < \omega < f_{max}/H_i \\ p(\Omega|H_i) = \frac{1}{\Omega_{max}-\Omega_{min}} p(\Omega|H_i) = 0 & \text{for } \omega < \Omega_{min} \text{ or } \omega > \Omega_{max} \end{cases} \tag{2.70}$$

with ξ_i now:

$$\xi_i = \frac{\Omega_{max} - \Omega_{min}}{f_{max}/H_i - \Omega_{min}} \int_{\Omega_{min}}^{f_{max}/H_i} A(H_i\omega)d\omega. \tag{2.71}$$

The inputs of the method are thus an approximate range for the IAS, the meshing orders and the vibration signal. For every order a pdf is then constructed based on the signal's instantaneous spectrum and rescaled to the given range for the IAS. Next, the pdfs are multiplied to combine the information of all the orders so that the main corresponding estimate for the IAS becomes the most dominant peak in the pdf.

Currently, the pdfs are still independently generated for each time step and thus do not guarantee any continuity of the IAS, which is a logical assumption for any mechanical system. Due to the inertia of the rotating shafts, strong acceleration or deceleration is improbable.

As such, to improve the results further, an a priori of continuity is introduced for the IAS. The concept relies on generating for each time step a pdf that is based on the pdfs of several time steps before and after the central pdf. Appropriate weighting of these pdfs is done by convolving the pdf with a centered Gaussian and the time relationship is introduced by letting the variance depend on the time between the considered pdf and the central pdf. The pdf at time step j generated by the pdf at time step $j+k$ is defined as:

$$p(\Omega_j)_{j+k} = \int_{\Omega_{min}}^{\Omega_{max}} p(\Omega_j|\Omega_{j+k})d\omega \propto \exp(\frac{\omega^2}{2\sigma_k^2}) * p(\Omega_{j+k}) \tag{2.72}$$

with $p(\Omega_j)_{j+k}$ the pdf at time j that can be obtained by convolution of the pdf at time $j+k$, $p(\Omega_{j+k})$ with a centered Gaussian, and $\sigma_k = |\gamma k \Delta_t|$ with Δ_t being the time step, γ the standard acceleration of the IAS. Similar to the previous step in which the pdfs corresponding to the different orders have to be multiplied for each time step to obtain a single combined pdf, there are now again multiple pdfs for every time step $j+k$ belonging to time steps before and after time step j. Thus the final step is to multiply again all the pdfs for every time step:

$$p(\Omega_j)_s \propto \prod_{k=-K}^{K} [\Omega_j]_{j+k}. \tag{2.73}$$

The instantaneous angular speed can then simply be obtained by calculating the expected value of every pdf.

2.4.5 ViBES method

The ViBES method is another STFT-based algorithm recently developed at MIT's Research Laboratory of Electronics and similar in approach to the MOPA method. It also views the STFT as a combination of probability density functions but has the user define ranges of the STFT that represent these pdfs. These ranges also track the instantaneous frequency estimates. For this method, the user defines an initial estimate of the vibration frequency, $\hat{f}_0$, and defines a set of vibration

mesh ranges of interest,

$$\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M\}, \tag{2.74}$$

where $\mathbf{r}_i = [r_{i,\min}, r_{i,\max}]$, i.e., all real values in the range $r_{i,\min}$ to $r_{i,\max}$.

The set of mean values of these ranges,

$$\bar{\mathbf{R}} = \{\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \ldots, \bar{\mathbf{r}}_M\} \tag{2.75}$$

should correspond to strong (in the signal-to-localized-noise ratio sense) individual vibration spectral components correlated to the frequency profile of interest (e.g. the frequency of a particular shaft's rotation). For convenience, these mesh ranges should be ordered from low to high and related such that,

$$\frac{\mathbf{r}_1}{\bar{\mathbf{r}}_1} = \frac{\mathbf{r}_i}{\bar{\mathbf{r}}_i}. \tag{2.76}$$

In this way, the user can define $\bar{\mathbf{R}}$ based on the expected locations of strong vibrational mesh content, and set the width of all the corresponding ranges based on the width of a single range. This defining range should be set wide enough so that all ranges sufficiently encompass their corresponding vibrational spectrum component but not too wide that any range also encompasses adjacent correlated vibrational spectrum content. For example, for harmonically related vibration meshes, each range should only encompass one harmonic.

The method operates by calculating the STFT as defined in Eq. 2.67 and obtaining the short-time amplitude spectra $A_t[k]$ of the vibration measurement, with t being the time at the center of the window.

At each time instance, t, the algorithm converts the M vibration mesh ranges of interest, $\mathbf{R}$, into frequency ranges as,

$$\mathbf{F} = \hat{f}_{t-\Delta t}\mathbf{R}, \tag{2.77}$$

where $\hat{f}_{t-\Delta t}$ is the previous time instance frequency estimate. The corresponding M sections of $A_t[k]$ are $A_t[\mathsf{K}]$ where $\mathsf{K} \ni k_i$, and,

$$k_i = N\frac{\hat{f}_{t-\Delta t}}{F_s}\mathbf{r}_i. \tag{2.78}$$

with F_s being the sample rate of the measurement.

The algorithm treats the magnitude profiles of these M ranges as scaled and sampled probability density function estimates of the location of the vibration frequency profile of interest at time t, i.e.,

$$\hat{p}_{t,i}\left(f_i\right) = \alpha_i |A_t[k_i]|, \tag{2.79}$$

where,

$$f_i = \frac{f[k_i]}{\bar{\mathbf{r}}_i}. \tag{2.80}$$

Here, α_i is the scaling factor to ensure $\sum \hat{p}_{t,i}(f_i) = 1$.

If the mesh ranges, $\mathbf{R}$, are ordered from low to high and related according to (2.76), then each f_i will cover a similar range, though f_M will have the highest frequency resolution. Each probability density function, $\hat{p}_{t,i}(f_i)$ can then be interpolated to match the sample points, f_M, so that the M resulting probability density function estimates $\hat{p}_{t,i}(f_M)$ have the same length and thus can be arithmetically combined to form a single composite probability density function,

$$\hat{p}_t(f_M) = g\left(\hat{p}_{t,i}, f_M\right). \tag{2.81}$$

For the results presented here, the function $g()$ calculates the joint probability density function assuming each $p_{t,i}$ as independent, i.e.,

$$g(\hat{p}_{t,i}, f_M) = \prod_{i=1}^{M} \hat{p}_{t,i}(f_M). \tag{2.82}$$

Finally, the frequency estimate for time, t, is calculated as the expected value of the composite probability density function,

$$\hat{f}_t = \sum_{f_i \in f_M} f_i \hat{p}_t(f_i). \tag{2.83}$$

This process is repeated for the length of the vibration signal $x(n)$ to form the full estimated frequency profile, $\hat{f}$.

2.4.5.1 Adaptation 1: Illegal Regions

To improve robustness, the ViBES method allows defining "illegal" frequency regions unavailable for the analysis. This helps protect against the amplifying influences of resonances as well as vibration disturbances from extraneous sources, e.g., 50 Hz or 60 Hz "hum" corresponding to the electrical line frequency. These regions can be defined as,

$$\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_L\}. \tag{2.84}$$

At each time instance, t, if any of the frequency regions of $\mathbf{F}$ (calculated in (2.77)) overlap with any regions of $\mathbf{Z}$, the algorithm removes the overlapping regions from $\mathbf{F}$ and removes the corresponding mesh range from $\mathbf{R}$. That is, (2.77) becomes,

$$\mathbf{F} = \left\{ \hat{f}_{t-\Delta t}\mathbf{r}_i \mid \hat{f}_{t-\Delta t}\mathbf{r}_i \cap \mathbf{z}_j = \varnothing \quad \forall \quad \mathbf{z_j} \in \mathbf{Z} \right\}, \tag{2.85}$$

and $\mathbf{R}$ is updated as,

$$\mathbf{R} = \frac{\mathbf{F}}{\hat{f}_{t-\Delta t}}. \tag{2.86}$$

The analysis then continues through (2.78) - (2.83).

2.4.5.2 Adaptation 2: Variance-Based Lock-in Tracking

The ViBES method also allows the automated toggling between two modes, *wait* and *track*, in response to a criteria metric. For example, the variance in the composite probability density function,

$$\sigma^2 = \sum_{f_i \in f_M} \left(f_i - \hat{f}_t \right)^2 \hat{p}_t(f_i), \tag{2.87}$$

can be thought of as a confidence metric of the estimate, $\hat{f}_t$. A low value corresponds to when the individual probability density functions, $\hat{p}_{t,i}(f_M)$, have high signal-local-noise ratios with content co-aligned in f_M. A high value occurs when the signal-local-noise ratio is low and/or when their vibration content is not well co-aligned. If the previous time-instance estimate of frequency, $\hat{f}_{t-\Delta t}$, is good at time

instance t, then σ^2 will be small and the analysis should save the updated estimate. However, if the estimate is poor, then the content will not align well and σ^2 will be high. Thus, maintaining a binary state variable S_t and setting it conditional to σ^2 incorporates an algorithmic decision to start, keep, or stop tracking frequency. That is,

$$S_t = \begin{cases} 0, & \text{if } S_{t-\Delta t} = 0 \text{ and } \sigma^2 > \sigma_l^2 \\ 0, & \text{if } S_{t-\Delta t} = 1 \text{ and } \sigma^2 > \sigma_h^2 \\ 1, & \text{if } S_{t-\Delta t} = 0 \text{ and } \sigma^2 \leq \sigma_l^2 \\ 1, & \text{if } S_{t-\Delta t} = 1 \text{ and } \sigma^2 \leq \sigma_h^2 \end{cases}, \tag{2.88}$$

where σ_l^2 and σ_h^2 are empirical threshold values with $\sigma_h^2 > \sigma_l^2$. In (2.88), $S_t = 0$ indicates *wait*-mode and $S_t = 1$ indicates *track*-mode. (2.83) can then be altered to depend on S_t,

$$\hat{f}_t = \begin{cases} \sum_{f_i \in f_M} f_i \hat{p}_t(f_i), & \text{if } S_t = 1 \\ \varnothing & \text{if } S_t = 0 \end{cases}, \tag{2.89}$$

so that the profile $\hat{f}_t$ only contains estimates when in *track* mode.

Under this operation, (2.77) and (2.78) need to be adapted by replacing $\hat{f}_{t-\Delta t}$ with $f_{t,0}$ as $\hat{f}_{t-\Delta t}$ can take on the null set. By setting $f_{t,0}$ as,

$$f_{t,0} = \begin{cases} \hat{f}_{t-\Delta t}, & \text{if } S_t = 1 \\ f_w, & \text{if } S_t = 0 \end{cases}. \tag{2.90}$$

where f_w is the initial frequency guess when in *wait* mode and defined by the user.

2.4.6 Cepstrum-based multi-order approach

This chapter also investigates the performance of combining the cepstrum transformation with the multi-order probabilistic approach. This technique is inspired by the approach proposed by F. Bonnardot at the data contest [236] of the Surveillance 8 conference which was held at

the Roanne Institute of Technology in France. In this contest Bonnardot uses the cepstrum to find an initial estimate for the speed. Instead of a Time-Frequency Representation (TFR), a Time-Quefrency Representation (TQR) is generated based on a short-time cepstrum transform. The technique then makes use of a tracking algorithm based on the maxima of the first five rhamonics in the TQR and based on using linear prediction to find the expected quefrencies.

Background about cepstrum

The complex cepstrum is defined as the inverse Fourier transform of the log spectrum. It can be expressed in terms of the amplitude and the phase of the spectrum:

$$C_c(\tau) = \mathscr{F}^{-1}\{\ln(X(f))\} = \mathscr{F}^{-1}\{\ln(A(f)) + j\phi(f)\} \tag{2.91}$$

where $X(f)$ is the frequency spectrum of the signal $x(t)$:

$$X(f) = \mathscr{F}\{x(t)\} = A(f)e^{j\phi(f)} \tag{2.92}$$

By setting the phase to zero in Eq.(3.1), the real cepstrum can be obtained:

$$C_r(\tau) = \mathscr{F}^{-1}\{\ln(A(f))\} \tag{2.93}$$

Here, τ is a measure of time, referred to as "quefrency", however it is not defined in the same sense as a signal in the time domain. A peak at a certain quefrency corresponds to the inverse period of a series of periodic harmonics in the spectrum. For example, if the sampling rate of a signal is 20 kHz and the cepstrum displays a quefrency peak at 1000 samples, the peak indicates that there is a family of harmonics present in the spectrum with a spacing of 20 Hz (20 kHz/1000 samples).

An important property of the cepstral domain is that the convolution of two time domain signals can be expressed as an addition of their cepstra. Consider the output signal $y(t)$ of a physical system represented by the convolution of an input signal $x(t)$ and an impulse

response h(t) of the system:

$$y(t) = x(t) * h(t) \tag{2.94}$$

Because of the convolution theorem, this time domain expression transforms into a multiplication in the frequency domain:

$$Y(f) = X(f)H(f) \tag{2.95}$$

In turn, taking the logarithm of Eq.(3.5) transforms the multiplication into a sum:

$$\ln(Y(f)) = \ln(X(f)) + \ln(H(f)) \tag{2.96}$$

Since the Fourier transform is linear, the addition remains valid in the cepstral domain.

$$C(\tau) = \mathscr{F}^{-1}\{\ln(Y(f))\} = \mathscr{F}^{-1}\{\ln(X(f))\} + \mathscr{F}^{-1}\{\ln(H(f))\} \tag{2.97}$$

This property indicates the possibility to deconvolve a signal if one of the factors is known. As such the logarithmic transformation allows the separation of the influence of the excitation source and the transmission path of the system in the cepstral domain. This property opens up possibilities for modal analysis in the cepstral domain [237], but this is not the focus of this chapter.

TQR-based speed estimation

The usage of the cepstrum for speed estimation is slightly different in this chapter compared to Bonnardot's approach. Instead of using a tracking approach based on the first five harmonics, the TQR is regarded as a probability density function map of the rhamonic orders similar to MOPA.

The TQR of a signal or Short-Time Cepstrum Transform (STCT) is essentially based on the STFT. The STCT is then simply the inverse Fourier transform of the natural logarithm of the absolute values of every spectrum in the STFT. The amount of windows remains the same:

$$STCT_x(m,\tau) = \frac{1}{N}\sum_{k=0}^{N-1} \ln(|STFT(m,k)|e^{-2\pi j\tau\frac{k}{N}}) \tag{2.98}$$

with τ the quefrency index. This STCT or "cepstrogram" is then used as input for the same formalism as defined in Section 2.4.4 for MOPA. The key difference between MOPA on the STFT and STCT is based on the fact that a decrease in rotation speed will actually lead to an increase in the quefrency peak related to that rhamonic order. This means that the STCT looks like it is inverted compared to the STFT. All the frequency intervals defined for the spectra are also inverted since the minimum and maximum expected rotation frequency correspond to the maximum and minimum rotation quefrency respectively.

2.4.7 Maximum tracking combined with Vold-Kalman filter

The final technique to be investigated is a combination of two commonly employed techniques and is based on the three-step procedure described in [220]. First, the spectrogram of the signal is calculated and used for a maximum tracking procedure. There exist quite a few maximum tracking algorithms, but the one showcased in this chapter is the one described in [220]. Second, this initial speed estimate based on the maxima serves as input for the Vold-Kalman filter, which is regarded as a time-varying band-pass filter in this context with the center frequency being the initial speed estimate. Third, the resulting, filtered signal should then be a mono-component signal and thus suitable for phase estimation through its analytic signal.

Maximum tracking

The idea behind maximum tracking is very straightforward: the amplitudes of a speed-related harmonic (or set of harmonics) are tracked over time in the spectrogram in an automated way by simply looking at the peaks (maxima) near the expected frequency and assuming that the found peaks are related to the harmonic to track. The manner in which this automated search and tracking is implemented varies greatly within literature [238] and there still does not seem to be a consensus as to what the optimal approach consists of. Regardless,

the approach in this chapter (as proposed in [220]) is based on solving a constrained optimisation problem in the form of:

$$\underset{k}{\text{minimize}} \; = -|STFT(n,k)|^2 \tag{2.99}$$

$$subject\ to\ (k\Delta f - f_c(n))^2 \leq \Delta f_c^2 \tag{2.100}$$

with f_c the center of the constraint, n the time index, Δf_c the bandwidth of the constraint, k the frequency index, and Δf the frequency resolution. It is assumed that:

$$f_{IF}(n\Delta t) \approx f_{max}(n) \tag{2.101}$$

with f_{IF} the actual instantaneous frequency of the harmonic to track, Δt the time resolution of the spectrogram, and f_{max} the frequency that corresponds to the maximum amplitude. In order to make the tracking process more robust, the acceleration of the instantaneous frequency (IF) can be taken into account. The Taylor series expansion of the IF gives:

$$f_{IF}(t) = f_{IF}(t-\Delta t) + \Delta t \frac{d}{dt} f_{IF}(t-\Delta t) + \frac{1}{2}\Delta t^2 \frac{d^2}{dt^2} f_{IF}(t-\Delta t) + ... \tag{2.102}$$

The gradients of the IF are of course unknown (and assumed to be continuous) and the actual f_{IF} is also unknown. Therefore, the gradients are estimated using a simple finite difference scheme based on the previous IF estimates $f_{max}(n-1)$, $f_{max}(n-2)$, etc. Next, it is assumed that:

$$f_{max}(n) = f_{IF}(n) + \nu(n) \tag{2.103}$$

with ν representing the deviation due to smearing of the STFT and noise in the signal. It is assumed that this deviation has a Gaussian distribution $\nu \sim N(0, \sigma^2)$ so that the estimated IF, f_{max}, can be related to the true IF, f_{IF}, as follows:

$$p(f_{max}(n)|f_{IF}(n), \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(f_{max}(n)-f_{IF}(n))^2}{2\sigma^2}} \tag{2.104}$$

A model of the true IF $f_{IF,w}$ is used since f_{IF} is unknown. An N^{th}-order polynomial is used for $f_{IF,w}$:

$$f_{IF,w} = w_0 + w_1 t^1 + w_2 t^2 + ... + w_N t^N \tag{2.105}$$

The polynomial weights $\mathbf{w} = [w_0, w_1, w_2, ..., w_N]^T$ are estimated through a maximum likelihood procedure:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}} \prod_{i=n-N_m}^{n-1} p(f_{max}(i)|f_{IF,w}(i), \sigma^2) \tag{2.106}$$

with N_m the number of previous time steps taken into account. The maximization of the likelihood function under a conditional Gaussian noise distribution for a linear model is equivalent to minimizing a sum-of-squares error function [239]. This boils down to solving a least squares problem which uses a Moore-Penrose pseudo-inverse. The resulting weights can then be found using following matrix expression:

$$\hat{\mathbf{w}} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{f}_{max} \tag{2.107}$$

with $\mathbf{f}_{max}$ the $(N_m \times 1)$ vector containing the previous N_m IF estimates, and $\mathbf{Q}$ the design matrix of the polynomial:

$$\mathbf{Q} = \begin{pmatrix} 1 & (n-1)\Delta t & \dots & ((n-1)\Delta t)^N \\ 1 & (n-2)\Delta t & \dots & ((n-2)\Delta t)^N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (n-N_m)\Delta t & \dots & ((n-N_m)\Delta t)^N \end{pmatrix} \tag{2.108}$$

In this dissertation a first order polynomial is used to minimize potential errors in the extrapolation. Also a limited number of previous points N_m is taken into account, namely five, for the cases described in this chapter to make sure that the computation time remains acceptable. The method does require estimates at the first time index for the IF and its gradient. The initial gradient is assumed zero and the initial IF estimate is obtained by visual inspection of the spectrogram.

The constrained minimization problem is reformulated using a penalised unconstrained cost function (see [220] for more details):

$$\kappa(\rho,n,k) = -|STFT(n,k)|^2 + \rho_1 \cdot max[0,(k\Delta f - f_{max}(n-1))^2 - \Delta f_{c1}^2] + \rho_2 \cdot max[0,(k\Delta f - f_{IF,w}(n))^2 - \Delta f_{c2}^2] \quad (2.109)$$

This cost function is minimized using a brute force approach which is computationally feasible as long as the number of points N_m is not too high. The bandwidth for the tracker is denoted by Δf_{c1} and Δf_{c2}. The parameters ρ_1 and ρ_2 are chosen in such a way that the cost function is dominated by the constraint terms in case the constraints get violated. The estimated IF is then given by:

$$f_{max}(n) = \Delta f \operatorname*{argmin}_k \kappa(\rho,n,k) \quad (2.110)$$

Vold-Kalman filter

The IF estimate returned by the maximum tracking is still quite rough due to the resolution limitations of the spectrogram. Therefore the Vold-Kalman filter (VKF) is employed as a time-varying band-pass filter with a center frequency based on the IF estimate returned by the maximum tracking. The VKF allows for defining a bandwidth of the band-pass filter such that the provided rough speed does not have to be perfect. The larger the bandwidth however, the more noise and extraneous components are included. The implementation used here is based on the one-pole angular-displacement filter as recommended by [220]. The angular-displacement VKF tries to estimate the envelope of the mono-component signal and should be fairly robust to crossing orders compared to the angular-velocity VKF. The full background of the VKF is not provided here, but interested readers are referred to [220, 240–242] for more details. Finally, after the VKF, the instantaneous speed estimate is found by using the analytic signal as described in Section 2.4.1.

2.5 Performance comparison on experimental data

The seven methods highlighted in Section 2.4 and the proposed multi-harmonic demodulation method are examined for their accuracy and ease of use on three different experimental data sets. One data set originates from a wind turbine gearbox, one from an aircraft engine, and one from a ship generator. Each data set has very different characteristics and thus their analysis can provide some interesting insights into the subtleties of each method. Since each method and data set require new input parameter settings, describing all of them each time would be quite cumbersome. Therefore only the most important settings that change between the different cases are reported and the remaining parameters are displayed in tables. A general overview of all the input parameters per method is provided in Table 2.1. The methods are abbreviated as follows:

- SHD = Single Harmonic Demodulation
- MHD = Multi-Harmonic Demodulation
- ISHD = Iterative Single Harmonic demodulation
- FDEO = Frequency Domain Energy Operator
- MOPA = Multi-Order Probabilistic Approach
- C-MOPA = Cepstrum-based Multi-Order Probabilistic Approach
- max.tr. + VKF = maximum tracking in the STFT with Vold-Kalman filtering

The list of input parameters per implemented method is undoubtedly subjective since another user might want to add or reduce certain inputs as to increase the flexibility for their particular case. However, the input parameters here are defined from a perspective that each method should be easy to use in an automated manner, thus with the least amount of manual effort possible. By looking at Table 2.1, it can be noticed immediately that the methods based on the STFT need at least three additional parameters just for the calculation of the STFT.

The method with the most required inputs is the combination of maximum tracking with the Vold-Kalman filter since this method is probably also the most complex method to implement out of the eight methods tested.

TABLE 2.1: Overview of the input parameters of each method.

Method name							
SHD	MHD	ISHD	FDEO	MOPA	ViBES	C-MOPA	max. tr. + VKF
Fs	Fs	Fs	Fs	Fs	Fs	Fs	Fs
ω_{init}	ω_{rough}	ω_{init}	ω_{init}	ω_{min}	N_w	ω_{min}	ω_{init}
Bw	Bw	Bw	Bw	ω_{max}	N_{FFT}	ω_{max}	Bw_{max}
N_w	N_w	N_w	$\{H_i\}$	$N_{overlap}$	$\{H_i\}$	$\{H_i\}$	N_w
	$\{H_i\}$		N_w	$\{R\}$	N_w	N_{FFT}	
				N_{FFT}	$\{Z\}$	N_{FFT}	$N_{overlap}$
				$N_{overlap}$	σ_l^2	$N_{overlap}$	N_p
				K_w	σ_h^2	K_w	N_m
				γ		γ	Bw_{VKF}
							N_{VKF}

with Fs the sampling rate in Hz, ω_{init} the IF at the first time index, ω_{rough} an initial rough IF estimation, Bw the bandwidth of the band-pass filter, N_w the window size used, $\{H_i\}$ the list of harmonic orders, N_{FFT} the size of the FFT in samples, $N_{overlap}$ the amount of overlap between windows in samples, K_w is the number of windows used for the continuity smoothing in MOPA, γ is the expected acceleration of the IAS in MOPA, $\{R\}$ the set of mesh ranges of interest with each range defined by a minimum and maximum value, $\{Z\}$ the set of illegal frequency regions, σ_l^2 & σ_h^2 the variance threshold values for beginning and stopping track- and wait-mode of ViBES, N_p the order of the polynomial used for maximum tracking, N_m the number of previous time steps taken into account, ρ_1 & ρ_2 are the weights for the penalised unconstrained cost function, Bw_{max} & Bw_{VKF} are the bandwidths for the maximum tracking and the vold-kalman filter respectively, and finally N_{VKF} the order of the vold-kalman filter.

2.5.1 Wind turbine gearbox data set

This well-documented data set originated from a diagnosis contest held at the International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMNO) in 2014 [187]. The provided vibration signal was measured on the gearbox housing of a wind

turbine near the epicyclic gear train and sampled at 20 kHz. The goal was to estimate the IAS of the high-speed shaft (carrying gear #7 in Fig.2.9). This estimate was then compared with a reference speed signal measured by an angle encoder. The length of the measurement was approximately 550 seconds. The spectrogram of the full signal can be seen in Fig 2.10 and is generated using a Hanning window of 1 second with 50% overlap.

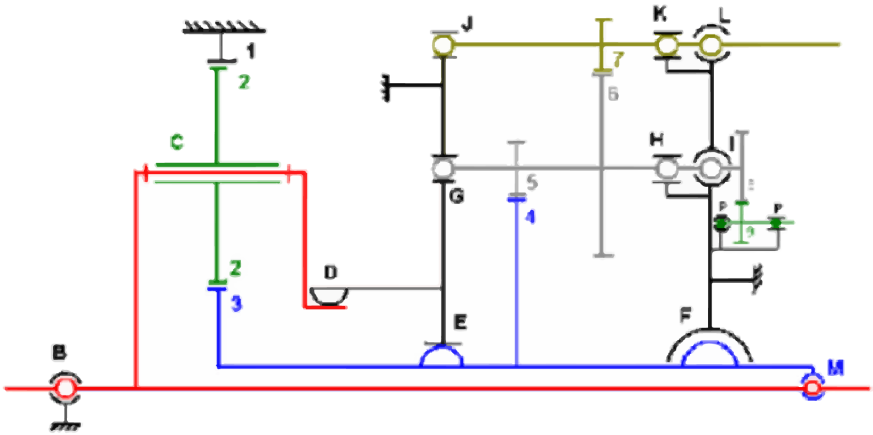

FIGURE 2.9: Visualization of the wind turbine gearbox used in the CMMNO 2014 diagnosis contest.

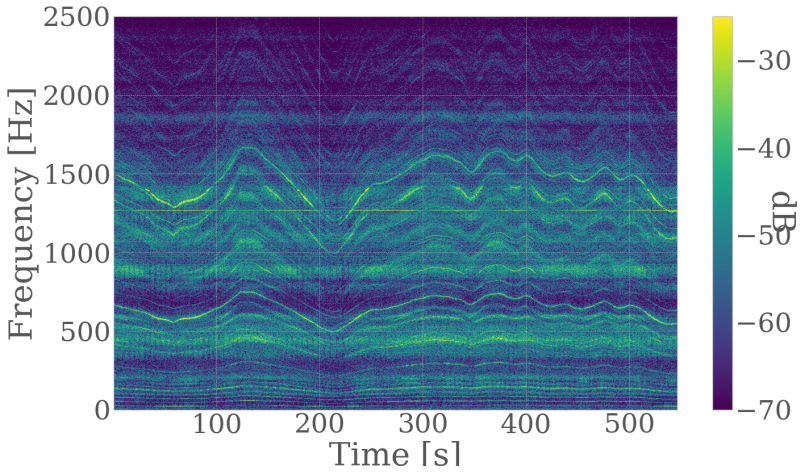

FIGURE 2.10: Spectrogram of the CMMNO 2014 diagnosis contest data.

2.5.1.1 Parameter settings

The methods based on phase demodulation mainly need one high SNR, speed-related harmonic as their most important input. In this case, the second harmonic of the planet gearmesh frequency around 55 Hz is chosen due to the good results obtained using it, which is also corroborated by the report in [187]. The bandwidth and window size for this case are set at 8 Hz and 10 seconds respectively. The methods based on the STFT all use the same settings for the STFT:

- uniform weighting window
- Window length of 1 second
- FFT size of 2 seconds (thus 1 second of zero padding)
- Overlap of 95% of the short time window length

The fundamental harmonic orders taken into account for the multi-order probabilistic approach are displayed in Table 2.2. The first ten harmonics of every fundamental order are considered for MOPA. The

number of windows K_w used for the a priori continuity introduction is set to 20 and the acceleration tolerance of the speed γ is set to $0.4\ \frac{Hz}{s}$. The benefit of using MOPA is that it allows the estimation of the speed of the full signal at once. The base frequency interval $\{\omega_{min}, \omega_{max}\}$ used here is chosen to be $\{15, 35\}$. The mesh ranges for the ViBES method are set to $R = \{[0.9290, 1.1219], [1.8580, 2.2438], [2.7871, 3.3657], [4.8167, 5.8167]\}$ and no illegal frequency ranges are defined. The first three of these ranges correspond to the fundamental, and 2nd and 3rd harmonics of the vibration content created by gear pairs 1/2 and 2/3 (Table 2.2), respectively, while the fourth range corresponds to the fundamental vibration content created by gear pair 4/5 (Table 2.2).

TABLE 2.2: Fundamental orders related to high-speed shaft.

Gear pair	**Order value**
1	1
2/3,1/2	1.025459229
4/5	5.316666667
6/7	29
8/9	15.225
10/11	6.619565217

Next, the multi-harmonic demodulation uses an initial rough speed estimation obtained by the zero crossings of the third harmonic of the 1.025th order. It then employs harmonics of all orders given in Table 2.2.

Finally, for the maximum tracking, the third harmonic of the planet gearmesh frequency which starts at about 75 Hz is chosen for ω_{init}. The bandwidth Bw_{max} is set to 2 Hz, the polynomial order N_p to 1, and the number of previous estimates to take into account N_m to 5. Lastly, the chosen bandwidth Bw_{VKF} is 4 Hz and the VKF order N_{VKF} is 2. An overview of every single input parameter is shown in Table 2.3.

TABLE 2.3: Overview of the input parameters for the CMMNO data set

Method name							
SHD	MHD	ISHD	FDEO	MOPA	ViBES	C-MOPA	max. tr.+VKF
$\omega_{init} = 53$ Hz	$Bw = 2$ Hz	$\omega_{init} = 53$ Hz	$\omega_{init} = 53$ Hz	$\omega_{min} = 15$ Hz	$\omega_{init} = 25$ Hz	$\omega_{min} = 15$ Hz	$\omega_{init} = 75$ Hz
$Bw = 8$ Hz	$N_w = 10000$	$Bw = 8$ Hz	$Bw = 8$ Hz	$\omega_{max} = 35$ Hz	$N_{FFT} = 2^{14}$	$\omega_{max} = 35$ Hz	Bw_{max}=2Hz
$N_w = 50000$	$\{H_i\}$ =Table 2.2	$N_w = 50000$	$N_w = 50000$	$\{H_i\}$ =Table 2.2	$N_{overlap}$=90%	$\{H_i\}$=Table 2.2	$N_w = 5000$
		$\{H_i\} = \{2, 10.62\}$		N_w=5000	N_w=10000	N_w=5000	$N_{FFT} = 10^4$
				$N_{FFT} = 10^4$	$\{R\} = \{[0.9290, 1.1219], [1.8580, 2.2438], [2.7871, 3.3657], [4.8167, 5.8167]\}$	$N_{FFT} = 10^4$	$N_{overlap}$=95%
				$N_{overlap}$=95%	$\{Z\} = \{\}$	$N_{overlap}$=95%	N_p=1
				K_w = 20	σ_h^2=[ ]	K_w= 20	N_m=5
				$\gamma = 0.4 \frac{Hz}{s}$	σ_l^2=[ ]	$\gamma = 0.4 \frac{Hz}{s}$	Bw_{VKF}=4 Hz
							N_{VKF}=2

2.5.1.2 Results

Some intermediary results are shown first in order to illustrate better the internal workings of the different methods. As described in Section 2.4.4, MOPA views the spectrogram as a probability density function (pdf) map and relates all of the pdf intervals belonging to different harmonic orders back to the fundamental speed interval, in this case $\{15, 35\}$. Figure 2.11 shows the obtained pdf map after summation and continuity introduction for the fundamental speed interval on the CMMNO data. The instantaneous speed profile can then be obtained by taking the expected value of every pdf at every time step.

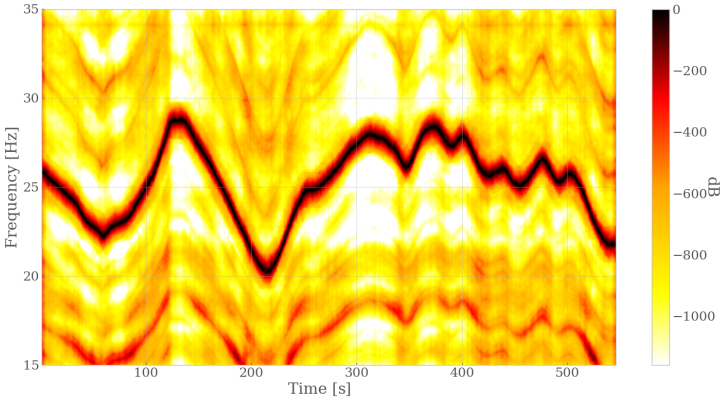

FIGURE 2.11: pdf map of the speed profile based on the CMMNO spectrogram after continuity introduction.

The same approach is employed for the cepstrum-based MOPA, but the STCT or cepstrogram is used instead of the spectrogram. Figure 2.12 displays the used cepstrogram with quefrency on the y-axis. The structure of the rhamonics clearly corresponds to the inverse of the instantaneous speed profile. Therefore the obtained pdf map, as shown in Fig. 2.13, also returns an inverted speed estimate. By taking the expected quefrency values of the pdf map and simply inverting them, the instantaneous frequency values are recovered.

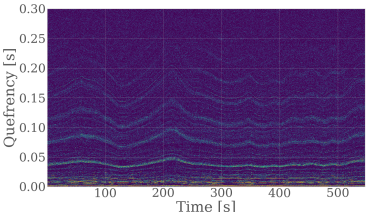

FIGURE 2.12: Cepstrogram of the CMMNO data.

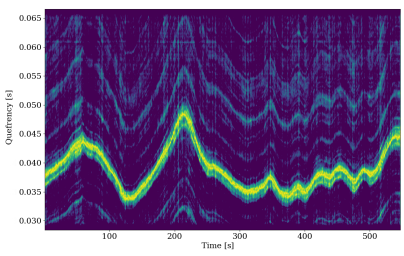

FIGURE 2.13: pdf map of the speed profile based on the CMMNO cepstrogram after continuity introduction.

The maximum tracking method basically calculates a 2D cost map

based on a penalized unconstrained cost function and the spectrogram. Figure 2.14 shows the cost map as an overlay in red over the spectrogram in the background. The cost map is red where the cost is very high and becomes more transparent where the cost is low. Clearly, the cost is lowest around the third harmonic of the planet gearmesh frequency since this harmonic lights up through the cost map layer. The input rough speed estimate for the VKF is then acquired by taking the frequency that corresponds to the minimum cost at every time step.

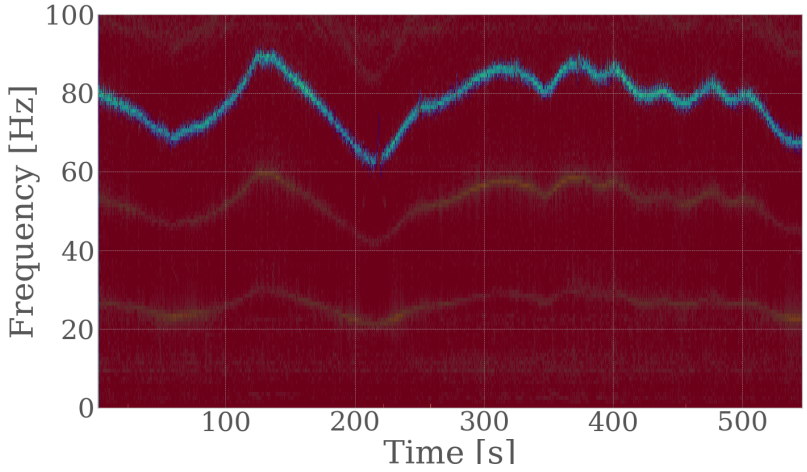

FIGURE 2.14: Cost map of the CMMNO data as calculated by the maximum tracking algorithm using a penalized unconstrained cost function.

The extracted speed profiles of the eight different methods are then compared to the reference speed signal measured by the angle encoder in Fig. 2.15. It can be seen that most methods perform quite well in tracking the overall profile of the speed. The cepstrum-based approach shows the largest discrepancies to the encoder speed. This is mainly attributable to the poor quefrency resolution and the fact that not all rhamonics are as well-pronounced as their harmonic spectral counterparts since the cepstrum is strongly influenced by noise. The maximum tracking and VKF approach exhibits one major deviation

of the encoder speed around 217 seconds. This deviation is primarily due to the maximum tracking being influenced by a very short, sudden drop in energy of the tracked harmonic. This is also the main problem of this method. If the noise is too high, if there are crossing orders, or if there is a short drop in the amplitude of the tracked harmonic, the maximum tracking can jump quickly to the wrong frequency bins and this is not always straightforward to control in an automated manner.

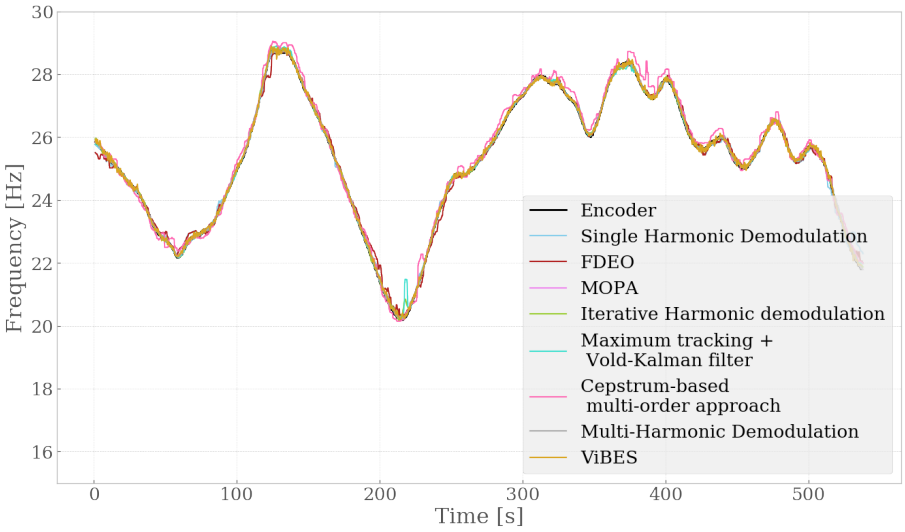

FIGURE 2.15: Estimated instantaneous speed profiles of every method on the CMMNO data.

Since it is fairly difficult to assess the accuracy of each method visually in this manner, the mean and median absolute errors are calculated and displayed in Fig. 2.16. In general, the errors are quite low apart from the one of the cepstrum-based MOPA as is explained earlier. The best result is obtained by the newly proposed multi-harmonic demodulation, closely followed by MOPA. The main reason why the MHD and the spectrum-based MOPA work so well is thanks to the well-defined harmonic structure in the spectrogram and the large number of different harmonic orders that do not overlap excessively. The MHD method benefits heavily in this case from taking into account more

harmonics than the SHD. Likewise, averaging the resulting pdfs together of over 60 harmonics produces a very accurate and smooth result for MOPA. The ViBES method only performs a bit worse than MOPA which is probably due to the difference in implementation and the different harmonic orders chosen. It comes as no surprise that the iterative phase demodulation performs better than the single step phase demodulation and FDEO. The latter two methods mainly suffer from the fact that a fixed window size was chosen for the entire signal. These methods would get closer to the result of the iterative demodulation if custom window sizes and filter bandwidths were defined for every signal part depending on the speed fluctuation. The iterative phase demodulation does not suffer from this drawback since after the first angular resampling it is possible to define a narrow filter bandwidth for the entire signal at once. A similar reasoning can be followed for the VKF approach since after the VKF the signal is essentially a mono-component signal, meaning that phase demodulation of the full-bandwidth signal is possible.

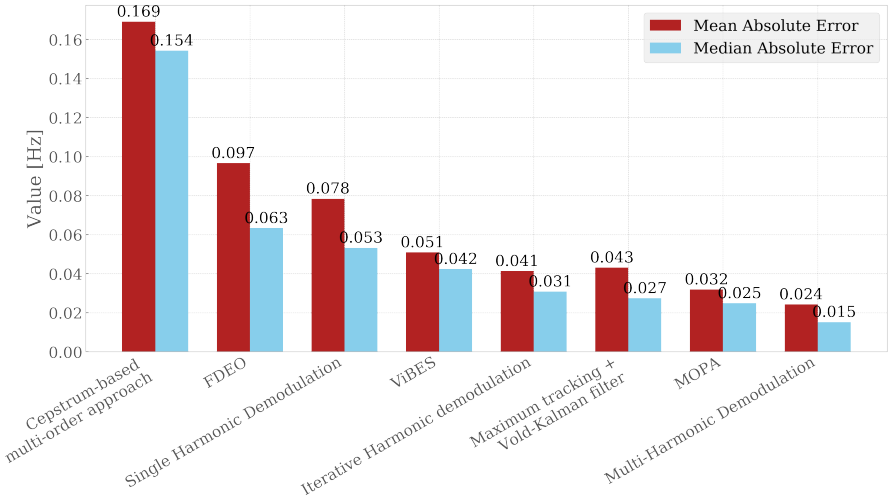

FIGURE 2.16: Mean and median absolute errors for every method on the CMMNO data using the encoder as reference.

To provide some insight into the computational effort required by

the eight methods, the execution time of every method is returned in Fig. 2.17. All the computations were done on the same computer using single threaded computation on one CPU core of an i5-5300U processor with 16GB of RAM. Deriving all the computational complexities of every single method is not the main focus here, but Fig. 2.17 aims to present a ballpark indication of the computation time such that individuals who are interested in implementing a fast speed estimation method have something to judge from.

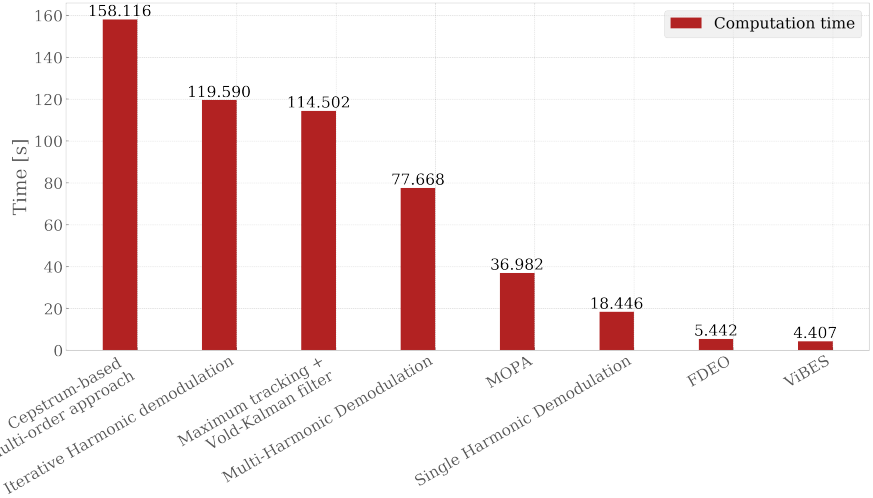

FIGURE 2.17: Computation time for every algorithm on the CMMNO data.

2.5.2 Aircraft engine data

The second data set to be analyzed originates from the Safran contest at the Surveillance 8 conference, held in Roanne, France [236]. The provided data consists of vibration and tachometer signals acquired during a ground test campaign on a civil aircraft engine with two damaged bearings.
A general overview of the engine with the damaged bearings and the sensors locations is displayed in Fig. 2.18. The engine has two main

shafts and an accessory gearbox with equipment such as pumps, filters, alternators, and starter. The accessory gearbox is linked to the high-pressure shaft HP by a radial drive shaft RDS and a horizontal drive shaft HDS. The kinematics of the gearbox and the rotating speeds of its shafts are described in Fig. 2.18. A spectrogram of the analyzed signal of accelerometer 2 can be seen in Fig. 2.20. It is generated using a Hanning window with a length of 2^{11} samples with an overlap of 95%.

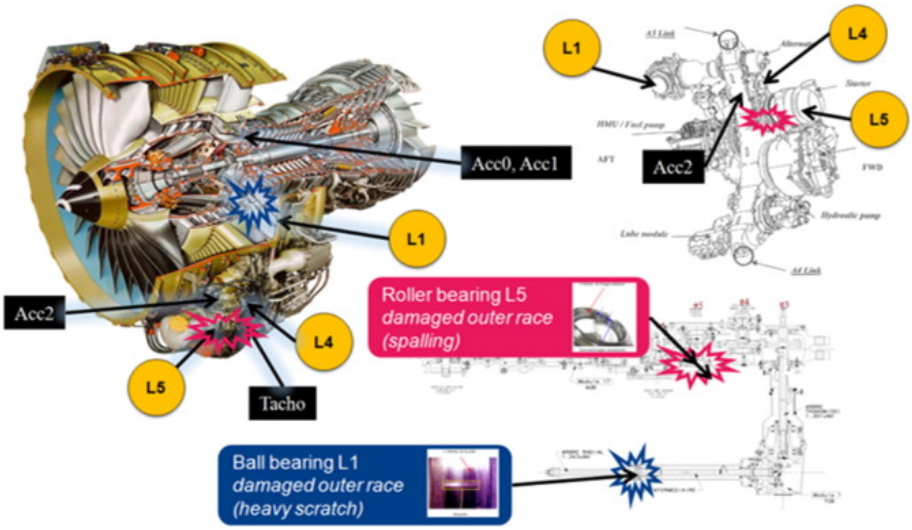

FIGURE 2.18: General overview of the engine and the accessory gearbox. Shafts are identified by labels in amber color.

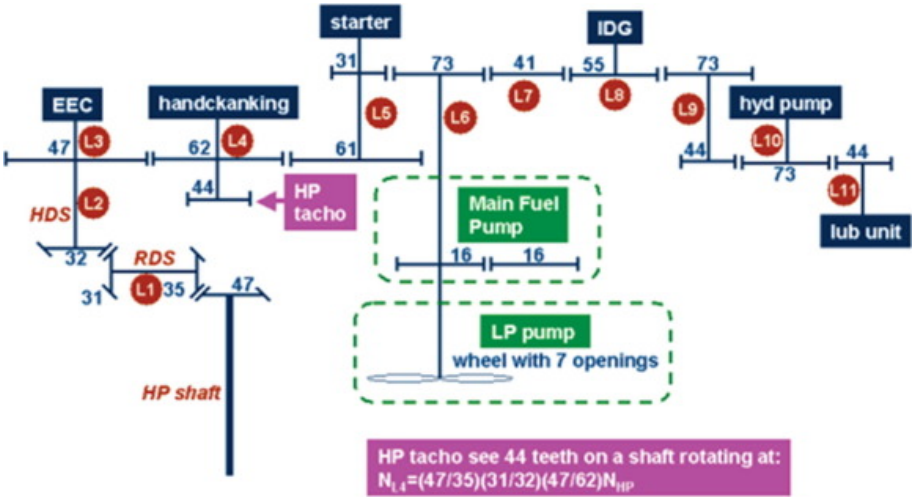

FIGURE 2.19: Diagram of the kinematics of the gearbox.

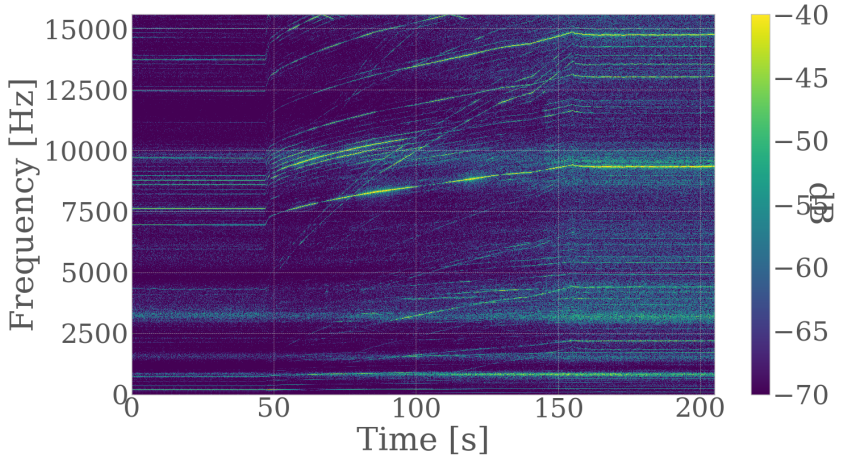

FIGURE 2.20: Spectrogram of the analyzed Surveillance 8 aircraft engine vibration data.

2.5.2.1 Parameter settings

The main harmonic used for the single harmonic phase demodulation techniques is the 38th harmonic of the high-pressure shaft. The chosen window size is only 1 second due to the very rapid speed increase. The bandwidth is therefore also quite large at 100 Hz. The multi-harmonic demodulation uses the first 60 harmonics of the high-pressure shaft with a moving window size for the MLE part of 2^{17} samples. The rough speed is estimated by using the zero crossings of the filtered first harmonic of the high-pressure shaft.

The parameters for MOPA are as follows:

- harmonics orders used: first 60 harmonics of the HP shaft & L1 shaft
- $\{\omega_{min}, \omega_{max}\}$ = $\{175$ Hz, 230 Hz $\}$
- N_w = N_{FFT} = 2^{14}
- $N_{overlap}$ = $0.9N_w$
- K_w = 20
- γ = 0.4

The mesh ranges for the ViBES method were set to correspond to the fundamental through 8th harmonic of the vibration created by the HP shaft. Table 2.4 displays all of the input parameter settings for this case. The main differences are the different harmonic orders and the adjustments necessary to deal with the very rapid speed increase.

TABLE 2.4: Overview of the input parameters for the Surveillance 8 aircraft engine data set

Method name							
SHD	MHD	ISHD	FDEO	MOPA	ViBES	C-MOPA	max. tr.+VKF
$\omega_{init} = 6960$ Hz	$Bw = 5$ Hz	$\omega_{init} = 6960$ Hz	$\omega_{init} = 6960$ Hz	$\omega_{min} = 175$ Hz	$\omega_{init} = 178$ Hz	$\omega_{min} = 175$ Hz	$\omega_{init} = 6960$ Hz
$Bw = 100$ Hz	$N_w = 2^{17}$	$Bw = 100$ Hz	$Bw = 100$ Hz	$\omega_{max} = 230$ Hz	$N_{FFT} = 2^{15}$	$\omega_{max} = 230$ Hz	Bw_{max}=50Hz
$N_w = 2^{14}$	$\{H_i\} = \{1\}$	$N_w = 2^{14}$	$N_w = 2^{14}$	$\{H_i\} = \{1, 1.342\}$	$N_{overlap}$=90%	$\{H_i\} = \{1, 1.342\}$	$N_w = 2^{14}$
		$\{H_i\} = \{38, 75\}$		N_w=2^{14}	N_w=31200	N_w=2^{14}	$N_{FFT} = 2^{14}$
				$N_{FFT} = 2^{14}$	$\{R\} = \{[0.9375, 1.0625], [1.8750, 2.1250], [2.8125, 3.1875], [3.7500, 4.2500], [4.6875, 5.3125], [5.6250, 6.3750], [6.5625, 7.4375], [7.5000, 8.5000]\}$	$N_{FFT} = 2^{14}$	$N_{overlap}$=90%
				$N_{overlap}$=90%	$\{Z\} = \{\}$	$N_{overlap}$=90%	N_p=1
				K_w = 20	σ_h^2=[ ]	K_w= 20	N_m=5
				$\gamma = 0.4 \frac{Hz}{s}$	σ_l^2=[ ]	$\gamma = 0.4 \frac{Hz}{s}$	Bw_{VKF}=4Hz
							N_{VKF}=2

2.5.2.2 Results

Figure 2.21 shows the obtained pdf map after summation and continuity introduction for the fundamental speed interval on the Surveillance 8 aircraft engine data. The instantaneous speed profile is again obtained by taking the expected value of every pdf at every time step. It can be seen from the continuity map that the very sudden increase in rotation speed around 50 seconds causes the certainty of the extracted expected value to decrease due to the smearing of all the harmonics, illustrated by the vertical red line.

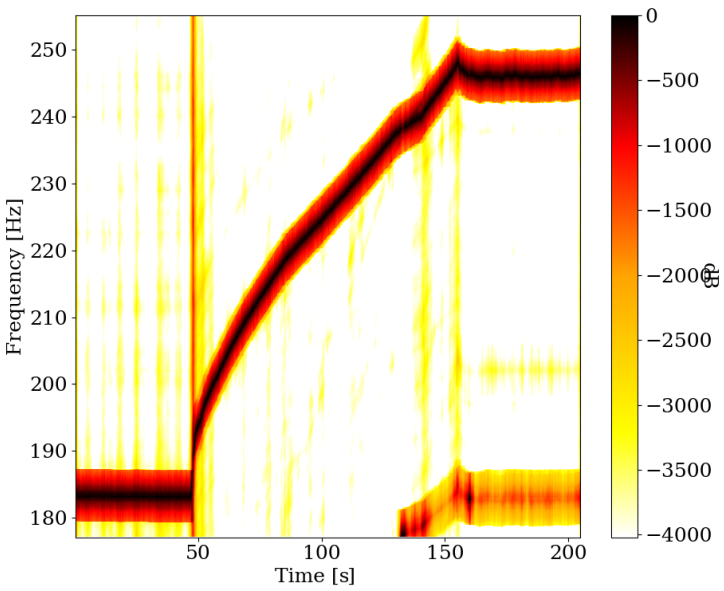

FIGURE 2.21: Pdf map of the speed profile based on the Surveillance 8 spectrogram after continuity introduction.

Figure 2.22 displays the cepstrogram. The obtained pdf map after continuity introduction, as shown in Fig. 2.13, portrays quite clearly the instantaneous speed profile, but suffers from the same problems as the spectrum-based MOPA around 50 seconds.

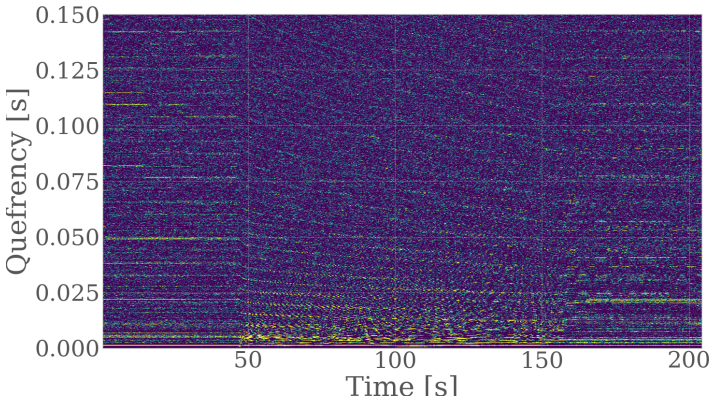

FIGURE 2.22: Cepstrogram of the Surveillance 8 aircraft engine data.

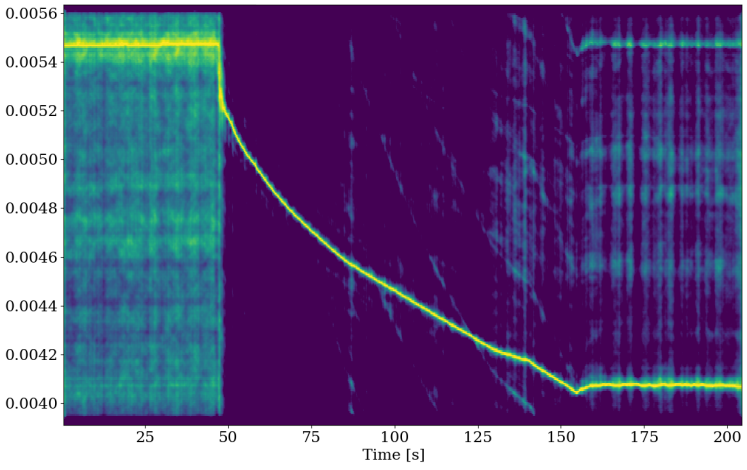

FIGURE 2.23: Pdf map of the speed profile based on the Surveillance 8 cepstrogram after continuity introduction.

Figure 2.24 shows the cost map for the Surveillance 8 data as an overlay in red over the spectrogram in the background. The cost map

is red where the cost is very high and becomes more transparent where the cost is low. The cost is lowest around the 38th harmonic of the HP shaft frequency illustrated by the visibility of the spectrogram coloring from 6.9 kHz to 9.3kHz.

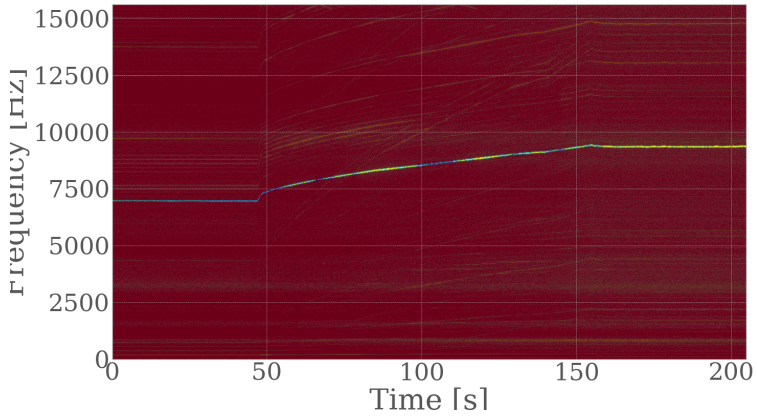

FIGURE 2.24: Cost map of the Surveillance 8 data as calculated by the maximum tracking algorithm using a penalized unconstrained cost function.

A comparison of the estimated speed profiles is provided in Fig. 2.25. All methods are successful in estimating the general outline of the speed, albeit with varying degree of accuracy. The cepstrum-based MOPA again suffers from resolution issues causing the estimated speed to be quite choppy. The FDEO method manages to track the speed very well visually but suffers from a temporary drop in amplitude and thus signal-to-noise ratio of the 38th harmonic around 105 seconds.

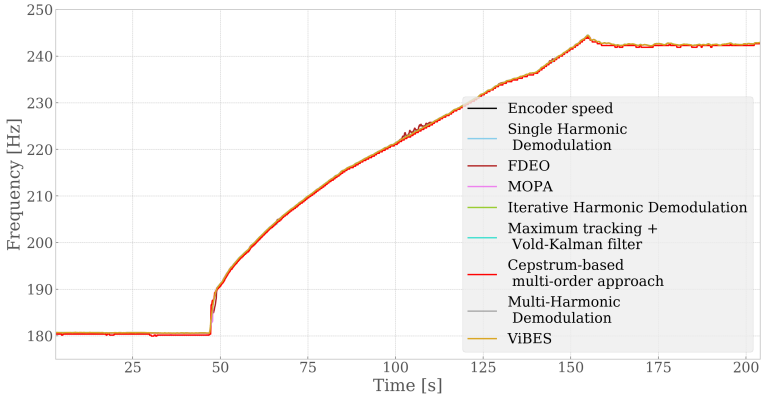

FIGURE 2.25: Estimated instantaneous speed profiles of every method on the Surveillance 8 aircraft engine data.

Again the mean and median absolute errors are calculated and shown in Fig. 2.26 to assess the performance more objectively. The cepstrum-based MOPA performs the worst, which is expected when looking at the speed profiles. Interestingly, the phase demodulation methods perform the best. This can probably be explained by the fact that the 38^{th} harmonic has a very high signal-to-noise ratio overall and does not have any significant crossing orders, leading to very clean demodulation results. The FDEO method also has a very low median absolute error but has a large mean error due to the erroneous tracking around 105 seconds. The best results are obtained by the iterative harmonic phase demodulation method. Almost identical results are found using the multi-harmonic demodulation, albeit ever so slightly worse. In this case it seems like incorporating more harmonics than the most dominant ones does not yield significant improvements. This is confirmed by looking at the harmonic weights estimated by the MLE optimization of the MHD method shown in Fig. 2.27. This figure indicates that also the MHD method prefers to mainly use the 38^{th} harmonic for demodulation.

Another observation experienced during the testing of the methods based on the spectrogram is that the resulting errors were quite sensitive to the choice of the set of harmonic orders, more so than for the CMMNO case. Increasing the set of harmonic orders did not always yield a better result compared to simply choosing the highest amplitude orders, which suggests potential for order set optimization.

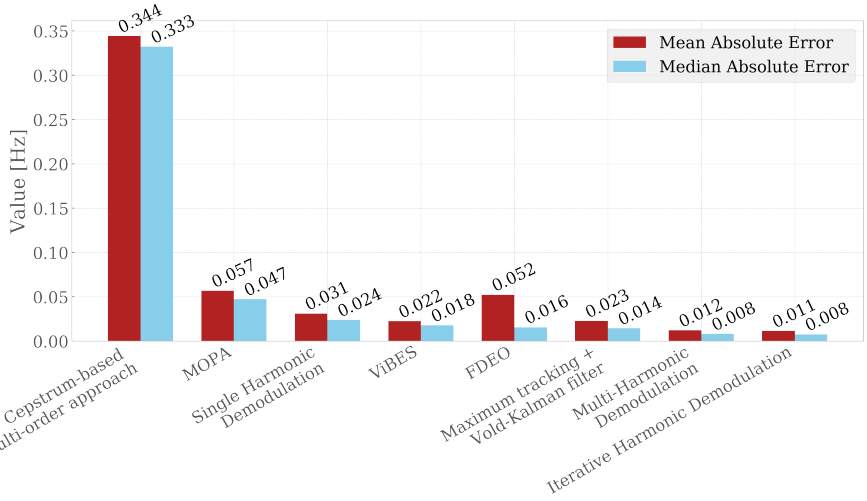

FIGURE 2.26: Mean and median absolute errors for every method on the Surveillance 8 data using the encoder as reference.

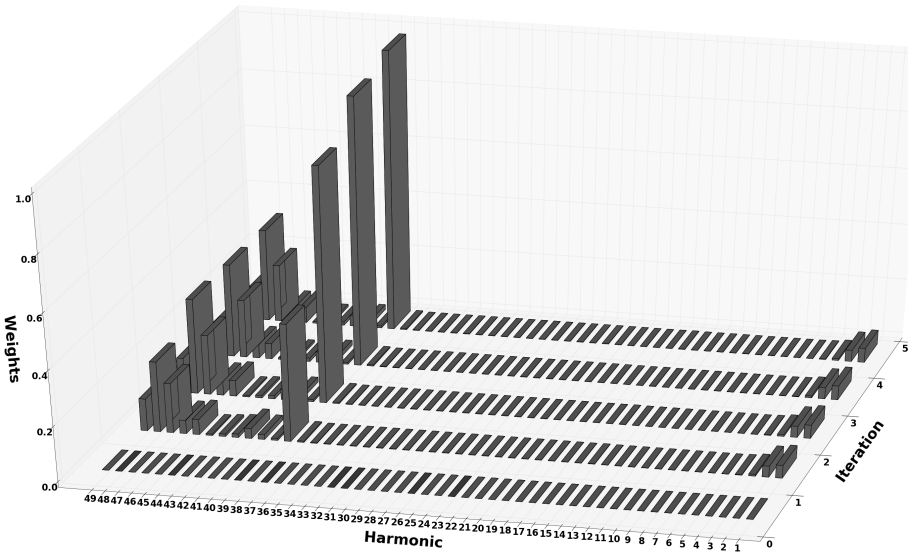

FIGURE 2.27: Estimated harmonic weights for the MHD method.

Lastly, the computation times are displayed in Fig. 2.28. The calculation of the cost map for the maximum tracking is the main contributor for the maximum tracking with VKF combination, due to the large size of the spectrogram.

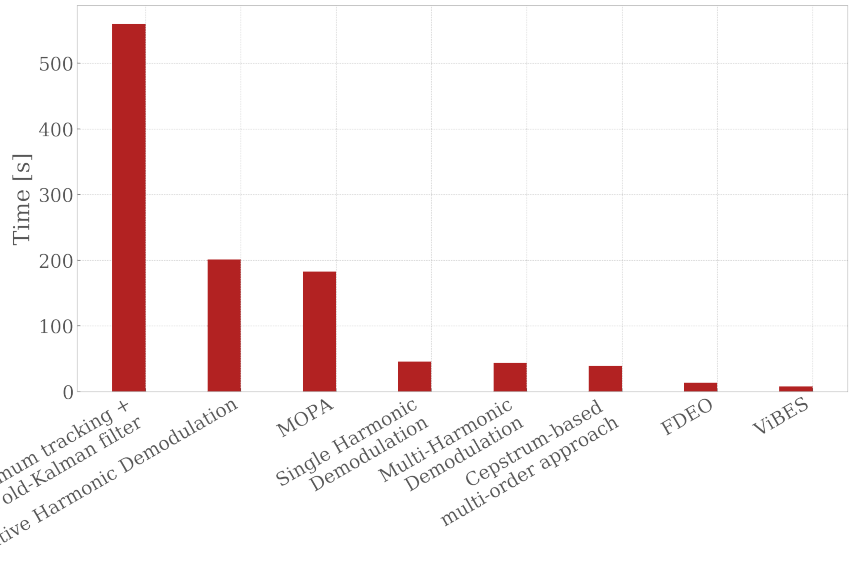

FIGURE 2.28: Computation time for every algorithm on the Surveillance 8 data.

2.5.3 Ship generator data

The third and last data set to be examined is measured on one of the main generators of a YP-700 US navy ship as seen in Fig. 2.29. Each Yard patrol craft has two generators but only one runs at a time. The generator, shown in Fig. 2.30, is a Detroit Diesel 3-71 and has a power output of 50 kW at 450 V_{AC}. The electrical system is a three-phase system at 60 Hz. The cause for the excessive speed fluctuation of the generator is a bad governor, displayed in Fig. 2.31. An accelerometer was placed close to the base of the generator as shown in Fig. 2.32. The used measurement is approximately 200 seconds in duration at a sample rate of 5 kHz. Interestingly, it contains two fast run-downs and one very fast run-up. These extreme speed fluctuations form quite a significant hurdle to overcome for most of the speed estimation methods as there is also a complete standstill part in-between the first run-down and run-up. The spectrogram of the signal is shown in Fig. 2.33 and exhibits a large amount of speed-related harmonics. Regrettably,

there is no angle encoder or speed reference available for this data set. Therefore a quantitative assessment of estimation errors is not possible, but the authors do believe that highlighting the potential issues that come into play when analyzing this type of data is of some importance to investigate.

FIGURE 2.29: Navy ship on which the generator is installed.

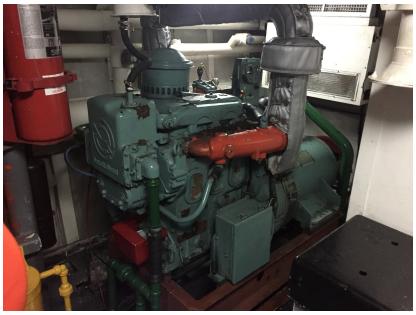

FIGURE 2.30: Generator.

FIGURE 2.31: Faulty governor.

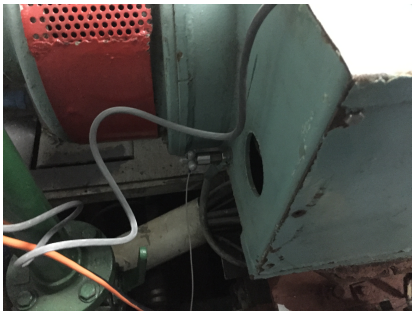

FIGURE 2.32: Accelerometer placement.

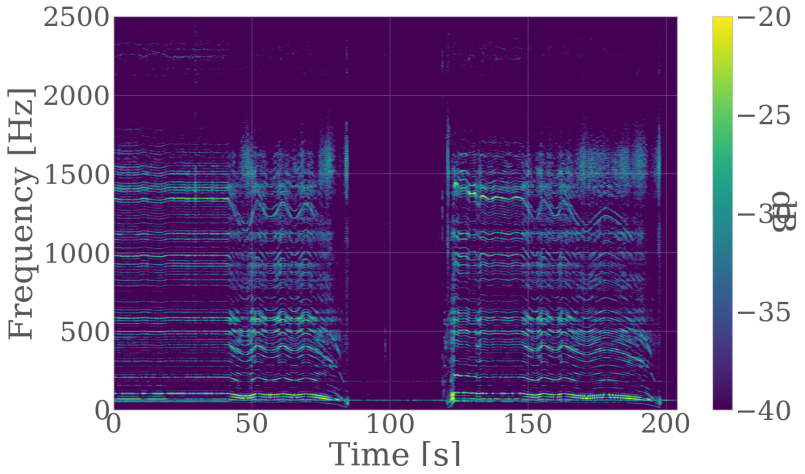

FIGURE 2.33: Spectrogram of the analyzed ship generator vibration data.

2.5.3.1 Parameter settings

Since there is no reference speed, only the single stage phase demodulation is performed. This means neither the iterative nor the multi-harmonic demodulation method is showcased here because both suffer from the drawback that demodulation methods have here, being the difficulty of defining a filterband for the temporary standstill around 100 secs. The input parameters are shown in Table. 2.5. For the demodulation the first harmonic of the generator shaft is tracked which always stays right below the electrical line frequency of 60 Hz. The MOPA method takes into account the sub-harmonics of the fundamental frequency at multiples of one third its frequency. The mesh ranges for the ViBES method were set to correspond to the first seven components of the vibration spectrum that are correlated to the generator shaft speed. The ViBES method applies the additional adaptations explained in Section 2.4.5 in the analysis of the ship generator data. The first adaptation, related to the "illegal" frequency regions, is applied by defining $\mathbf{Z}$. $\mathbf{Z}$ contains two regions, $\mathbf{z}_1 = [58.5, 61.5]$ Hz and $\mathbf{z}_2 = [178.5, 181.5]$ Hz. These "illegal" regions of the vibration

spectrum contain disturbances corresponding to the fundamental and third harmonic of the electrical line frequency, respectively. Also the second adaptation, related to the wait-and track-mode, is employed. The thresholds are set to σ_l^2 = 0.2 Hz^2 and σ_h^2 = 0.4 Hz^2.

TABLE 2.5: Overview of the input parameters for the ship generator data set

Method name					
SHD	FDEO	MOPA	ViBES	C-MOPA	max. tr. + VKF
$\omega_{init} = 53$ Hz	$\omega_{init} = 53$ Hz	$\omega_{min} = 12$ Hz	$\omega_{init} = 53$ Hz	$\omega_{min} = 12$ Hz	$\omega_{init} = 53$ Hz
$Bw = 10$ Hz	$Bw = 10$ Hz	$\omega_{max} = 60$ Hz	$N_{FFT} = 2^{13}$	$\omega_{max} = 60$ Hz	$Bw_{max} = 3$ Hz
$N_w = 5000$	$N_w = 5000$	$\{H_i\} = \{0.333\}$	$N_{overlap}$=96%	$\{H_i\} = \{0.333\}$	$N_w = 5000$
		N_w=5000	N_w=5000	N_w=2^{14}	$N_{FFT} = 5000$
		$N_{FFT} = 5000$	$\{R\}$= $\{[0.9444, 1.0556]$, $[1.2593, 1.4074]$, $[1.5741, 1.7593]$, $[1.8889, 2.1111]$, $[2.2037, 2.4630]$, $[2.5185, 2.8148]$, $[2.8333, 3.1667]\}$	$N_{FFT} = 2^{14}$	$N_{overlap}$=99%
		$N_{overlap}$=98%	$\{Z\}$=$\{[58.5, 61.5]$, $[178.5, 181.5]\}$	$N_{overlap}$=90%	N_p=1
		K_w = 20	σ_h^2=$0.4Hz^2$	K_w= 20	N_m=5
		$\gamma = 0.4\frac{Hz}{s}$	σ_l^2=$0.2Hz^2$	$\gamma = 0.4\frac{Hz}{s}$	Bw_{VKF}=3Hz
					N_{VKF}=2

2.5.3.2 Results

The run-down and run-up approximately in the middle of the data record make it difficult for some of the techniques to track the speed of the signal in one go, at least not without some additional measures like for example the second adaptation of the ViBES method in Section 2.4.5. The second adaptation of the ViBES method allows the algorithm to track the second region of operation starting at the 123 second mark after losing tracking of the first operation after 82 seconds. Since the demodulation methods require an initial speed estimate to start tracking a harmonic, this leads to erroneous results when these methods try to track the run-up again due to the standstill in-between causing the tracking to go haywire. Therefore, the data is

processed in two separate parts for the phase demodulation, FDEO and maximum tracking method. The MOPA method does not have this drawback because only a speed interval needs to be defined and it automatically gets back "on track" the moment the most likely speed estimate emerges within that speed interval again. This property is usually sufficient for most cases since the focus is mostly on the operating regimes at higher rotation speeds as there is often little to learn from near standstill data from a fault monitoring perspective.

The cepstrogram is shown in Fig. 2.34 together with the generated pdf map after continuity introduction in Fig. 2.35. As expected, the probability density functions jump around in frequency due to smearing during the standstill part around 100 seconds and at the end of the record after the run-down. The maximum tracking is done in two separate parts due to the necessity for an initial frequency estimate and the inability of the method to track the very fast run-up at 120 seconds. This run-up goes from 0 Hz to 55 Hz in approximately 4 seconds which causes all the harmonics to smear together in the spectrogram, making maximum tracking practically impossible. The same issue occurs for the phase demodulation methods since the signal has a low SNR in the beginning of the run-up and due to the difficulty in defining a proper band-pass filter. Figure 2.36 & 2.37 display the cost maps for the first and second part of the signal on the fundamental harmonic order. This order is chosen because it is reasonably well separated from the other harmonics and the deceleration of the run-down is not too extreme as compared to higher harmonics.

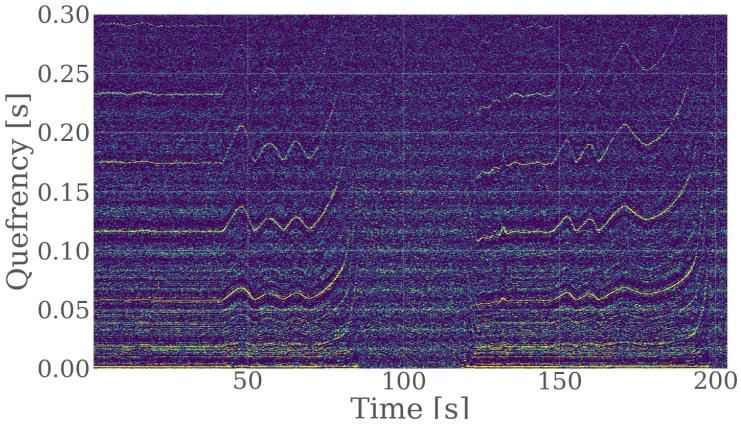

FIGURE 2.34: Cepstrogram of the ship generator data.

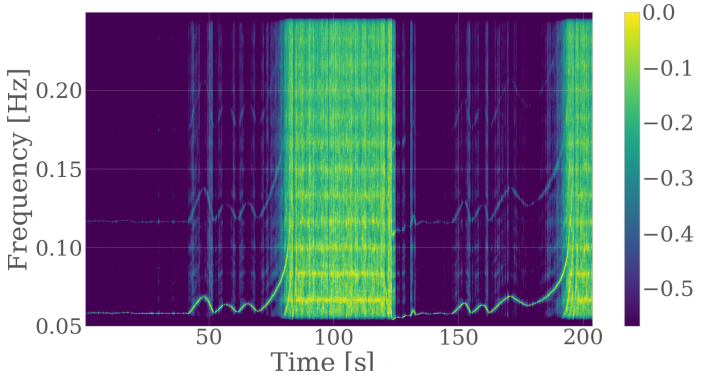

FIGURE 2.35: Pdf map of the speed profile based on the ship generator cepstrogram after continuity introduction.

Unfortunately, there was no reference speed provided with these measurements, so only a qualitative visual assessment of the speed profiles can be made. Figure 2.38 shows the estimated speed profiles and it distinctly showcases the issues that arise when a fast run-down

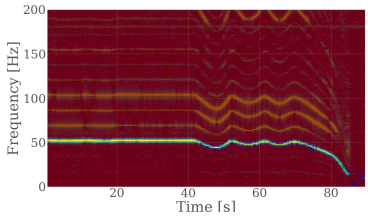

FIGURE 2.36: Cost map of the first part of the ship generator data as calculated by the maximum tracking algorithm using a penalized unconstrained cost function.

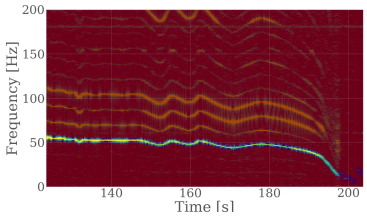

FIGURE 2.37: Cost map of the second part of the ship generator data as calculated by the maximum tracking algorithm using a penalized unconstrained cost function.

or run-up occurs. The MOPA method is able to continuously track the speed but produces meaningless results during the standstill. The phase demodulation method, maximum tracking with VKF, and FDEO method are unable to track the speed in one go, but perform better in tracking the run-down than MOPA as they produce a sensible result down to 18 Hz approximately. Clearly, the methods based on tracking a specific harmonic within a frequency band need additional built-in intelligence in order to make them cope with sudden run-downs and run-ups as this would make them more flexible to use in an industrial setting.

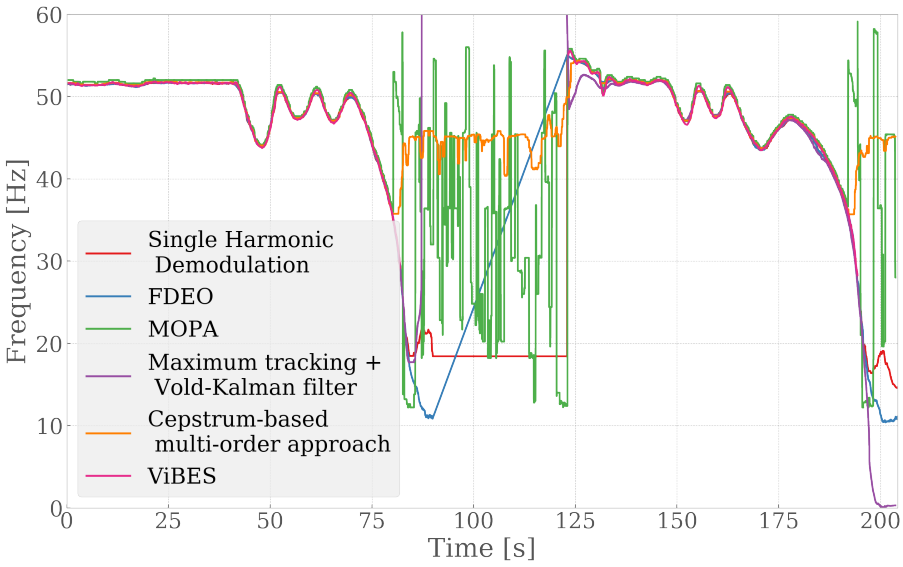

FIGURE 2.38: Estimated instantaneous speed profiles on the ship generator data.

2.6 Discussion

Based on the experimental results of the speed estimation comparison, there is no absolute outcome as in which method is the most accurate for every case and which method is always the easiest to use. Nevertheless, some general comments can be made.

The signal-to-noise ratio, with the harmonics being the signal and all the other signal content being the noise, logically influences all the speed estimation methods mentioned in this chapter. The methods that are based on a single harmonic are more dependent on the signal having a high SNR, well-separated harmonic as compared to the multi-harmonic methods. Therefore, in situations where there is not a single harmonic that is well excited, the multi-harmonic methods should outperform the single harmonic methods due to a reduction in variance by taking into account multiple harmonics.

The single harmonic phase demodulation method produces very satisfactory results as long as there is a high SNR harmonic present in the signal without any significant crossing orders and well separated from nearby harmonics. Doing an additional step of angular resampling and phase demodulation of a higher harmonic order can improve the result even further but the limit of the accuracy gain is usually reached after one to two iterations. The multi-harmonic demodulation method is similar in approach as the iterative demodulation approach, but allows for more leniency towards the end-user in choosing the right harmonics. A whole range of potential harmonics can be given as an input. The maximum likelihood estimation will then decide for you which harmonics are best for the speed estimation. The downside to the demodulation methods is the need for an initial or rough speed estimate. This means that when there is a total standstill such as in the third data set, the phase demodulation methods fail to start tracking the right harmonic again by themselves since they have been demodulating noise during the standstill. This could be improved by adding some intelligence to the tracking such as described by the variance-based lock-in tracking adaptation in Section 2.4.5. The same remarks are valid for the FDEO method.

The methods based on a TFR or TQR of the signal are able to continuously track the speed since they estimate the most likely speed at every time index and they only require an approximate speed range. The performance of the cepstrum-based MOPA is consistently the worst due to the poor resolution and sensitivity to noise, but it is easy to use since the harmonic orders are usually known a priori from the kinematics of the investigated system. The spectrum-based MOPA and ViBES method are two similar takes on perceiving the spectrogram as a probability density function map, but with different implementations and adaptations. The accuracy is very similar for both methods and depends mostly on the chosen inputs. The main drawback of this type of techniques is the dependency on the frequency and time resolution of the TFR or TQR. If the speed fluctuates very strongly, the harmonics tend to smear together in the short-time spectrum making it difficult to distinguish them and causing the obtained pdfs to be smeared as well. It is also difficult to capture small, but fast speed fluctuations

since the time resolution is often too coarse to identify such variations.

In the future a possibility to automate this search for the optimal window length choice can be investigated. An optimization procedure could be implemented that can for example be based on the maximum acceleration of the speed or be based on a statistical quantity of the time-frequency representation. Another potential improvement would be to remove the a-priori assumption about the maximum physically possible speed acceleration. Instead, the maximum acceleration could be tracked and updated over time based on the previous time windows. The idea here is to make the Gaussian pdfs used in MOPA as narrow as possible for each time step.

The combination of the maximum tracking and the Vold-Kalman filter tries to circumvent this issue by using the VKF as a time-varying band-pass filter such that the phase can be obtained from the resulting analytic signal. The main issue however in utilizing the VKF is the large number of adjustable parameters that can impact significantly the performance of the method. This encumbers somewhat the practicality of the approach and makes it probably the most complex technique out of the eight tested methods. Also the maximum tracking used in this chapter is based on calculating a cost map of the spectrogram which can take quite a while to calculate and needs an initial speed estimate. A potential improvement can probably be made by combining a technique such as MOPA with a time-varying band-pass filter such as the VKF.

Finally, if there is no tacho or angle encoder available to compare the estimated speed to, then the only sources of information available are the estimated speed signal itself and the vibration signal. It is reasonable to assume that the speed profile needs to be continuous and cannot be jumping around like e.g. in Fig.2.38. A possible way to grade the estimation therefore could be based on the acceleration of the speed signal. The vibration signal itself can also be employed: if for example the RMS of the signal drops below a certain threshold, you can assume the machine is in standstill condition and thus the estimated speed can be ignored. Such approaches offer a more automated way of assessing the quality of the estimated speed compared

to visually inspecting it.

2.7 Conclusion

This chapter proposes a novel instantaneous speed estimation method, coined the multi-harmonic demodulation (MHD) method. It also investigates seven other instantaneous angular speed estimation methods based on the most commonly used principles, namely phase demodulation and tracking in a time-frequency representation of the vibration signal and compares them on experimental data. It is clear from Section 2.2 that there exists a large number of possible variations and extensions to these two basic principles. While it is impossible to investigate all of these variations, this chapter still aims to shed some light on the strengths and drawbacks of different method implementations by assessing their performance on experimental data. A short overview of the different theoretical backgrounds of each method is provided in Section 2.4. Three experimental vibration data sets are investigated: one was measured on a wind turbine gearbox, one on an aircraft engine, and one on the generator of a ship. Section 2.5 discusses the results obtained by applying the eight methods on these three data sets. While every method is able to track the general speed profile for each case, the level of manual involvement in tweaking the input parameters for every method differs greatly, as does the resulting accuracy. The performance of every method is discussed in Section 4.5. Based on the assessment provided in this chapter, it is evident that effective speed estimation methods already exist, but that there are still improvements to be made. It should be noted though that the proposed MHD method shows promising results in both accuracy and applicability.

In general, the main challenges for speed estimation methods continue to be sudden fast speed fluctuations, operating regime changes that influence the harmonic structure, and accurate continuous tracking, even at low speeds.

Chapter 3

Discrete-random signal separation

3.1 Introduction

An essential step in many vibration-based condition monitoring approaches is to separate contributions from components that produce deterministic/discrete signal content, such as gears and shafts, from components that produce stochastic/random content, such as rolling element bearings (REB). In many cases, this separation is considered to play a significant role in the proper diagnosis of bearing faults and it is therefore often employed and mentioned in bearing fault detection research. This chapter discusses this signal separation with a focus on specifically this bearing fault detection. More in particular, it investigates the automation of a cepstrum-based separation procedure.

Separating bearing faults from masking signal content assumes that a bearing fault signal is stochastic due to the random variation on the fundamental period of the fault frequency. This can be attributed to the random slip of the rolling elements. This random slip causes a smearing of the bearing frequencies in the amplitude spectrum (shown in Fig. 3.1), while the deterministic signals manifest themselves as discrete peaks. This property forms the basis for the cepstrum editing methods investigated in this chapter for signal separation. The term 'jitter' used in this chapter is sometimes used instead of slip. However, the random variable used in the signal models is in fact the spacing between the impulses itself, rather than a fixed

jitter on the fundamental bearing fault period. This means the uncertainty increases with projected time into the future. It also means that the bearing fault signals simulated in this chapter are pseudo-cyclostationary, rather than cyclostationary.

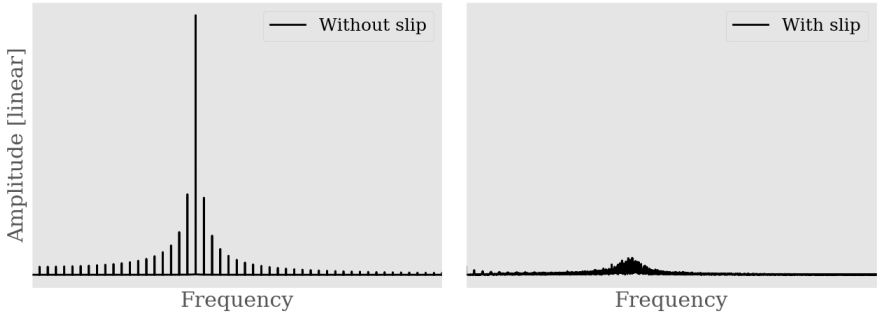

FIGURE 3.1: Amplitude spectrum of a simulated outer race bearing fault, (left) without any slip, (right) with 2% slip .

It is now understood that the real cepstrum can be used to edit the log amplitude spectrum of stationary signals and combined with the original phase to achieve edited time signals. This particular finding has given rise to the development of cepstrum editing methods for the separation of deterministic signal content from stochastic content. Initially, most of the research focused on developing a cepstrum editing procedure to selectively set certain cepstral peaks belonging to masking discrete frequencies to zero [167, 173, 243]. The idea here is mainly to filter out the deterministic frequencies while preserving the rest of the signal's content. Lately however, there has been an increasing usage of a so-called cepstrum pre-whitening method [244–246]. Instead of setting only a selection of peaks to zero, this method sets the whole real cepstrum to zero, except for the zero quefrency. This technique is very easy to implement and has a very low computational cost, but alters the signal content substantially in a rather uncontrolled fashion.

In this chapter a comparison is made between these two developments in cepstral processing of vibration measurements, namely

between an automated cepstrum editing procedure (ACEP) and cepstrum pre-whitening (CPW). While this comparison investigates some important influences on the performance of both methods, it is not an exhaustive review of all possible influencing factors. It assesses the potential benefits and/or disadvantages of each method in order to make future end-users aware of this when choosing between the two methods. This chapter presents an overview of the performance of both methods on virtual and experimental signals.

3.2 Methodology

Before illustrating the algorithms of the two studied methods, a small summary of the theory behind the cepstrum is presented. Next, the automated cepstrum editing procedure and the cepstrum pre-whitening method are explained further in sections 3.2.2 & 3.2.3. To get a better view on the method sensitivities, both methods are first used on simulated signals and investigated for varying signal-to-noise ratios of additive white Gaussian noise, noisy resonances and number of harmonics. An envelope spectrum feature is defined to help in evaluating the performance of the two methods. Finally, the ACEP and CPW techniques are examined on experimental data originating from the *wind turbine gearbox condition monitoring round robin study* organized by the National Renewable Energy Laboratory (NREL) in the US.

3.2.1 Cepstrum: Theory

The complex cepstrum is defined as the inverse Fourier transform of the log spectrum. It is here expressed in terms of the amplitude and the phase of the spectrum:

$$C_c(\tau) = \mathscr{F}^{-1}\{ln(X(f))\} = \mathscr{F}^{-1}\{ln(A(f)) + j\phi(f)\} \tag{3.1}$$

where $X(f)$ is the frequency spectrum of the signal $x(t)$:

$$X(f) = \mathscr{F}\{x(t)\} = A(f)e^{j\phi(f)} \tag{3.2}$$

By setting the phase to zero in Eq.(3.1), the real cepstrum can be obtained:

$$C_r(\tau) = \mathscr{F}^{-1}\{ln(A(f))\} \tag{3.3}$$

Here, τ is a measure of time, referred to as "quefrency", however it is not defined in the same sense as a signal in the time domain. A peak at a certain quefrency corresponds to the inverse period of a series of periodic harmonics in the spectrum. For example, if the sampling rate of a signal is 20 kHz and the cepstrum displays a quefrency peak at 1000 samples, the peak indicates that there is a family of harmonics present in the spectrum with a spacing of 20 Hz (20 kHz/1000 samples).

An important property of the cepstral domain is that the convolution of two time domain signals can be expressed as an addition of their cepstra. Suppose an output signal y(t) of a physical system that is the convolution of an input signal x(t) and an impulse response h(t) of the system:

$$y(t) = x(t) * h(t) \tag{3.4}$$

Because of the convolution theorem, this time domain expression transforms into a multiplication in the frequency domain:

$$Y(f) = X(f)H(f) \tag{3.5}$$

In turn, taking the logarithm of Eq.(3.5) transforms the multiplication into a sum:

$$log(Y(f)) = log(X(f)) + log(H(f)) \tag{3.6}$$

Since the Fourier transform is a linear transform, the addition remains valid in the cepstral domain.

$$C(\tau) = \mathscr{F}^{-1}\{log(Y(f))\} = \mathscr{F}^{-1}\{log(X(f))\} + \mathscr{F}^{-1}\{log(H(f))\} \tag{3.7}$$

This property indicates the possibility to deconvolve a signal if one of the factors is known. As such the logarithmic transformation allows to separate the influence of the excitation source and the transmission path of the system in the cepstral domain.

3.2.2 Automated cepstrum editing procedure (ACEP)

The cepstrum has the interesting property that it is able to concentrate periodic spectral components (e.g. harmonics) into a smaller number of impulses called *"rahmonics"*. The first rahmonic is considered to be the most important one since it represents the average protrusion level of harmonics in the spectrum above the noise level [176]. The higher rahmonics are affected by artefacts, e.g. the used window for the frequency analysis. Removal of these rahmonic peaks, also called *"liftering"*, corresponds to a decrease of the log amplitude of the periodic components in the signal. It was recently shown by Randall & Sawalhi [243] that the real cepstrum is an effective way of editing the log amplitude spectrum of signals containing harmonic components. Recombining this edited amplitude spectrum with the original phase produces the edited time domain signal. This cepstrum editing procedure (CEP) has been further investigated [244, 245, 247] where it was shown that the CEP method is a simple method to implement for separating deterministic signal content from random or cyclostationary signal content. While the editing of certain frequencies introduces some phase distortion, this is generally negligible compared to the significant reduction in amplitude at those frequencies. A general scheme of the CEP method is shown in Fig.3.2a. It can be seen that the most crucial part in the CEP method is the editing of the real cepstrum. To enhance the applicability of the CEP method to industrial environments, an automated editing procedure is used in this research instead of a manual one. It was shown by Ompusunggu [248] that it is possible to implement a robust and efficient automated liftering of the harmonic signal content. Figure 3.2b displays an overview of the steps to automate the editing. It can be seen that the real cepstrum is enhanced before the actual peak selection. The reason for the enhancement is simply to make the peak detection step easier and more straightforward. First, the cepstrum is long-pass liftered as to prevent liftering of low quefrency content. Second, a wavelet denoising and spectral subtraction (SS) method is used to reduce the amount of noise present in the signal. Afterwards, a notch lifter is generated based on the automatic peak detection on the denoised cepstrum. Lastly, the original real cepstrum is liftered with the resulting notch lifter which gives the detected peaks zero amplitude in the cepstrum,

before transforming back to the frequency and time domain. The end result should possess a similar amplitude spectrum as the original signal, yet without the deterministic components.

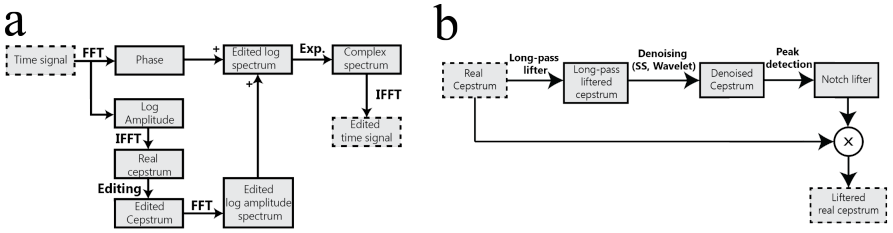

FIGURE 3.2: (a) Schematic diagram of the cepstrum editing procedure. (b) Schematic diagram of the automated cepstrum editing step.

3.2.2.1 Long-pass lifter

Transforming the signal to the cepstral domain leads to a concentration of modal information in the low quefrency region. These system resonances are preferable to retain. Thus, the low quefrency region is not taken into account for the notch lifter generation step, implying that there will be no detected peaks corresponding to modal content.

For a general case it is difficult to determine the appropriate cut-off quefrency. The decision can be based on the resonance frequencies that are present in the investigated system, but this involves manual inspection of the signal. To automate the procedure, it is reasonable to construct a long-pass lifter with a relatively high cut-off quefrency as to have a safety margin and to prevent liftering away too much modal content. A user should take into account the type and size of the machine on which they perform the analysis.

If $n, N_{cut-off}$ and L are respectively the sample quefrency index, the cut-off quefrency index and the sample length of the cepstrum, the long-pass lifter can be defined as follows:

$$l_{LP}(n) = \begin{cases} 0 & n = 1 : N_{cut-off} \\ 1 & n = N_{cut-off} + 1 : L \end{cases} \tag{3.8}$$

Applying the lifter to the unedited real cepstrum $c(n)$ produces the long-pass liftered cepstrum:

$$c_{LP}(n) = c(n)l_{LP}(n). \tag{3.9}$$

3.2.2.2 Noise reduction

After the long-pass liftering step, the cepstrum is denoised using two conventional denoising methods. As proposed by [248], the first applied method is spectral subtraction (SS). To smooth the cepstrum further, wavelet denoising is used afterwards.

Spectral subtraction

The spectral subtraction method was originally introduced by [249] as an acoustic speech enhancement technique and later an ample amount of variations have been developed [250, 251]. In this work the multi-band spectral subtraction proposed by Kamath et al. [252] is used.

The spectral subtraction method estimates the noise spectrum and the average signal spectrum and then subtracts them from each other, improving the average signal-to-noise ratio (SNR). Usually noise is colored and not perfect white Gaussian noise. The multi-band spectral subtraction method allows to some degree the presence of colored noise because the filtering of the signal spectrum is separated into different frequency bands.

Being a transformation of the measured time signal, the cepstrum contains noise as well and can be regarded as an addition of discrete signal components $\tilde{c}(n)$ and noise $d(n)$:

$$c_{LP}(n) = \tilde{c}(n) + d(n). \tag{3.10}$$

A diagram of the multi-band spectral subtraction technique is shown in Fig.3.3. The SS technique makes use of a window basis for calculating the Fourier transform and performing the denoising procedure. The amplitude $|C(f,k)|$ is employed for estimating the average noise spectrum in every frequency band iteratively. The estimated amplitude $|\tilde{C}(f,k)|$ is obtained by subtracting the noise estimate from the original amplitude $|C(f,k)|$. Finally, the de-noised cepstrum $\tilde{c}(n)$ is

reconstructed using the original phase $\angle C(f,k)$ and the amplitude estimates of all window frames.

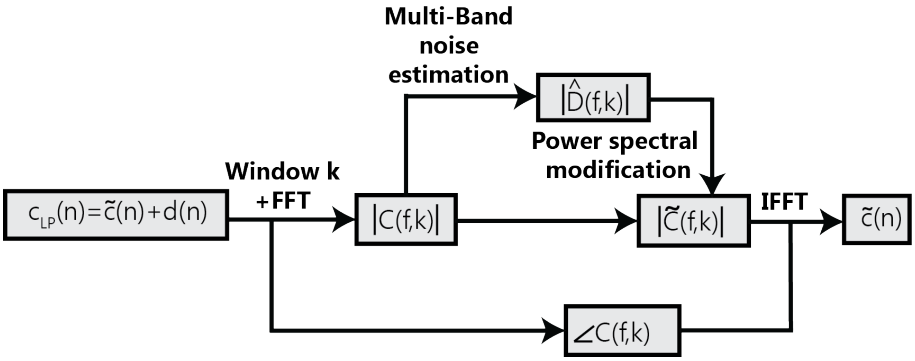

FIGURE 3.3: Schematic diagram of the multi-band spectral subtraction method.

Wavelet denoising

One of the primary applications of wavelets is denoising of signals. Wavelets allows for denoising in both time and frequency domain simultaneously and has even been used before as standalone method for bearing fault detection [176]. Wavelet denoising employs thresholding methods where the wavelet coefficients are thresholded to remove the noisy signal content and is mostly based on research conducted by [253]. A 'hard thresholding' method leaves the retained coefficients unchanged, while a 'soft thresholding' method, as is used for the cepstrum, subtracts the noise estimate (threshold value) from the retained coefficients. An example of wavelet denoising is presented in Fig. 3.4.

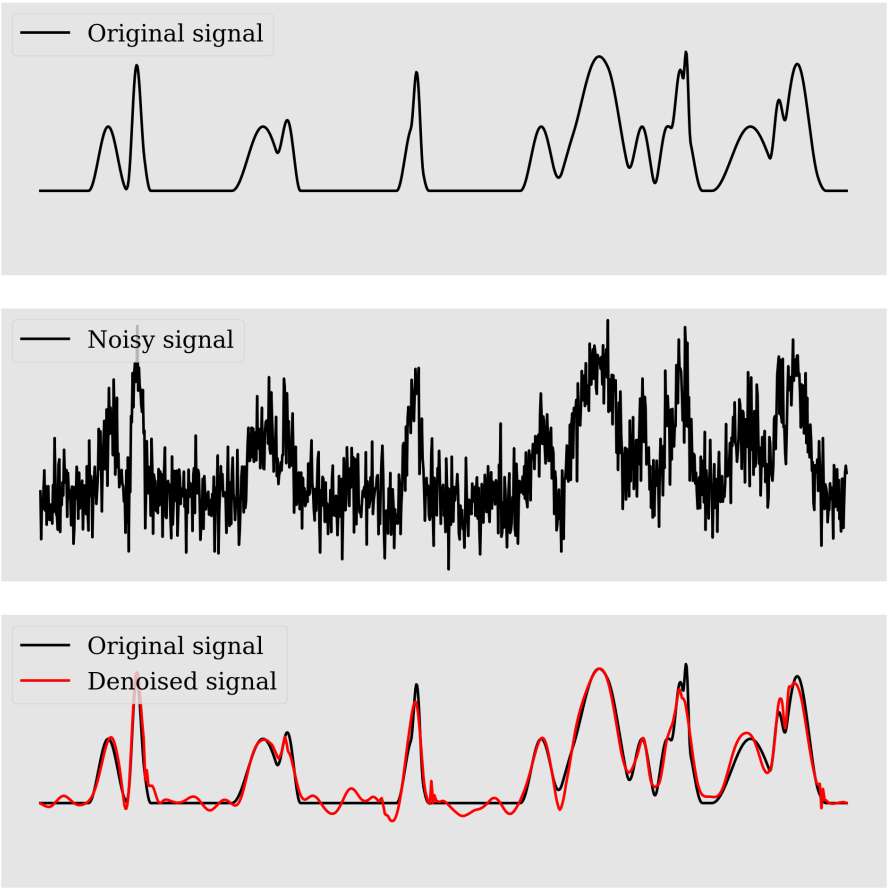

FIGURE 3.4: (Top) Original signal, (Middle) Signal with additive white Gaussian noise at an SNR of 3.45 dB, (Bottom) Denoised signal with an improved SNR of 13.44 dB.

The wavelet denoising method employed for the CEP method makes use of Daubechies3 wavelets and of universal soft thresholding. This chapter does not go into further detail about wavelet denoising however since it is not the primary focus of this dissertation. For more information on wavelet denoising, the interested reader is referred to Refs [176, 254, 255].

3.2.2.3 Peak detection

After long-pass liftering and denoising, a fixed threshold of three standard deviations (3σ) is calculated from the residual cepstrum $c_r(n)$ as follows:

$$threshold = E[c_r(n)] + 3 \times std[c_r(n)], \tag{3.11}$$

where $E[.]$ and $std[.]$ denote respectively the expectation operator and the standard deviation operator. A vector m is constructed, based on all values greater than the threshold, containing all the corresponding sample indices:

$$m = \{\forall n | c_r(n) > threshold\} \tag{3.12}$$

While periodic signal components show up as sharp peaks in the cepstrum, second-order cyclostationary components like bearing faults normally do not show up as strong peaks in the real cepstrum and thus they should not be detected nor liftered away.

3.2.2.4 Liftering

The final editing step is the generation of the notch lifter, based on the cepstral peaks of the denoised cepstrum, and the actual liftering of the unaltered real cepstrum with this lifter. It should be noted however that the notch width has a significant influence on the performance of the ACEP method. If there is speed variation present in the signal, it can be desirable to choose a larger notch width, but it should not be chosen too large either since this can lead to excessively distorting the amplitude spectrum.

3.2.3 Cepstrum pre-whitening

Another approach of cepstral editing, usually called cepstrum prewhitening, is far more radical in liftering the real cepstrum [167]. Instead of liftering only the detected cepstral peaks and setting a fairly limited amount of samples to zero, this technique sets the whole real cepstrum to zero, except for the zero quefrency. Both discrete frequencies and resonances are removed from the amplitude spectrum using this approach. The zeroed cepstrum is then transformed back to the

time domain, after recombination with the wrapped phase of the original signal. This method is actually easily implemented by dividing the complex frequency domain signal by its absolute value and after inverse transforming it back to the time domain:

$$x_{CPW}(t) = \mathscr{F}^{-1}\{\frac{\mathscr{F}(x(t))}{|\mathscr{F}(x(t))|}\} \tag{3.13}$$

This technique "whitens" the signal's spectrum by decreasing the amount of variation between adjacent frequency bins. As described by Blackman & Tukey [256], this introduces "compensation" to the spectrum of the signal and the flattening does not need to be that precise either. Essentially it decreases the rate of change of power spectral density with frequency, giving the same weighting to all frequency bands, in the hope of emphasizing impulsive events. Figure 3.5 shows a diagram of the straightforward processing procedure using the CPW method.

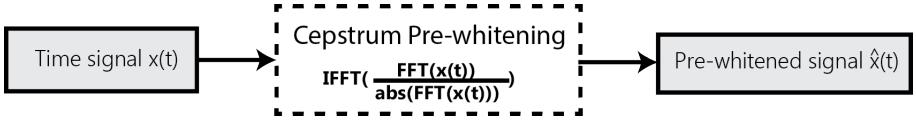

FIGURE 3.5: Schematic diagram of the cepstrum pre-whitening procedure.

3.3 Comparison on simulated signals

In order to make a qualitative comparison of the two mentioned techniques and to get a better knowledge of the influencing factors, the results of applying the methods to virtual signals are first assessed. Vibration measurements on complex machinery usually contain a mixture of signals originating from a plethora of sources. All these sources influence the measurements in their own way. As such, this chapter does not cover all the possible influencing factors, rather it investigates the sensitivity and performance of the two methods to a chosen set of affecting elements.

3.3.1 Squared envelope spectrum feature

Assessing the performance of two methods raises the need for a performance metric or indicator, enabling an easy comparison between the techniques under investigation. Since one of the most used techniques for bearing fault diagnosis is the *squared envelope spectrum* (SES), a feature is defined on the normalized squared envelope spectrum which gives a measure of the discernibility of the bearing fault frequencies. This feature is defined as the average amplitude of the first five harmonics of the fault frequency f_{fault} in the normalized envelope spectrum subtracted with a threshold that is based on the noise floor level in the envelope spectrum. Mathematically, this is simply expressed as follows:

$$feature = \frac{1}{5}\sum_{i=1}^{5}(A(f_{fault}i) - A_{threshold}) \tag{3.14}$$

Note also that the envelope is calculated over the full spectral bandwidth of the signal and no additional filtering takes place.

3.3.2 Influence of additive white Gaussian noise

The first factor that is examined is the influence of noise. While noise is always present in experimental measurements, it does depend on the application and the measurement equipment whether the noise levels will play a deciding role in the fault detection or not.
For this case, the simulated signal consists solely out of an outer race bearing fault signal, additive white Gaussian noise and harmonics. The model used for the bearing signal assumes that there is only a single fault present and that it is rotating at a constant speed. The bearing fault signal $x(t)$ can then be expressed as the convolution of an impulse train with a transfer path impulse response function $h(t)$. This gives:

$$x(t) = h(t) * \sum_{i=1}^{N} A_i\delta(t - iT_{fault}) \tag{3.15}$$

The fundamental fault period T_{fault} is however not a constant, it varies slightly in a random manner due to the bearing slip [161]. Next to this

random slip on the fundamental period, there is also a random amplitude modulation of impulses A_i, and the impulse responses $h(t)$ take into account system resonances excited by the bearing fault with a corresponding damping ratio.
The simulated bearing signal used in this chapter is synthesized as a single-degree-of-freedom (SDOF) system with a resonance frequency of 2000 Hz and a damping ratio of 5%. The period of the Dirac impulses corresponds to a bearing fault frequency of **32 Hz** with 5% random slip and 10% random amplitude modulation. The resulting bearing signal in the time domain, sampled at 10 kHz, is shown in Fig. 3.6a and in the frequency domain in Fig. 3.6b.

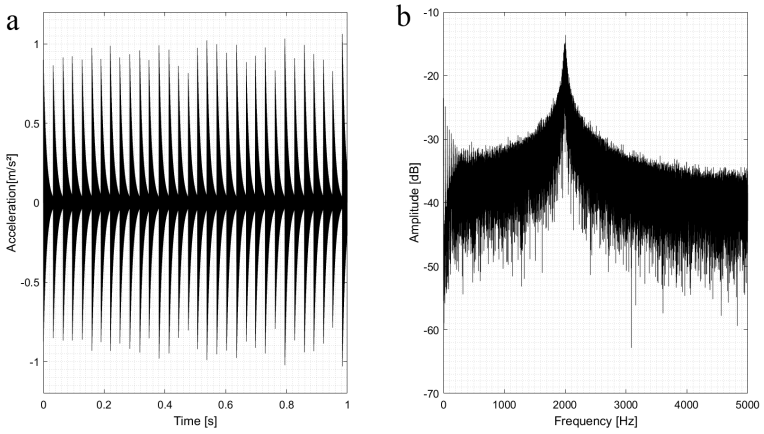

FIGURE 3.6: (a) Time domain representation of the simulated bearing signal. (b) Amplitude spectrum of the bearing fault signal

Since the CPW and ACEP method are designed to remove the influence of deterministic signal content in the envelope spectrum, a set of harmonics is added to the bearing signal. In total three harmonic families are added with fundamental frequencies of 22 Hz, 37 Hz and 50 Hz.

Finally, white noise is added to the bearing signal with a predetermined signal-to-noise ratio (SNR), with the bearing fault being the signal of interest and the white Gaussian noise the noise. The SNR is

then varied from -50 dB to +50 dB and the envelope feature is calculated for every SNR value. This enables the possibility of discerning the differences between CPW and ACEP. Figure 3.7 shows the results of this simulation for the unedited original signal, consisting of the bearing signal, harmonics and additive noise, and for the same signal after applying the CPW and ACEP methods.

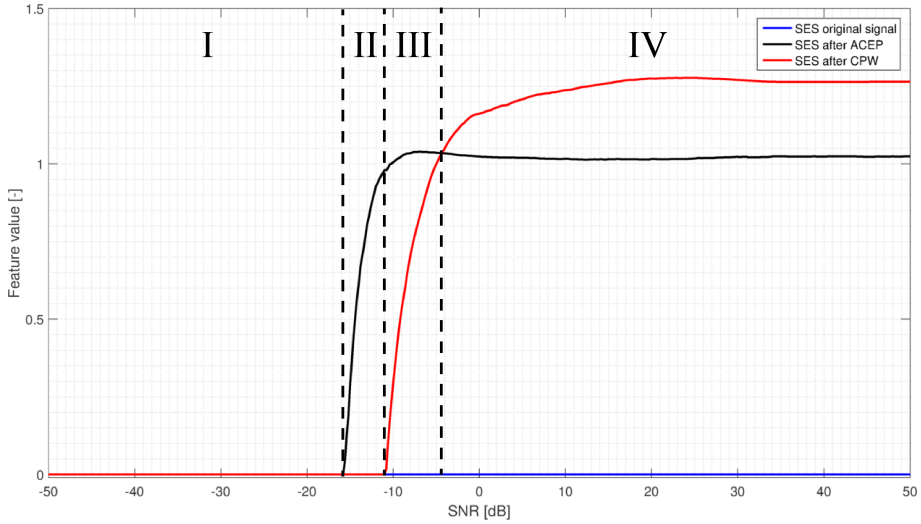

FIGURE 3.7: Evolution of SES feature value in function of the signal-to-noise ratio of the bearing fault signal to the white Gaussian noise.

It can be seen that for SNR values below -16 dB the ACEP method is unable to detect any presence of the bearing fault frequency in the squared envelope spectrum. However, this is still an improvement compared to the lower detection limit of -11 dB for the CPW method. The graph indicates that there are four regions discernible:

I. Below -16 dB SNR, where neither the CPW or the ACEP method are able to detect the bearing fault.

II. Between -16 dB and -11 dB SNR, where only ACEP detects the bearing fault.

III. Between -11 dB SNR and -4 dB SNR, where the feature value for the CPW method is increasing but remains smaller than the ACEP method.

IV. Above -4 dB SNR, where the feature value for the CPW method crosses the ACEP method and converges towards a higher value than the ACEP method.

The first region is not of interest since it only shows the presence of noise in the envelope spectra. Therefore the envelope spectra corresponding to an SNR value in each of the last three regions are further investigated. Three signals are generated with an SNR of -15 dB, -10 dB and +5 dB, each corresponding to one of the three regions. In Fig. 3.8a the amplitude spectrum is shown of the unaltered signal for the -15dB SNR case. The resonance around 2000Hz can barely be identified here. Figure 3.8b shows the spectrum of the signal after the ACEP method (*black*) and the CPW method (*red*). Unsurprisingly, the spectrum is just a horizontal line after CPW, since this is the only manipulation this method administers. Figure 3.9 shows the squared envelope spectra for the original signal, the signal after applying the ACEP method, and the signal after CPW. A visual inspection of these three envelope spectra corroborates the simulation results of Fig. 3.7. Only the envelope spectrum after ACEP is able to distinguish the 32 Hz BPFO component properly. Higher harmonics cannot be seen, but the spectrum after CPW does not exhibit any clear peak at the BPFO frequency. This is not really surprising since the main benefiting factor for the detection of the bearing fault signal is the resonance around 2 kHz. The CPW method removes this resonance and thus nullifies part of the bearing information protruding above the noise floor. It relies solely on the residual phase information, which alone is too noisy to give rise to a fault frequency peak in the envelope spectrum. The envelope spectrum of the original signal is, as expected, totally dominated by the strong deterministic components and does not show any presence of the bearing fault frequency. The ACEP method is able to remove these harmonics while still preserving the bearing resonance. This preservation helps in the detection of the BPFO component as seen on the figure.

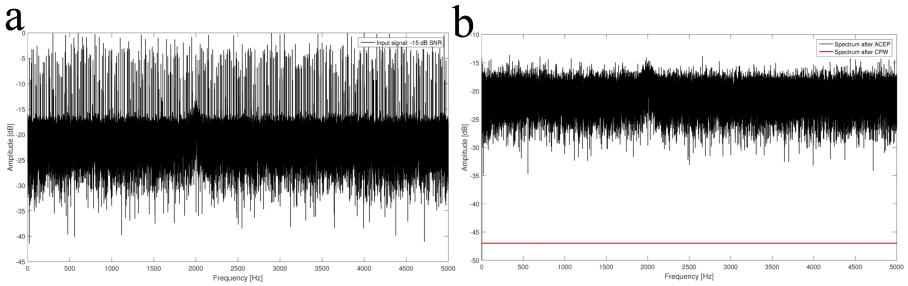

FIGURE 3.8: (a) Amplitude spectrum of the simulated signal for the -15dB SNR case. (b) Amplitude spectrum after the ACEP and CPW method.

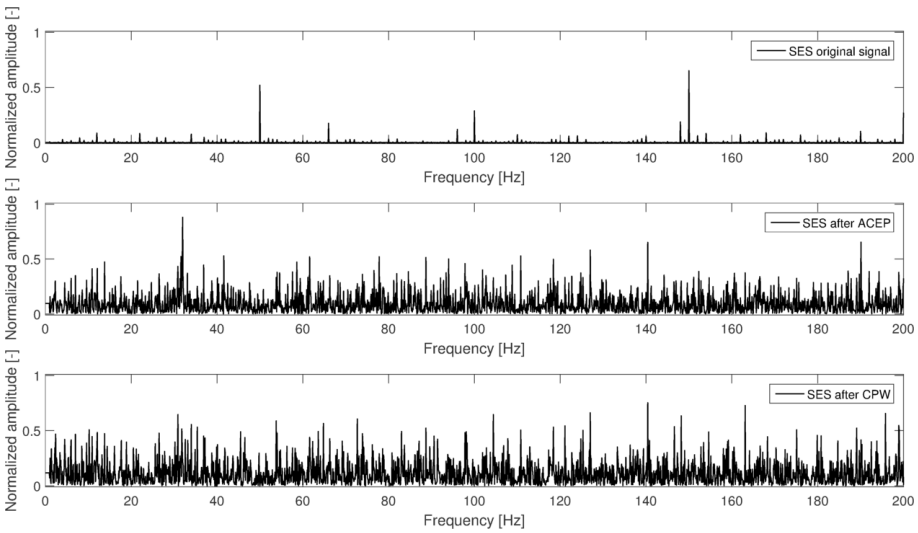

FIGURE 3.9: Squared envelope spectra of the original signal, the signal after ACEP, and the signal after CPW for the -15 dB SNR case.

The third region is assessed by investigating a signal with an SNR of -10 dB. The SES feature indicates that both methods should already detect the bearing frequency quite well, but the ACEP method should still outperform the CPW method. Figure 3.10 confirms this suspicion.

The ACEP envelope spectrum gives a clear view of the 32 Hz fault frequency and its second harmonic, while only the fundamental fault frequency is discernible in the CPW envelope spectrum.

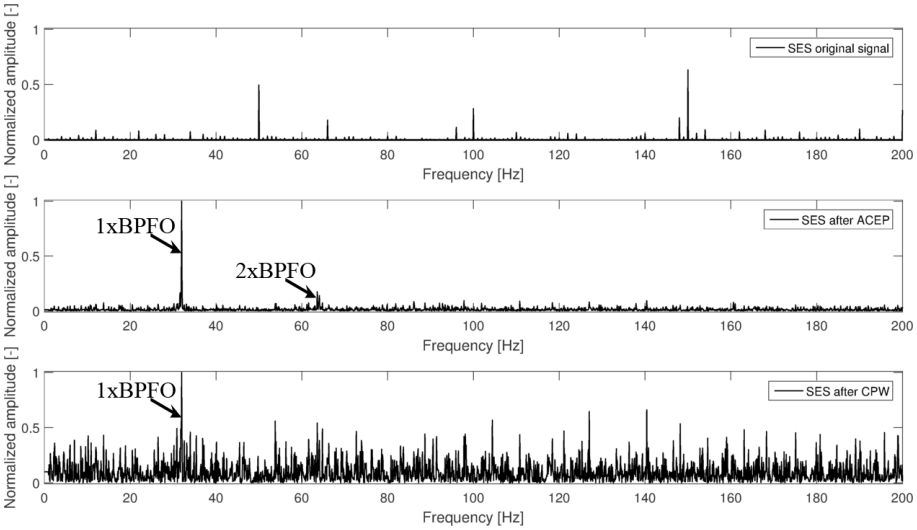

FIGURE 3.10: Squared envelope spectra of the original signal, the signal after ACEP, and the signal after CPW for the -10 dB SNR case.

The fourth and last region indicates that the CPW envelope spectrum should display a stronger presence of the fault frequency than the ACEP spectrum. Again, this belief is validated as can be seen in Fig. 3.11. The amplitude of the higher harmonics of the bearing fault frequency are more easily distinguishable after the CPW technique than after the ACEP method. This can be explained by the fact that the CPW method removes most of the resonance's influence in the time signal, making the reconstructed time signal essentially a somewhat noisy Dirac comb with a period between the peaks equal to the bearing fault period. The envelope spectrum of a Dirac comb is again a Dirac comb with a frequency spacing at the bearing fault frequency. This effect thus amplifies the detection possibilities compared to the ACEP method since it does retain the bearing resonance, causing a smearing effect of the envelope time signal and reducing the amplitude of the higher harmonics in the envelope spectrum.

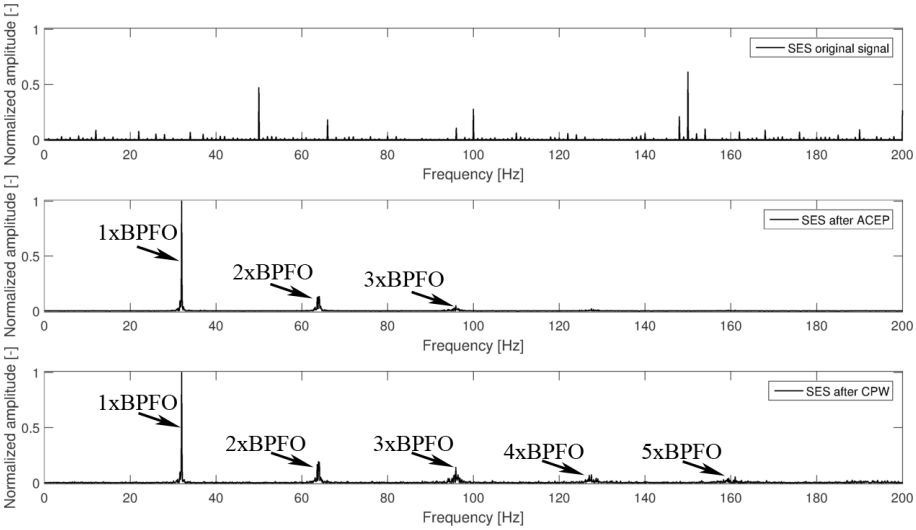

FIGURE 3.11: Squared envelope spectra of the original signal, the signal after ACEP, and the signal after CPW for the +5 dB SNR case.

In general, the analysis exemplifies that for low signal-to-noise ratios, i.e. strong additive white Gaussian noise compared to the bearing signal, the ACEP method can outperform the CPW method due to its capabilities of retaining more of the bearing fault information. While for high signal-to-noise ratios the CPW method can outperform the ACEP method due to the removal of resonance frequencies, emphasizing the bearing fault impulses.

3.3.3 Influence of non-fault related resonances

In the previous section it was already mentioned that the presence of system resonances has an influence on the detection capabilities of the studied techniques. Hence, this influence is further examined through a similar analysis as for the additive white noise case.

A virtual signal is constructed consisting of harmonic components, the same bearing fault signal as in the previous section, and a system resonance around 1 kHz which is not related to the fault in any way.

This resonance is constructed in the frequency domain using the following transfer function:

$$H(s) = K\frac{s}{(s-p_1)(s-\overline{p}_1)} \tag{3.16}$$

with $s = j\omega$ the complex frequency, K the transfer gain, $p_1 = -10 + 1000i$, and $\overline{p}_1$ its conjugate pole. This results in a resonance around 1 kHz with a damping ratio of 1%. The output acceleration signal $A(s)$ is the result of using complex white Gaussian noise as force input for the transfer function:

$$A(s) = H(s)F(s). \tag{3.17}$$

This frequency domain signal is then simply transformed back to the time domain using an inverse discrete Fourier transform.

Again, the evolution of the SES feature value is tracked for increasing SNR values, going from -50 to +50 dB SNR, with the 1 kHz resonance signal being the noise this time. Figure 3.12 displays the simulated results, which look similar to the ones of the previous section. However, now it is the CPW method that succeeds in detecting the bearing frequency for lower SNR values.

In order to reduce the size of the represented results, only the envelope spectra for an SNR value in the second and third region are examined for validating the simulation results. An SNR value of -20 dB SNR, corresponding to the second region, is inspected further, together with an SNR value of -10 dB SNR for the third region.

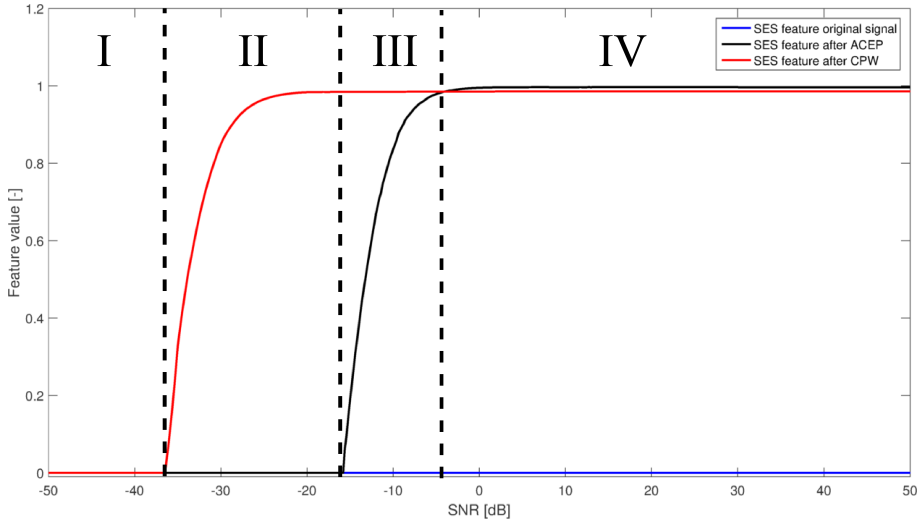

FIGURE 3.12: Evolution of SES feature value in function of the signal-to-noise ratio of the bearing fault signal to the 1 kHz resonance signal.

Figure 3.13a shows the spectrum for the -20 dB SNR case before cepstral editing, and Fig. 3.13b after applying the ACEP and CPW methods. Inspection of the envelope spectra in Fig. 3.14 ratifies the evolution analysis results. The CPW envelope spectrum clearly exhibits the bearing fault harmonics, while the ACEP spectrum does not show any significant peaks near the BPFO at all and is mainly dominated by noise. Since the envelope spectrum is calculated for the full bandwidth of the signal, this is to be expected. The spectrum after ACEP is for the most part still overshadowed by the 1 kHz resonance with noise as force input due to it higher amplitude. band-pass filtering around the 2 kHz bearing resonance would alleviate this problem, but this is not the research intent here specifically. The CPW method circumvents the influence of the resonance by equally weighting the amplitudes of all the frequencies. Therefore it enhances the detection of the bearing impulses much earlier than the ACEP method by relatively reducing the 1 kHz resonance amplitude and increasing the 2 kHz one.

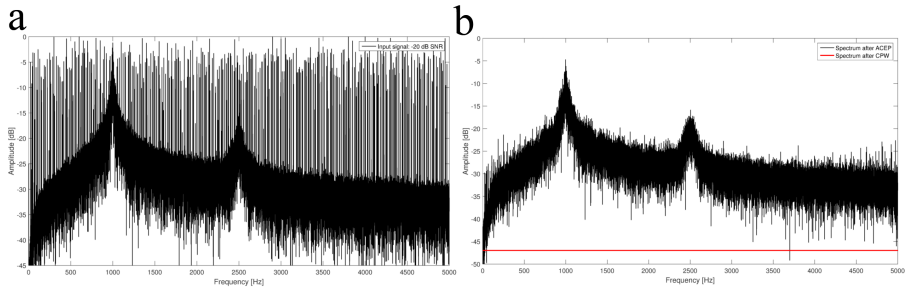

FIGURE 3.13: (a) Amplitude spectrum of the simulated signal for the -20 dB SNR 1 kHz resonance case. (b) Amplitude spectrum after the ACEP and CPW method.

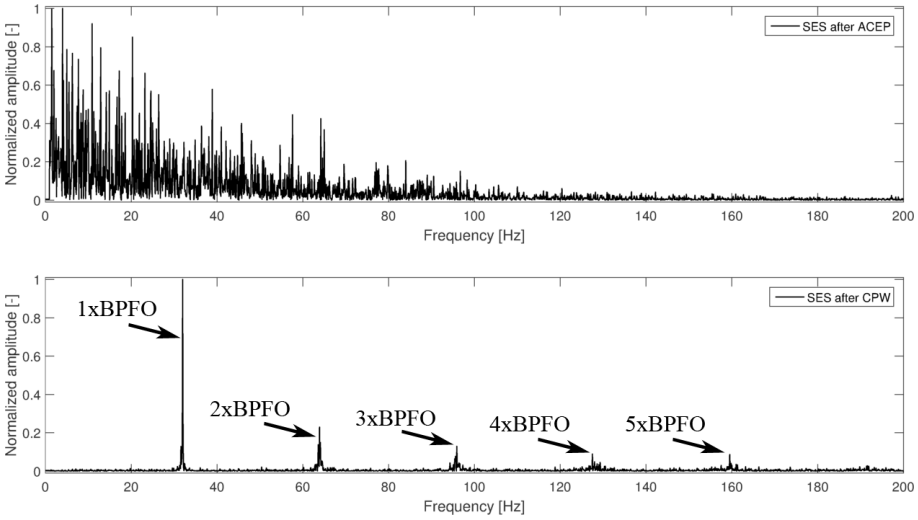

FIGURE 3.14: Envelope spectra for -20 dB SNR case with a 1 kHz resonance for the simulated signal after ACEP and CPW.

The second analysis is done for -10 dB SNR, which is located in the third region. Fig. 3.15 reveals that the ACEP method indeed starts to detect the bearing frequency for higher SNR values. In general it can be concluded that the influence of resonances that carry no bearing information, is quite significant for full bandwidth enveloping and since

practical setups are usually multiple-degree-of-freedom (MDOF) systems, it is very likely that the bearing information is clouded by these resonances, making the CPW method a more interesting candidate for such full bandwidth analyses. This analysis was also performed for multiple resonances at different frequencies, but all gave similar results to the ones presented here.

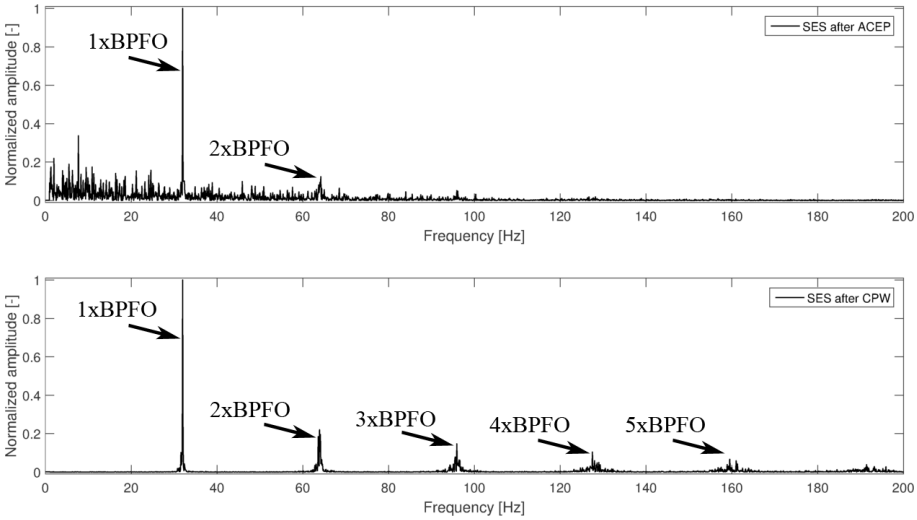

FIGURE 3.15: Envelope spectra for -10 dB SNR case with a 1 kHz resonance for the simulated signal after ACEP and CPW.

3.3.4 Influence of harmonics

In order to prove the possible equal performance of both described methods in filtering out harmonic content, a quick evolution analysis is done by varying the amount of harmonics. This amount influences the amplitude of the corresponding peak in the cepstral domain and therefore also the effectiveness of the ACEP method in detecting these peaks. However, it is found that with the right parameters the ACEP method displays similar performance compared to the CPW method. Figure 3.16 shows that the feature value remains constant for both the ACEP and CPW method ,while the feature value decreases significantly for an increasing number of harmonics without cepstral editing. The

signal used for this analysis consisted of a bearing fault signal, additive white Gaussian noise, and harmonics of 10 Hz. Other harmonic frequencies and multiple non-harmonic deterministic components were investigated as well and gave analogous results, although the ACEP method does need more fine-tuning compared to the CPW method.

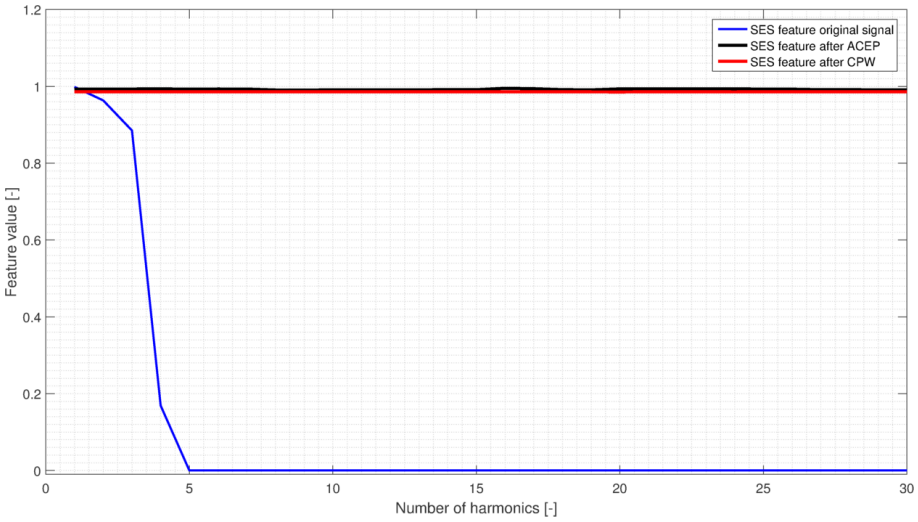

FIGURE 3.16: Evolution of SES feature value in function of the number of added harmonics for the three simulated signals.

3.3.5 Influence of additional parameters

As can be deduced from the previous sections, it would be interesting to do a more in-depth analysis of some additional key parameters of the method and of the simulated signals. As such it was decided to investigate the influences of the following extra parameters:

1. The height of the threshold used in calculating the feature indicator
2. The number of peaks used in calculating the feature indicator
3. The amount of slip on the fundamental bearing fault period
4. The amount of random amplitude modulation of the fault impact peaks

5. The damping ratio of the noise resonances
6. The damping ratio of the bearing fault resonance

The first two parameters are to showcase the sensitivity of the feature indicator to the calculation procedure, while the other parameters are independent design parameters of the simulated signals. These independent parameters are examined simultaneously later on in this section using a multivariable analysis. First however curves are identified for every parameter separately to get a feeling for the individual contributions to the only dependent variable here, namely the bearing fault indicator.

3.3.5.1 One-dimensional analysis of parameters

Indicator threshold The first parameter to be analyzed is the threshold used in the calculation of the envelope spectrum metric. This threshold plays a crucial role in separating noise peaks from real envelope peaks belonging to the bearing fault frequency. The threshold used in section 3.3 is simply four times the moving standard deviation of the envelope spectrum using a central windowing scheme and a window size equal to the sample rate. These values were arbitrarily chosen based on observations and should be adapted if used for other purposes, but for this analysis case they performed well enough and presented a straightforward thresholding approach. It should be noted though that the values of an envelope spectrum typically are not normally distributed (a chi-squared distribution would be more appropriate [257, 258]), but the use of standard deviation is purely from a practical point of view for defining a threshold. To illustrate the differences when a lower or higher value of standard deviation is chosen, Fig. 3.17 illustrates the curves for 1 to 7 standard deviations ("sigma") for the case where the amount of noise is varied from -50 dB to +50 dB SNR. The figures show that, as expected, increasing or decreasing the threshold mainly constitutes a corresponding shift and scaling of the indicator curve. The main concern here is avoiding to take a too low or too high value of standard deviation, since this could respectively lead to noise peaks protruding above the threshold or negating of the

fault peaks.

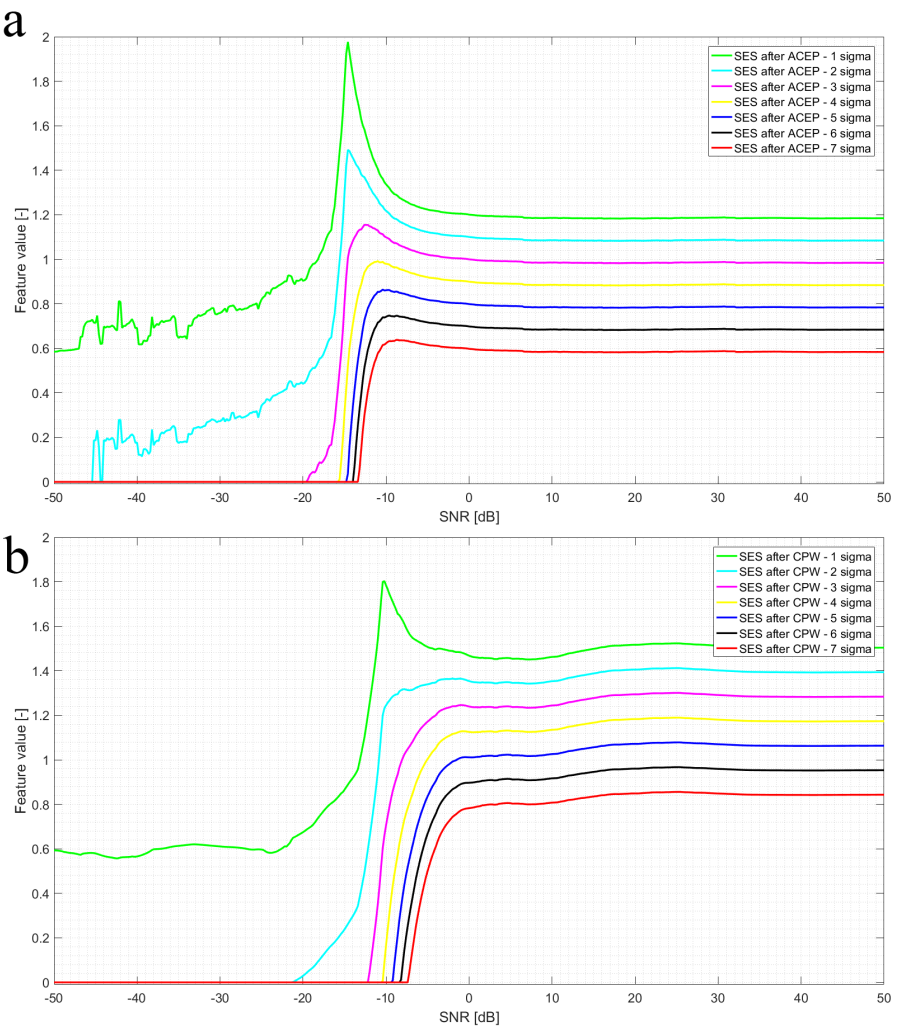

FIGURE 3.17: a) Feature evolution of the SES after ACEP for the case where the threshold is varied from 1 to 7 sigma. b) Feature evolution of the SES after CPW for the case where the threshold is varied from 1 to 7 sigma.

Number of peaks The second parameter, namely the number of peaks considered in the indicator calculation, does not influence the shape of the curves considerably. Mainly a scaling of the curves will occur, as can be seen in Fig.3.18 for the same case as the previous figures. This can be explained by the fact that primarily the first few peaks are high in amplitude and the sum of the amplitudes is simply divided by the total amount of considered peaks. This could be offset by introducing weighting factors for the higher order harmonics since they still protrude noticeably above the surrounding noise. However this was not investigated further.

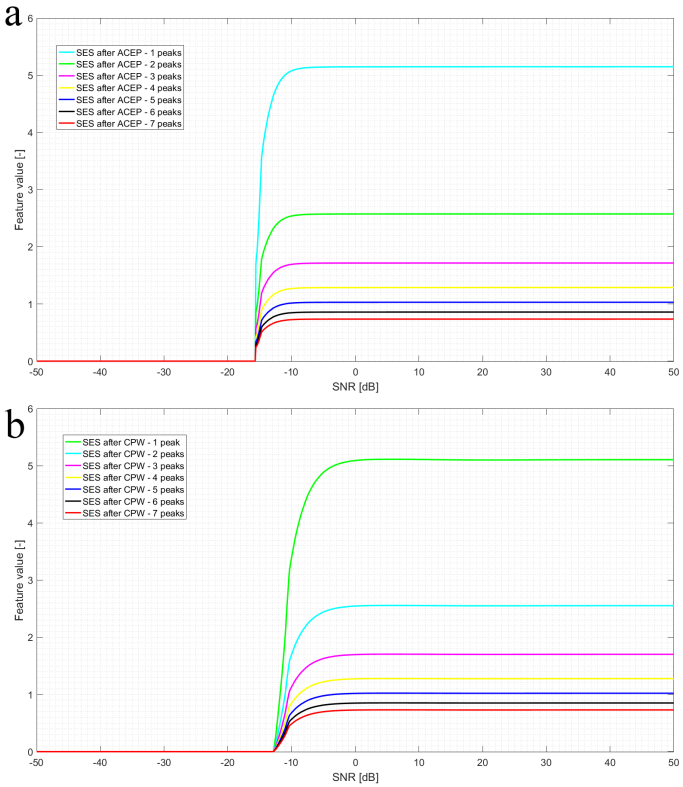

FIGURE 3.18: a) Feature evolution of the SES after ACEP for the case where the amount of peaks is varied from 1 to 7 peaks. b) Feature evolution of the SES after CPW for the case where the amount of peaks is varied from 1 to 7 peaks.

Slip The amount of slip on the fundamental fault period is the next parameter to be inspected. It is expected that the curves will start to lose their normal shape for an increasing amount of random jitter since this smears the envelope spectrum fault peaks and thus decreases the amplitude of the detected peaks. Figure 3.19 shows the result for letting the amount of the jitter vary from 1% to 10% of the fault period.

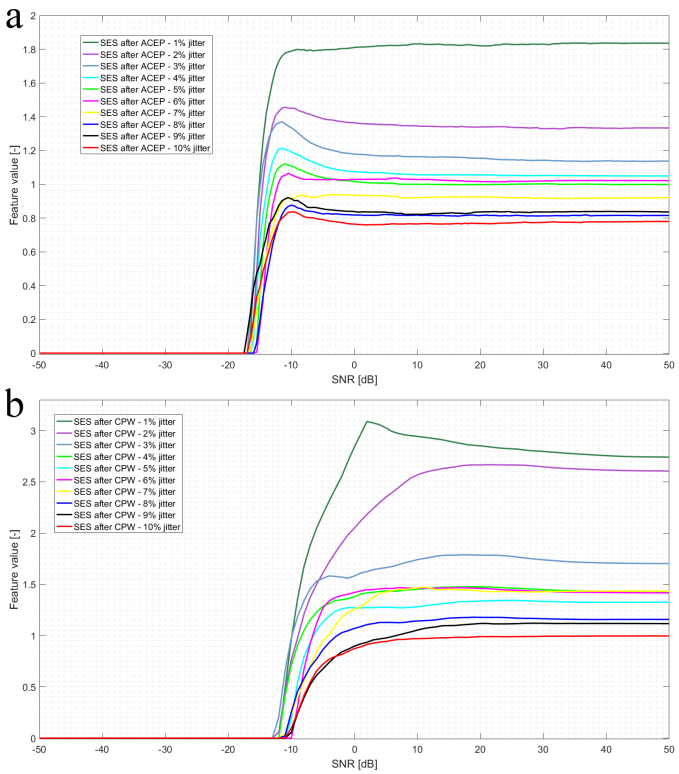

FIGURE 3.19: a) Feature evolution of the SES after ACEP for the case where the amount of jitter is varied from 1 to 10%. b)Feature evolution of the SES after CPW for the case where the amount of jitter is varied from 1 to 10%.

Amplitude modulation To analyze the influence of the amount of random amplitude modulation on the curves, the modulation is varied from 0% to 100%. It can be seen in Fig.3.20 that the random amplitude

modulation has very little effect on the indicator curves. Again, this is to be expected because on average the amplitude modulation doesn't influence the occurrence of the impulses or the spacing.

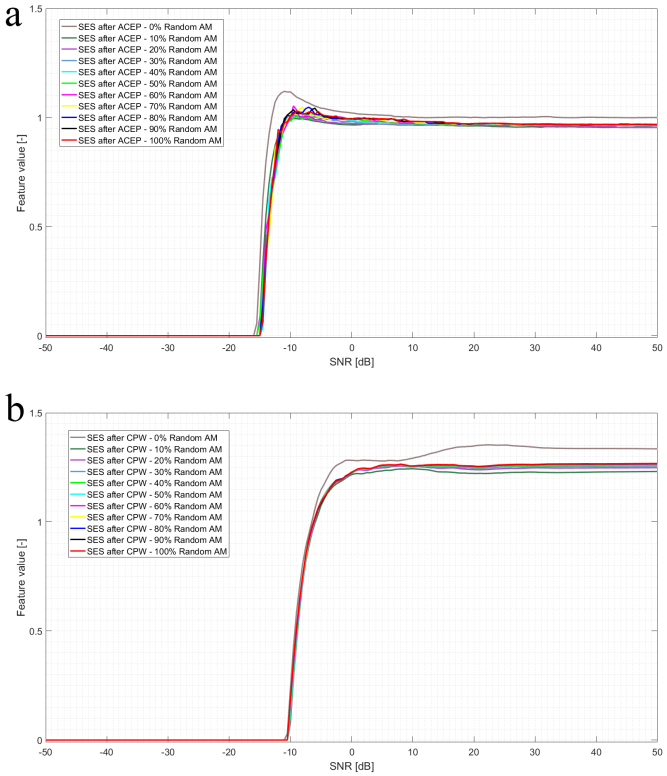

FIGURE 3.20: a) Feature evolution of the SES after ACEP for the case where the amount of random amplitude modulation is varied from 1 to 100%. b) Feature evolution of the SES after CPW for the case where the amount of random amplitude modulation is varied from 1 to 100%.

Damping ratios The last parameter to be investigated is the damping ratio of the resonances. Since there is a damping ratio for the bearing resonance and for the noise resonances, these two cases are investigated separately.
First the influence of the damping ratio is studied for the case where

the SNR is varied between the bearing fault and the 1 kHz noise resonance. It is shown in Fig.3.21 that while the damping ratio does not influence the CPW method significantly, the ACEP method is more sensitive to it. This is to be expected since the CPW method totally removes any amplitude influences of the resonances, while the ACEP method retains this information. As such, lower damping ratios and thus higher resonance amplitude peaks have a more substantial effect on the outcome of the ACEP method than the CPW method.

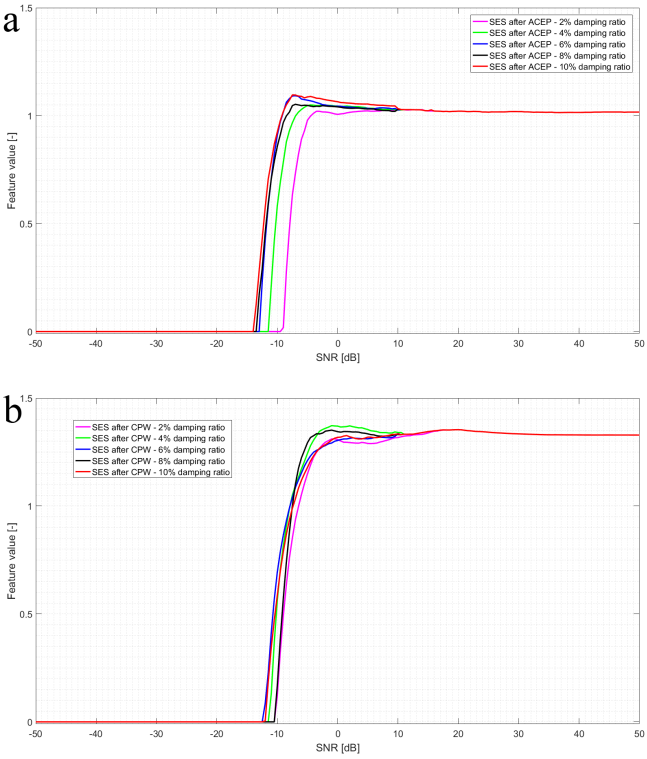

FIGURE 3.21: a) Feature evolution of the SES after ACEP for the case where the damping ratio of the noise resonance of 1 kHz is varied from 2 to 10%. b) Feature evolution of the SES after CPW for the case where the damping ratio of the noise resonance of 1 kHz is varied from 2 to 10%.

The final analysis is of the influence of the damping ratio of the bearing fault resonance on the curves. Figure 3.22 indicates that the damping ratio influences the ACEP method slightly less than the CPW

method. The low damping ratio makes the impulse peaks of the fault impacts less pronounced and causes the bearing fault signal to appear more smooth. This creates an envelope signal that exhibits the fault less clearly than for a high damping ratio. As can be seen, in the presence of low SNR values, the CPW method is more affected by this effect than the ACEP method.

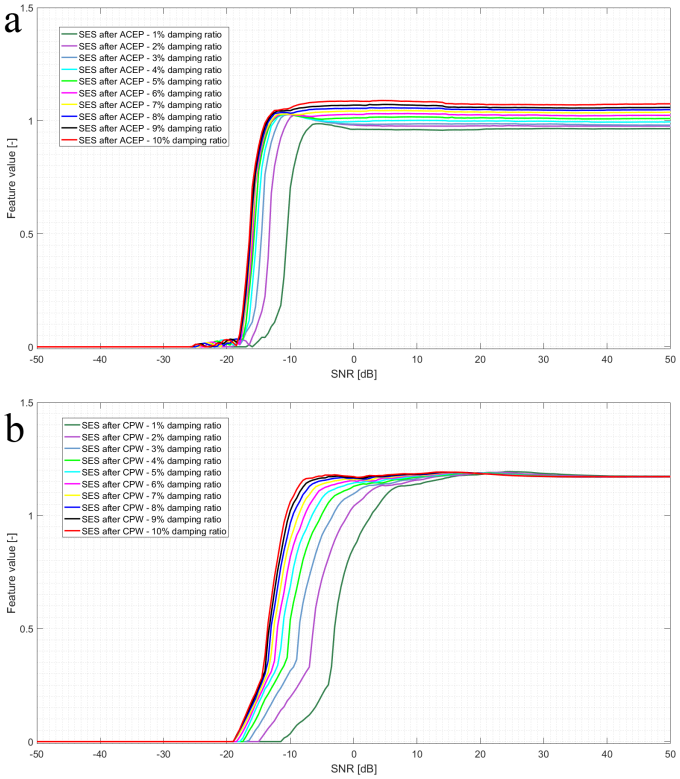

FIGURE 3.22: a) Feature evolution of the SES after ACEP for the case where the damping ratio of the bearing fault resonance of 2.5 kHz is varied from 1 to 10%. b) Feature evolution of the SES after CPW for the case where the damping ratio of the bearing fault resonance of 2.5 kHz is varied from 1 to 10%.

3.3.5.2 Modeling the feature indicator response

Based on the above findings, a multiple non-linear regression approach was chosen to fit a model to the results that resembles to the specific shape of the curves seen above. The chosen model is based on a Weibull distribution, more in particular the cumulative distribution function of a Weibull distribution. The shape of this function resembles strongly to the shape found in the feature indicator response and the Weibull distribution is a commonly used distribution in other research fields. Equation 3.18 presents the formula on which the model is based.

$$CDF_{Weibull} = \begin{cases} 1 - exp(-(x/\lambda)^k) & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3.18}$$

To take into account the non-linear influences that every parameter has on the feature indicator, this model was augmented by inspection of three dimensional surface plots representing the response variable as a function of two of its independent variables. It should be noted that for this model only three parameters are taken into account in order to not make the model excessively complicated. The chosen parameters are:

- Signal-to-noise ratio between bearing fault and additive white Gaussian noise
- The amount random jitter on the fundamental bearing fault period
- The amount of damping of the bearing fault resonance

As can be seen above, these are design parameters of the bearing fault and not method parameters of the feature indicator calculation. The following ranges were used for simulation and calculating the response feature indicator:

- SNR, from -30 dB to +30 dB, in steps of 0.5 dB
- Jitter, from 0.2% to 10%, in steps of 0.2%
- The amount of damping, from 0.2% to 10%, in steps of 0.2%

Based on the inspection of the three-dimensional response curves, the following model was made:

$$y = (\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \frac{\beta_4}{x_2^{\beta_5}}) \exp(-(\frac{10^{\frac{x_1+\beta_6}{10}}}{\frac{\beta_7}{x_3^{\beta_8}} + \beta_9})^{\beta_{10}}) \tag{3.19}$$

In the model x_1, x_2, x_3 and y stand for SNR, jitter, damping and the feature indicator respectively. This model was derived using the following observations:

- For high SNR values, the feature indicator converges towards a certain value. This value depends linearly on the damping and jitter, and is also reciprocal with respect to the jitter. This is in correspondence with the previously obtained results, that indicated a scaling of the indicator curves based on these parameters. The factor before the exponential in Eq.3.18 determines this convergence value and can thus be adapted to reflect this dependence. This explains the presence of coefficients $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5.
- To offset the negative SNR values (in dB), a coefficient β_6 is added.
- The point where the slope of the CDF is highest (approximately in the middle of the S-curve), corresponding to λ in Eq.3.18, is inversely dependent on the damping. This constitutes the need for coefficients β_7, β_8 and β_9. This relationship explains the fact that for decreasing damping values of the bearing fault resonance, it gets increasingly more difficult to detect the bearing fault.
- Lastly the shape parameter, k in Eq.3.18, is described by coefficient β_{10}.

In order to find the appropriate coefficients, the model is computed using MATLAB©. The convergence criterion on the coefficient estimates is $1e{-}8$ and the Levenberg$-$Marquardt algorithm is used to solve the non-linear least squares problem. The initial coefficient values can be found in Table 4.1 along with the estimated coefficients for both the ACEP and CPW method. It should be noted that the model

also converged to similar values of the coefficients for different starting values.

TABLE 3.1: Overview of the starting values and estimated values for the model coefficients for both the ACEP and CPW method.

Coefficient	Initial values	Estimated values ACEP	Estimated values CPW
β_1	1	48.4049	-115.8
β_2	1	62.5746	4.9167
β_3	1	-136.0759	-570.4539
β_4	1	0.00257811	2.3248
β_5	1	2.40694	1.1381
β_6	30	30	30
β_7	1	21.0799	-3.7842
β_8	2	2.2661	1.2026
β_9	1	132.9908	589.1836
β_{10}	2	0.03939	0.0071

The adjusted R-squared values for the model are 0.942 and 0.92 for respectively the ACEP and CPW model. The root-mean-squared-error is 2.42 for the ACEP model, and 3.79 for the CPW model. It should also be noted that all p-values for the coefficients were < 0.01 making them statistically significant. To asses the fit of the model to the actual data, Fig. 3.23 displays the relationship between two independent parameters and the dependent variable for the ACEP case. The left hand side of the figure shows the actual data and the right hand side the model response. Since it is impossible to show every possible combination of input parameters (2D plot for 3 input parameters) and because the main goal is to showcase the fit of the model, the values corresponding to the plot were chosen arbitrarily and are displayed above the subplots. As can be seen, the model shows good similarity in response compared to the actual data. Similar results are obtained for the CPW model in Fig.3.24, with the main difference being that for

very low jitter values (<0.5%) and high SNR levels the feature indicator is slightly lower than for higher jitter values (between 0.5% and 2.5%). This can be explained by the fact that for very low jitter values the bearing fault signature in the spectrum is less smeared out and the CPW method is just mainly filtering out that bearing fault signature for higher SNR levels, reducing the amplitude of the peaks in the envelope spectrum.

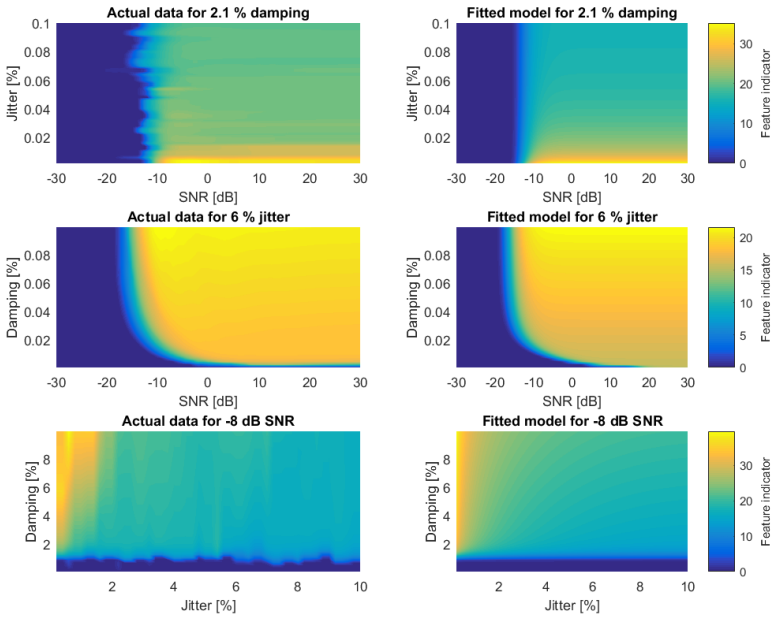

FIGURE 3.23: Comparison of the actual simulated response vs. the estimated model response for the ACEP method.

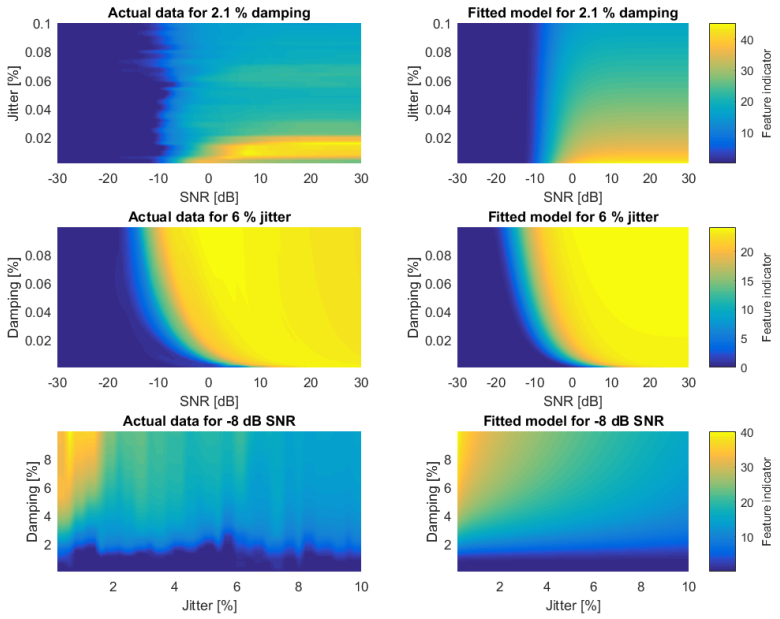

FIGURE 3.24: Comparison of the actual simulated response vs. the estimated model response for the CPW method.

This analysis shows the potential to develop a more complex model, taking into account more input parameters, and to refine the feature calculation process. This could lead to a useful evaluation tool for assessing the strengths and weaknesses of methods that try to improve bearing fault detection. This way you can get an overview of potential reasons why a new method does not work well on experimental data.

3.3.6 Possible other influences

The performed analysis covers far from all possible influences on the performance of the CPW or ACEP method, but it does provide insight on the strong and weak points of both methods for some key influencing factors. However, one can easily come up with possible other

important elements. Examples of such influencing factors could include:

- Resonances carrying impulsive information instead of noise.
- Randomly occurring impulsive noise events.
- Other modulating components, it should be noted that even though the CPW and ACEP method alter the signal content, they are not designed for multiplicative component removal.
- Other bearing fault signal parameters such as the type of the bearing fault: outer race fault, inner race fault, cage fault or ball fault.
- Coloured noise instead of white noise.
- Amount of speed variation.

3.4 Comparison of experimental results

To further assess the differences between the two methods, a comparison of the two techniques is also carried out on an experimental setup. Data from the *wind turbine gearbox condition monitoring round robin study* provided by NREL, the National Renewable Energy Laboratory, is used because of its well-documented and interesting setup. Only a short description of the experimental setup is provided since a more elaborate analysis procedure is investigated on this data set in Chapter 6. The focus in this chapter is on the performance comparison between CPW and ACEP.

3.4.1 Description of experimental setup

The test turbine is a three-bladed, stall-controlled, upwind turbine. It has a rated power of 750kW and the generator normally operates at 1800 rpm or 1200 rpm nominal. The complete drivetrain was installed in the NREL dynamometer test facility. It was hard fixed to the floor and missed the hub, rotor, yaw bearing and yaw drives. The used gearbox consists of one low speed (LS) planetary stage and two parallel

stages. The used nomenclature is the same as the one used by Sheng et al. [259] and can be seen in Fig.3.25. Two loss-of-oil events damaged its internal bearings and gears while in the field. Afterwards, it was again installed in the dynamometer facility and retested in a controlled environment. Various bearing types are used in the gearbox corresponding to the loading conditions and life requirements. Two full-complement cylindrical roller bearings (fcCRB) support the planet carrier and two cylindrical roller bearings (CRB) support the planet gears. The parallel shafts are each supported by a CRB on the upwind side and by two tapered roller bearings (TRB) on the downwind side of the assembly. A list of all bearing locations, manufacturers, part numbers and types can be found in [259] and Fig.3.25 illustrates the locations and used names of the different bearings. If the component is positioned upwind, this is denoted with an 'A' and if it is downwind, with a 'B' or a 'C'. The accelerometers were mounted on the outside of the gearbox and data was sampled at 40 kHz per channel. In total data sets of eight accelerometers (Model: IMI 626B02) were made available. Table 6.4 gives an overview of the fault frequencies present in the gearbox vibration signals and that are investigated in this chapter. In total, four bearing fault frequencies are examined sequentially.

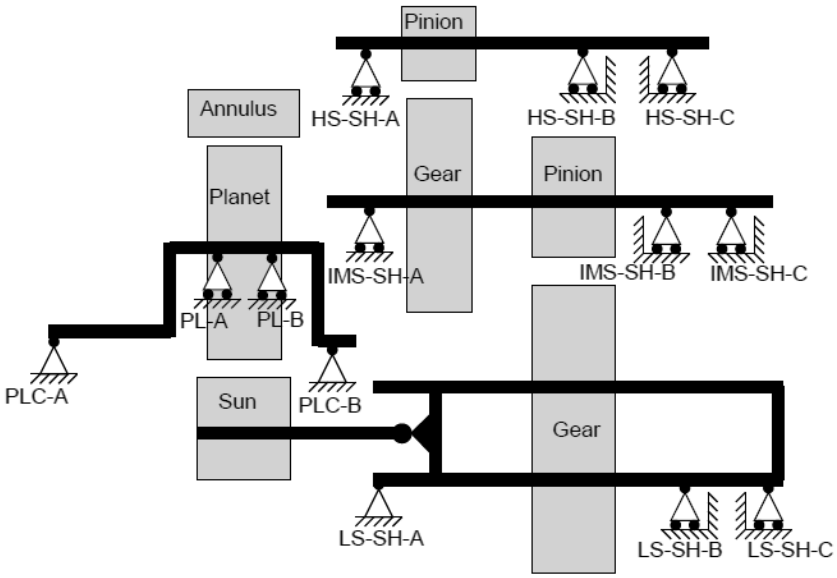

FIGURE 3.25: View of internal components of the test gearbox and the nomenclature and locations of the bearings and gears.

TABLE 3.2: Damaged bearings and their corresponding theoretical characteristic frequencies.

Bearing label	**Fault type**	**Detectable characteristic frequency [Hz]**
HS-SH downwind bearings (*HSS-B& C*)	BPFI FTF	345,29Hz 12,75 Hz
IMS-SH downwind bearings (*ISS-B& C*)	BPFO	105,75 Hz
IMS-SH upwind bearing (*ISS-A*)	BPFI	73,7 Hz

3.4.2 Results

The CPW and ACEP methods are compared for the previously mentioned bearing frequencies through the use of envelope analysis. It should also be noted that all signals are first order tracked in order to minimize the amount of speed variation present in the spectra.

3.4.2.1 High-speed downwind bearings

The first analysis examines the observed fundamental train frequency and inner race fault frequency of the HSS-B&C bearings. The data used originates from the AN7 accelerometer, which is located close to the high-speed shaft. Figure 3.26a displays the spectrum of the damaged AN7 signal before and after ACEP, below Fig.3.26b gives an indication of the reduction in dB of the first hundred shaft speed harmonics. It can be seen that a significant reduction is obtained. Figure 3.27 shows a zoom of the full bandwidth envelope spectra generated by using the developed methods around the FTF frequency of 12.75 Hz. It can be seen that the ACEP method amplifies significantly the fault peak here compared to the original signal and the CPW signal. Cepstrum pre-whitening only emphasizes the fault peak slightly.

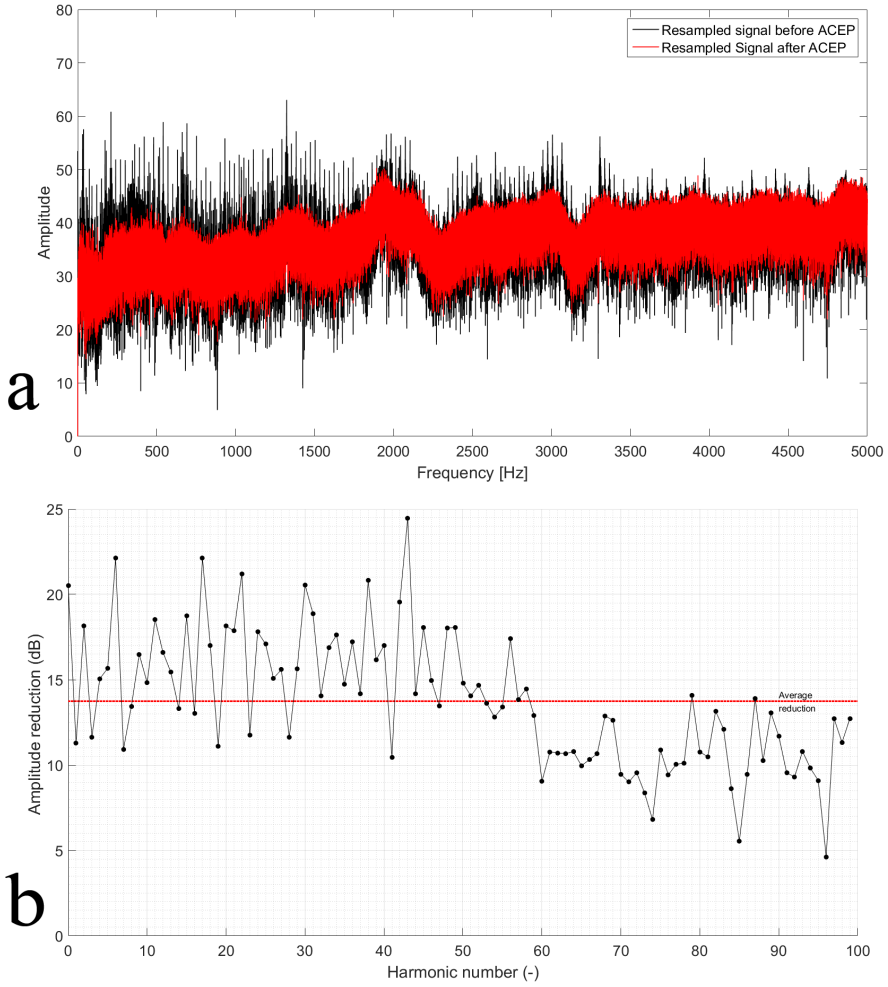

FIGURE 3.26: (a) Amplitude spectrum of the AN7 sensor signal before and after ACEP. (b) Graph showing the amplitude reductions, in dB, of the first 100 shaft speed harmonics.

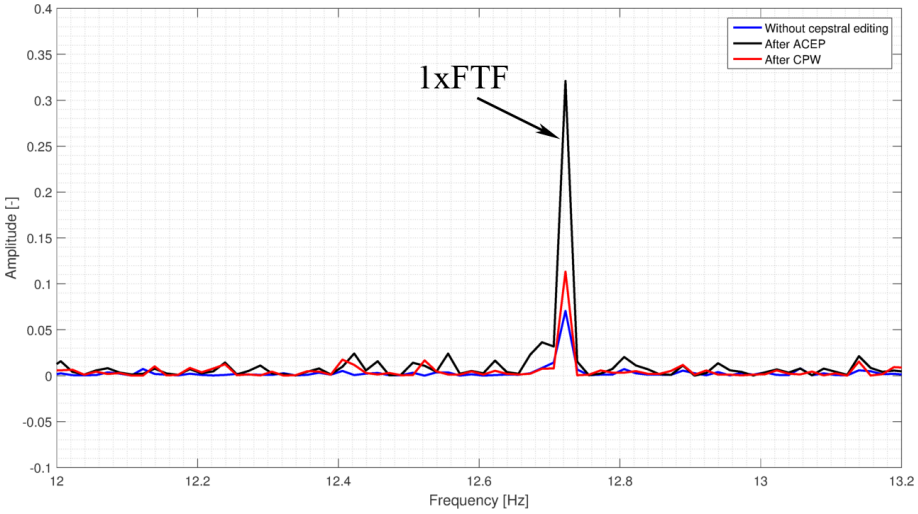

FIGURE 3.27: Zoom of the squared envelope spectra around the FTF frequency of 12,75 Hz of the HSS-B&C bearings for the signal without cepstral editing, and after ACEP and CPW.

However, the opposite is true for the ball pass fault frequency of the inner race around 345.2 Hz. Figure 3.28 exhibits a clear peak for the CPW method, while the original signal and the ACEP signal show no sign of a BPFI frequency. It is noted that band-pass filtering around 17 kHz with a bandwidth of 2 kHz (based on the kurtogram) does reveal the fault for the ACEP method as well.

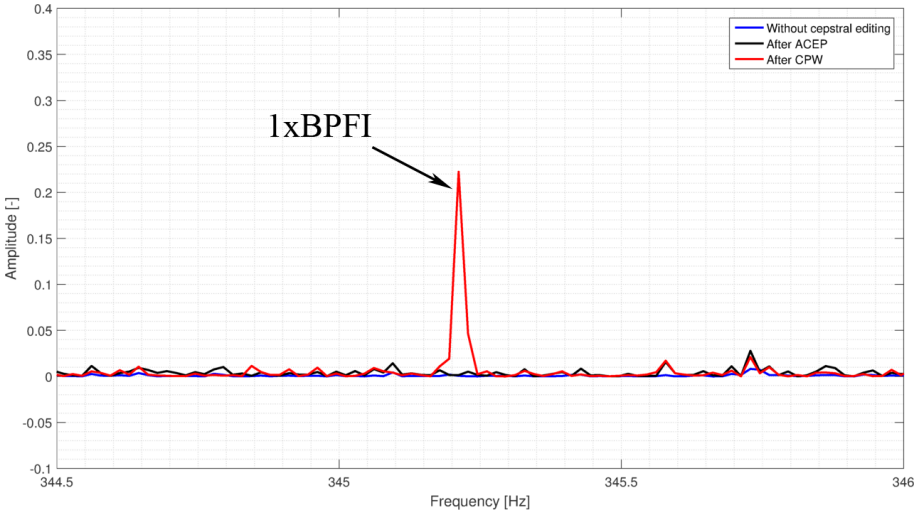

FIGURE 3.28: Zoom of the squared envelope spectra around the BPFI frequency of 345.2 Hz of the HSS-B&C bearings for the signal without cepstral editing, and after ACEP and CPW.

3.4.2.2 Intermediate-speed shaft upwind bearing

The next analysis looks at the inner race fault of the ISS-A bearing. It was stated in the damage report that this bearing showed signs of plastic deformation, debris dents, assembly damage, scuffing, false brinelling and contact corrosion. Figure 3.29 displays a zoom of the envelope spectra around the BPFI frequency of 72.93 Hz. It can be seen that the peak amplitudes remain small for all three signals, and that the CPW envelope spectrum contains more noisy content than the other two spectra. The two methods do not filter out this frequency however, indicating that it most likely does not belong to a deterministic source. Again, it should be noted that improving the peak amplitude is possible through band-pass filtering with a center frequency of 10 kHz with a bandwidth of 1 kHz. This filtering step is not the focus of this chapter however, so it is not illustrated here.

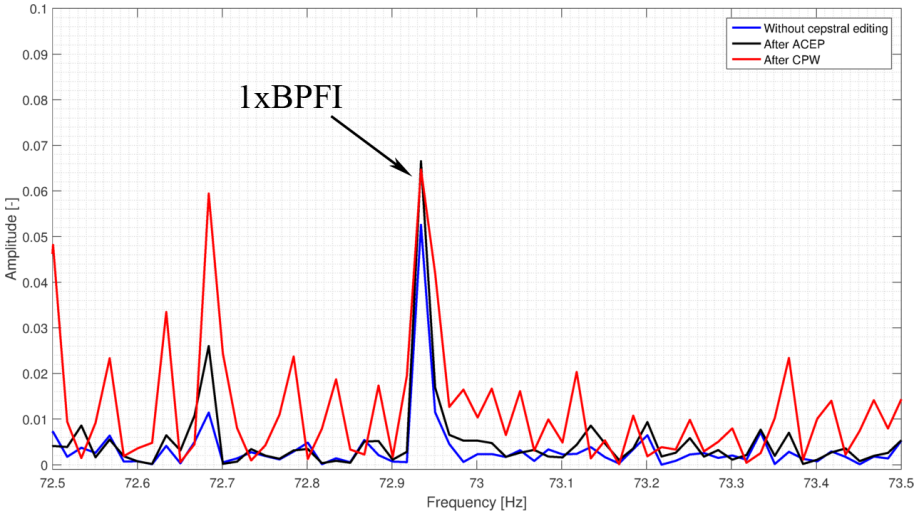

FIGURE 3.29: Zoom of the squared envelope spectra around the BPFI frequency of 72.93 Hz of the ISS-A bearing for the signal without cepstral editing, and after ACEP and CPW.

3.4.2.3 Intermediate-speed shaft downwind bearings

Lastly, the ISS-B&C bearings are inspected for their BPFO frequency of 105.2 Hz, using the data set of the AN6 accelerometer which is close to the intermediate-speed stage. The damage report declared that they observed plastic deformation, assembly damage and multiple dents for these bearings. The analysis of the fault frequency is however not as straightforward as the other ones. While most of the round robin research partners did find the second and higher even harmonics of the outer race fault frequency, they did not observe any strong indication of the fundamental fault frequency. Likewise it can be seen in Fig. 3.30a that also in this analysis there appears to be no high amplitude envelope peak around this frequency. The envelope spectrum after CPW exhibits the largest peak at this frequency, while the ACEP spectrum is too noisy to clearly discern the BPFO frequency. Analysis of the higher harmonics though reveal high amplitudes at the even harmonics (4x, 6x, 8x) of the BPFO frequency. Especially the ACEP method emphasizes the even BPFO peak as can be seen in Fig. 3.30b.

While it is not entirely certain why the fundamental frequency is so badly detected, it might be allotted to the presence of multiple dents on the outer ring. These multiple point defects can introduce cancellation and reinforcement of certain frequency components, which can modify the appearance of the envelope spectrum according to McFadden et al [260].

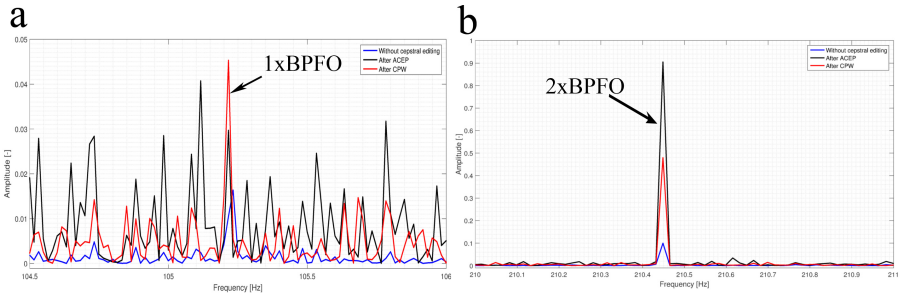

FIGURE 3.30: (a) Zoom of the squared envelope spectra around the fundamental BPFO frequency of 105.2 Hz of the ISS-B&C bearings. (b) Zoom of the squared envelope spectra around the second harmonic of the BPFO at 210.4 Hz.

3.5 Discussion

The question arises as to whether the ACEP method and the CPW method could be merged into a unified liftering method where liftering is done in such a way that it removes harmonics but does not equalize completely the resonances. A unified liftering approach that combines the advantages of both methods would be desirable. Such a method could make use of knowledge about the resonance damping and recently it has already been proposed for this purpose by Randall & Smith [261]. They suggest the use of an exponential lifter where the time constant is set using the damping of the lowest resonance frequency present in the signal. This exponential lifter smooths the spectrum to retain mainly the resonances and reduces the amplitude of the harmonics significantly. Compared to the ACEP method it is less forceful in removing the harmonics through the cepstrum though, so there might still be residual harmonics even after exponential liftering, since the

corresponding rahmonics are not totally set to zero. The performance of this exponential lifter has not been investigated here.

Another idea could be to optimize the liftering in such a way that it optimizes the feature indicator defined in Eq. 3.14. This approach does require a priori knowledge about the presence of a fault and the corresponding fault frequencies. Moreover the amount of bearing slip is usually not known and can cause variations around the theoretical fault frequency. This can be resolved by taking some tolerances into account for the indicator calculation.

For complex systems the sole step of cepstrum editing might be insufficient to properly detect the fault frequency, necessitating an additional band-pass filtering step (e.g. through the use of the kurtogram). In regard to the ACEP method, there is an additional risk of instability of the optimization when iterating over possible input parameter combinations for the method. However, it is definitely promising to develop an optimization scheme for cepstrum liftering, but depending on the implementation, the resulting method should try to keep the complexity as low as possible.

3.6 Conclusion

This chapter investigates two effective approaches for discrete-random signal separation based on the cepstrum. The performance of the automated cepstrum editing procedure (ACEP) and the cepstrum prewhitening (CPW) method is compared.

First, both methods are described theoretically and compared on simulated signals. The sensitivity for fault detection of both methods is assessed for varying SNR values of the bearing fault signal. An analysis is made comparing ACEP and CPW for the case of additive white Gaussian noise, where it is concluded that ACEP can outperform CPW for low SNR values due to the preservation of the bearing resonance. A similar analysis with additional noisy resonances is carried out. Here the CPW method outshines ACEP for lower SNR values. ACEP retains the resonances which heavily influences the envelope spectra when demodulating the full bandwidth of the signal. Other possible

influences are not further investigated, but suggestions are made for potential future studies.

Finally, the two techniques are examined on experimental data of the *wind turbine gearbox condition monitoring round robin study*, provided by the National Renewable Energy Laboratory (NREL). The results are less clear here than for the virtual signals. While the CPW method does indicate most of the bearing faults, the fault frequencies are sometimes surrounded by high levels of noise in the envelope spectrum, making it difficult to interpret. For some fault frequencies, the ACEP method performs better in amplifying the fault peaks than the CPW method, but it can totally miss a fault frequency peak without additional band-pass filtering, as is the case for the BPFI frequency of the HSS-B&C bearings. ACEP also requires more fine-tuning than CPW, making it a bit more cumbersome for automated industrial applications. Since CPW removes the influence of other high amplitude components like resonances, it is for some occasions able to increase the sensitivity of the full bandwidth envelope spectrum for these weaker components. This makes it potentially better suited for applications in which band-pass filtering is not employed.

Chapter 4

Signal filtering

4.1 Introduction

This chapter discusses the concept of vibration signal filtering from the perspective of increasing the signal-to-noise ratio of a potential fault signature. This may seem somewhat overlapping with the topic of the previous chapter since it discussed the *signal separation* of deterministic and stochastic content which in essence also can increase the fault signal-to-noise ratio. However, the techniques in this chapter are typically utilized after the signal separation step since the *signal filtering* described here make use of the properties of the residual stochastic signal that would otherwise be too negatively influenced by the deterministic content. In other words, the signal separation is considered a necessary precursor for further signal filtering.

After the signal separation, the residual stochastic signal normally still consists of a mixture of random signals, but the hope or assumption is that these signals occupy different frequency bands such that they can be separated by spectral filtering. Therefore, the filtering step often revolves around finding a filter that optimizes a certain criterion that characterizes the fault signature of interest such that the frequency band(s) of interest can be identified. In the case of impulsive events such as a bearing fault, the source signal is often assumed to be an impulse train at the repetition frequency of the bearing characteristic fault frequency. Two filtering approaches are commonly used for filtering out such an impulsive signature.

One way is to search for a good narrow-band filter that maximizes the kurtosis or impulsiveness of the filtered signal by iteratively going

through many filters with pre-defined center frequencies and bandwidths, i.e. the kurtogram [262]. More specifically, the vibration signal is put through a filterbank and the criterion of interest is calculated for the full filterbank. An example of the dyadic frequency grid for the filterbank used by the kurtogram is shown in Fig. 4.1.

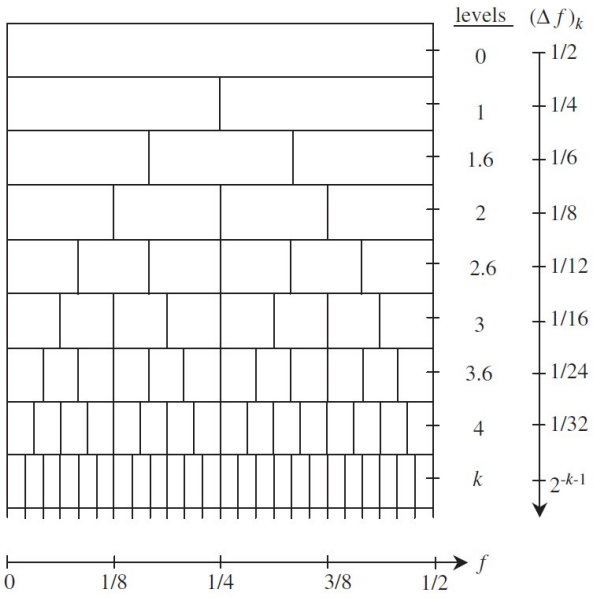

FIGURE 4.1: Filterbank used by kurtogram [262]

Another way is to drop the narrow-band assumption and to search for a filter that iteratively optimizes its own filter coefficients and converges to a filter that maximizes the filtered signal's kurtosis, i.e. the minimum entropy deconvolution (MED) technique [188]. The original idea employed by this latter MED approach is to recover the source signal of bearing fault impulses by finding an inverse filter that compensates the transmission path. An example of MED filtering is shown in Fig. 4.2. In this figure you can clearly see the recovered fault impulses after the MED filtering.

The methodology of the second approach is often called a *blind* approach since it allows to find an optimal filter without any prior knowledge. colorblueBlind filtering, as used in this work, is defined

similar to that of blind signal separation (BSS), meaning the separation of a signal or a set of signals from a set of signal mixtures, without the aid of information or with very little information about the signals of interest or the mixing process. This blind filtering concept is the main focus of this chapter and a new framework for deriving such blind filters for vibration analysis is proposed. First however, the need for blind filtering approaches is explained as to provide some background for its relevance in further research.

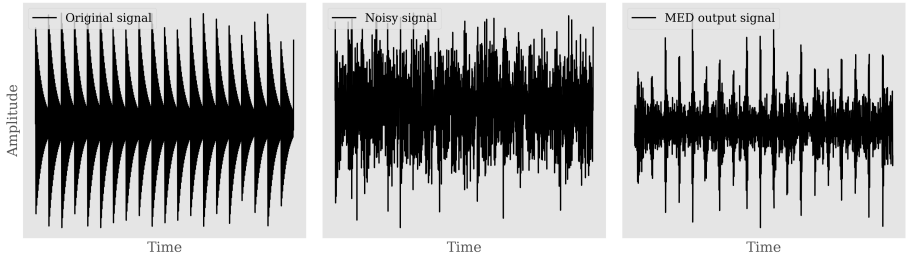

FIGURE 4.2: Example of application of Minimum Entropy Deconvolution [188] on a simulated noisy bearing fault signal, (Left) Original bearing fault signal, (Middle) With added white Gaussian noise at SNR of -4 dB, (Right) Resulting signal after MED filtering.

4.1.1 Background

The interest of industry in more advanced condition monitoring tools is growing thanks to the ever-improving computational power, expanding storage capabilities, and decreasing cost of modern day IT systems. Despite this trend, it sometimes remains difficult to implement some of the more advanced methods due to incomplete knowledge about the system of interest. Complex machines can consist of dozens of bearings and gears, modern examples are gearboxes of wind turbines and helicopters. These machines typically have one or more planetary gear stages in combination with parallel gear stages. Not all kinematic information of the manufacturer about the system might be available to the machine operator, or the information might be inaccurate due to reparations with new components. This issue constitutes

the need for methods capable of tracking the condition of these components without the need for prior knowledge about the kinematics or, in other words, blindly.

The concept of tracking signal characteristics blindly is already widely employed in industry thanks to the simplicity of time-domain statistical indicators. Nevertheless there is still room to improve further on this blind signal analysis concept and to bring it better in line with other state-of-the-art vibration analysis methods. One of the most popular approaches for fault detection is to look at the cyclostationary behavior of the vibration signal [176, 258, 263–269]. Inspired by this fact, this chapter investigates the possibility to utilize the cyclostationary content of a signal in a blind manner. Therefore, instead of just looking at the statistics of the time waveform, the squared envelope and its spectrum of the signal are employed as blind means to gain more information about potential defects. The envelope is probably the most frequently used tool to inspect the cyclostationary behavior of signals. From experience it is known that most mechanical faults of bearings or gears induce some form of cyclostationary behavior in the observed vibration signals [149, 265, 270]. This phenomenon causes the envelope time waveform to become more sinusoidal and thus more structured. Additionally, the envelope spectrum will exhibit discrete peaks at the corresponding fault frequencies. These peaks protrude from the background noise and thus their amplitude is distinguishable from the spectral noise floor. In other words, most mechanical faults increase the predictability of the envelope time waveform and the sparsity of the envelope spectrum.

4.1.2 Sparsity

Sparsity is a very useful concept that has become a widely used tool across various fields in recent years. In the context of 1D signal vectors, sparsity quantifies the inequality in the data. Intuitively, a sparse representation is one in which a small number of coefficients contain a large proportion of the energy. A concrete example of what is meant by this is given in Fig. 4.3. If the randomly generated signal on the left of the figure is considered to be values of the distribution of wealth

among citizens, it would be considered a fairly equal wealth distribution, i.e. not sparse. However, the signal on the right of the figure would be considered a lot less equal and in this case more sparse.

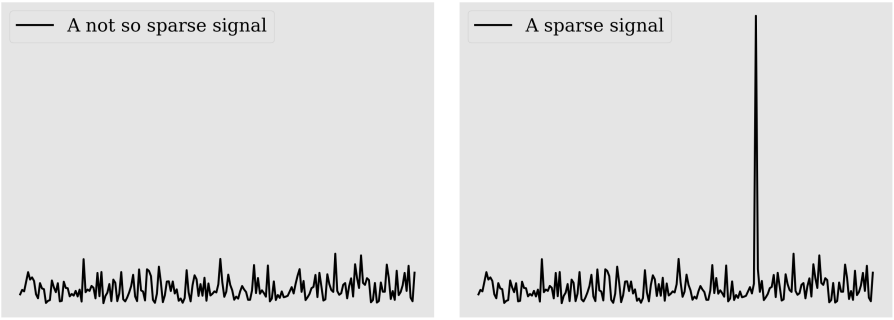

FIGURE 4.3: Illustration of sparsity principle used in this chapter, (left) Signal values that are mostly equal without any significant outliers, i.e. high equality and low sparsity, (right) The same signal but with one clear outlier being significantly higher in value, thus low equality and high sparsity.

The use of sparsity (or sparseness) in numerical analyses saw a significant increase in interest from scholars thanks to the introduction of compressed sensing in 2004. Candes et al. [271, 272] and D. Donoho [273, 274] proved that when knowledge is available about the sparsity of a signal, that signal can potentially be reconstructed with even fewer samples than strictly required by the Nyquist-Shannon sampling theorem. This idea is the basis of compressed sensing. Figure 4.4 shows how the usage of sparsity increased almost exponentially in research efforts over the last few years, as listed by the Web of Science [275]. This sparsity property can now also be exploited to find a filter that maximizes the sparsity of the envelope spectrum since it can be assumed that this corresponds to a mechanical fault and not to normal behavior.

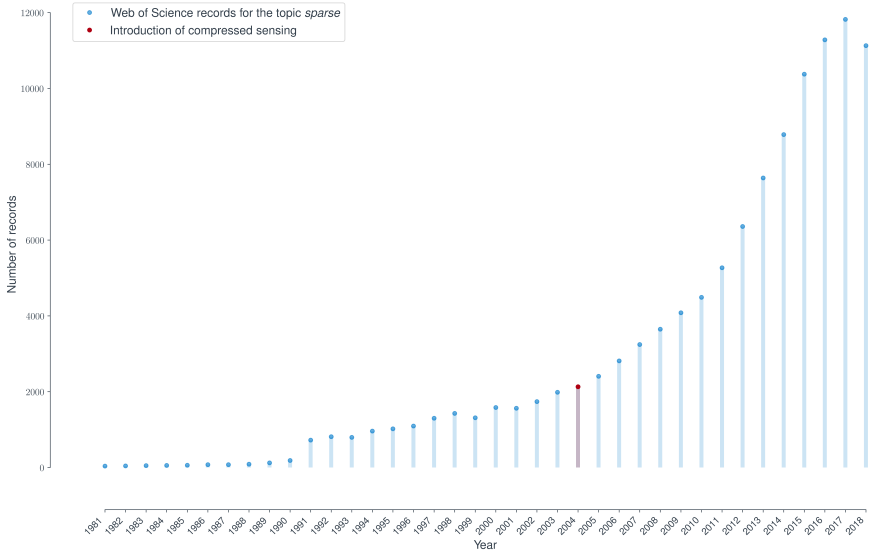

FIGURE 4.4: Evolution of the number of research records in the Web of Science related to the topic of 'sparsity'.

4.1.3 Blind filtering

This chapter illustrates the concept of blind filtering using sparsity indicators on the envelope spectrum and envelope predictability as a way of tracking the health of bearings. The proposed blind filtering approach is similar to the blind deconvolution idea employed by Minimum Entropy Deconvolution (MED) filtering [188] which uses the kurtosis of the time waveform as the statistic to maximize. MED filtering was orginally applied in the field of seismic signal processing.

Kurtosis however has some disdvantages as a measure to maximize [276]. Kurtosis tends to be maximized by a single large peak amplitude, which is rarely realistic for a rotating component fault. Therefore, improvements and variations on the MED approach were developed afterwards. Cabrelli [277] suggested to use the D-norm (MEDD), an alternative to kurtosis, as a way to find a direct solution to the blind deconvolution problem. Wang et al. reduced the computation time of MEDD by using an adaptive least-squares approach [278]. Broadhead

et al. compared the performance of MED and MEDD for varying signal-to-noise ratios (SNR) and found that MEDD outperforms MED in low to moderate SNR. W. Gray [279] analyzed normalized higher-order moments other than kurtosis. McDonald et al. [280] made a comprehensive investigation of MED and its flaws. They afterwards suggest several improvements and extensions to MED in an attempt to circumvent these flaws, such as using multi-point kurtosis instead of standard kurtosis. Obuchowski et al. [189] opted to use the Jarque-Bera statistic, a combination of skewness and kurtosis, for the detection of gear faults. Very often the MED-related techniques get paired with other pre-processing methods, like autoregressive filtering or cepstrum prewhitening, in order to remove deterministic content that could potentially hinder successful deconvolution of the fault signal [281–285]. The need for a novel approach, and the origin of the idea in this chapter for the proposed blind filtering approach, was actually identified by looking at the available literature about blind deconvolution filters and making the observation that a large majority of techniques is based on the use of kurtosis (or a derivative of kurtosis) as an optimization metric.

In recent years, the blind filtering research has experienced a shift in focus from looking at the time waveform statistics to looking at the cyclostationary properties. Maximum correlated kurtosis deconvolution (MKCD) and the multipoint optimal minimum entropy deconvolution (MOMEDA) [280] are technically cyclostationary approaches but were empirically introduced from the perspective of trying to maximize the impulsiveness linked to a certain fault period. Thanks to the recent efforts in further development and promotion of the concept of cyclostationarity for vibration analysis [257, 258, 269, 286, 287], the topic is now a lot more understood and more widely used. A newly proposed approach [288] employs an indicator of second-order cyclostationarity (ICS2) in the process of blind deconvolution. It allows the user to maximize the ICS2 in an efficient and flexible manner thanks to the versatility of the generalized Rayleigh quotient. However, the approach still requires prior knowledge of the fault frequencies in order to work.

4.1.4 Proposed blind filtering concept

One way to remedy this requirement of prior knowledge, is applying the concept of sparsity to the envelope spectrum, which is probably the most popular tool to analyze the cyclostationary content of a vibration signal. Wang et al. [289] suggested to use the $\frac{l_2}{l_1}$-norm on the envelope spectrum of filtered signals. They extend the filterbank idea, originating from the kurtogram, to utilize the $\frac{l_2}{l_1}$-norm instead of the kurtosis. The downside of this approach is however that it is still not fully blind in the sense that an end-user still has to inspect the filterbank with the $\frac{l_2}{l_1}$-norm values and choose the appropriate frequency band. Additionally, it makes the assumption that the signal of interest has a narrow-band frequency signature, which can limit the applicability. In contrast, the proposed methodology employs the discussed blind filtering approach such that a suitable filter that maximizes the envelope spectrum sparsity is found automatically and the filtered signal can be wideband. The main "selling" point of the proposed approach as compared to the ICS2 filter or similar methods is the fact that no prior information is needed about the characteristic fault frequencies. The proposed method provides a potential way to track the presence of any modulating fault signature using conventional finite-impulse response filters.

Using the proposed methodology which is primarily based on the usage of the Rayleigh quotient, another blind filtering approach is derived using the envelope predictability. The emergence of a cyclostationary fault and the subsequent introduction of modulating behavior in the signal means that the envelope waveform becomes more predictable and can thus be fitted with an autoregressive model. This property is exploited to find a filter that minimizes the relative prediction error of the squared envelope since it is assumed that a good fit corresponds to a mechanical fault and not to normal behavior.

An important remark about the proposed approach is that the blind filtering methodology described in this chapter cannot be categorized as blind deconvolution, blind signal separation, or denoising. The proposed approach namely does not attempt to deconvolve the signal in order to recover the source signal (e.g. impulses), nor does it attempt

to separate signals from a mixture or remove noise from the signal without distorting it. In fact, it actually does distort the signal such that the envelope is as predictable as possible or that the envelope spectrum is as sparse as possible. This is a rather new concept since from this perspective the algorithm does not care about restoring the signal or recovering the source signals. Instead the algorithm just tries to maximize the figure of merit. It is important to take this distinction into account when inspecting filtering results since the results might not correspond to what is expected.

4.1.5 Chapter summary

This chapter attempts to highlight several things: the utility of sparsity from a cyclostationary perspective, the potential of using the envelope waveform directly for filtering without any need of its spectral counterpart, and it emphasizes the versatility of the Rayleigh quotient regarding the indicator choice for the blind filtering step. First, the theoretical background is explained in Section 4.2. The indicator choice and the derivation of the Rayleigh quotients are described. Next, the methodology is validated on simulated signals in Section 4.3 and experimental data of a gearbox trending data set in Section 4.4. The results show that the proposed approach is capable of extracting a cyclostationary fault signature and that both the sparsity measure of the envelope spectrum and the envelope prediction error can be used as a tracking parameters.

4.2 Methodology

Before going into detail how to optimize the sparsity of the envelope spectrum or the prediction error of the envelope waveform, the mathematical quantities of the envelope and its spectrum are expressed in matrix notation in section 4.2.1. Afterward, the derivation of the blind filters is specified for the relative prediction error of the envelope and for four different sparsity indicators of the envelope spectrum.

4.2.1 Blind filtering

The concept of blind filtering is to find a filter that maximizes a certain criterion of the signal starting from a noisy measured signal $\boldsymbol{x}$:

$$\boldsymbol{s} = \boldsymbol{x} * \boldsymbol{h} \tag{4.1}$$

where $\boldsymbol{s}$ is the estimated input, $\boldsymbol{h}$ is the inverse filter, and $*$ refers to the convolution operation. It should be noted that vectors and matrices are set in bold font to illustrate the difference with scalars. The convolution is expressed as:

$$\boldsymbol{s} = \boldsymbol{X}\boldsymbol{h} \tag{4.2}$$

$$\begin{bmatrix} s_{N-1} \\ \vdots \\ s_{L-1} \end{bmatrix} = \begin{bmatrix} x_{N-1} & \dots & x_0 \\ \vdots & \ddots & \vdots \\ x_{L-1} & \dots & x_{L-N-2} \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ h_{N-1} \end{bmatrix}$$

with L and N the number of samples of $\boldsymbol{s}$ and $\boldsymbol{h}$ respectively.
Now the squared envelope $\boldsymbol{\epsilon_x}$ can be defined as follows:

$$\boldsymbol{\epsilon_x} = |\boldsymbol{s}|^2 = |\boldsymbol{X}\boldsymbol{h}|^2 \tag{4.3}$$

It can also be written as:

$$\boldsymbol{\epsilon_x} = \begin{bmatrix} s_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{L-N+1} \end{bmatrix}^H \boldsymbol{X}\boldsymbol{h} = diag(\boldsymbol{s^H})\boldsymbol{X}\boldsymbol{h} \tag{4.4}$$

with $\boldsymbol{s^H}$ being the Hermitian transpose of $\boldsymbol{s}$, and $diag(\boldsymbol{s^H})$ being a diagonal matrix with the values of the vector $\boldsymbol{s^H}$ on its diagonal. The squared envelope spectrum $\boldsymbol{E_x}$ is then the Fourier transform of this squared envelope:

$$\boldsymbol{E_x} = \boldsymbol{F^H}\boldsymbol{\epsilon_x} = \boldsymbol{F^H} diag(\boldsymbol{s^H})\boldsymbol{X}\boldsymbol{h} \tag{4.5}$$

with:

$$\boldsymbol{F} = \begin{bmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & e^{-2\pi j \frac{1}{L-N}} & \dots & e^{-2\pi j \frac{k}{L-N}} & \dots & e^{-2\pi j \frac{K-1}{L-N}} \\ \vdots & \vdots & \ddots & & & \vdots \\ 1 & e^{-2\pi j \frac{n}{L-N}} & \dots & e^{-2\pi j \frac{kn}{L-N}} & \dots & e^{-2\pi j \frac{(K-1)n}{L-N}} \\ \vdots & \vdots & & & \ddots & \vdots \\ 1 & e^{-2\pi j \frac{L-N-1}{L-N}} & \dots & e^{-2\pi j \frac{k(L-N-1)}{L-N}} & \dots & e^{-2\pi j \frac{(K-1)(L-N-1)}{L-N}} \end{bmatrix} \tag{4.6}$$

with $n = 0..L-N-1$, $k = 0..K-1$ and $K-1$ corresponding to the index of the maximum frequency of interest. $\boldsymbol{F}$ has dimensions $(L-N, K)$. The defined variables now allow deriving filters for the envelope-based indicators in a fairly efficient manner.

4.2.2 Derivation of LP-envelope filter

Before deriving the envelope spectrum sparsity-based filters, we take a look at the filter using the prediction error of the squared envelope in the time domain. The idea of the proposed methodology is to exploit the predictability of a fault modulation signature by trying to fit it with a linear prediction (LP) filter or auto-regressive (AR) filter. The prediction error of an auto-regressive all-poles model of the squared envelope serves then as the metric of interest. The better the AR model can fit the actual envelope, the more predictable and thus the less noisy it is. This means that if there is a signal component present with e.g. a clean sinusoidal amplitude modulation, the AR model is then capable of predicting future samples accurately which in turn corresponds to a low prediction error. This does indicate the potential need for prewhitening the signal to make sure the AR model does not try to fit the envelope of deterministic components in the signal.

The relative prediction error of the AR model is closely related to the spectral flatness as the AR model also maximizes the spectral flatness of the squared envelope prediction error [290]. The relative prediction error (RPE) of the AR model of the squared envelope is given

by:

$$RPE = \frac{\sigma_e^2}{\sigma_{SE}^2} \tag{4.7}$$

with σ_e being the prediction error of the squared envelope, and σ_{SE} being the standard deviation of the squared envelope. The autoregressive coefficients can be obtained by fitting a linear prediction model on the squared envelope. The standard LPC representation of a signal $x(n)$ for a model of order N is:

$$x(n) = \sum_{i=1}^{N} a_i x(n-i) + e(n) \tag{4.8}$$

with a_i the autoregressive coefficients, and where the error $e(n)$ can be obtained by:

$$e(n) = \sum_{i=0}^{N} a_i x(n-i) \; with \; a_0 = 1. \tag{4.9}$$

Figure 4.5 illustrates the straightforward process. The autoregressive filtering necessary to obtain the prediction error $\boldsymbol{e}$ of the squared envelope ϵ_x can be written in matrix notation as follows:

$$\boldsymbol{e} = \boldsymbol{A} diag(\boldsymbol{s^H}) \boldsymbol{s} \tag{4.10}$$

with $\boldsymbol{A}$ a band matrix containing the autoregressive coefficients as follows:

$$\boldsymbol{A} = \begin{bmatrix} a_0 & a_1 & \dots & a_N & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_N & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_0 & a_1 & \dots & a_N \\ & & & & \ddots & \ddots & \vdots \\ & & & & & a_0 & a_1 \\ 0 & & & \dots & & 0 & a_0 \end{bmatrix} \tag{4.11}$$

Since the intent is actually to minimize the relative prediction error in Eq. 4.7 and not maximize it, the ratio to be maximized is actually the

FIGURE 4.5: Input-output relationship between the prediction error and the sampled signal.

inverted RPE. Equation 4.7 can thus be written in the following manner:

$$\frac{\sigma_{SE}^2}{\sigma_e^2} = \frac{\boldsymbol{\epsilon_x}^H \boldsymbol{\epsilon_x}}{\boldsymbol{e}^H \boldsymbol{e}} = \frac{\boldsymbol{h}^H \boldsymbol{X}^H diag(\boldsymbol{s}) diag(\boldsymbol{s}^H) \boldsymbol{X} \boldsymbol{h}}{\boldsymbol{h}^H \boldsymbol{X}^H diag(\boldsymbol{s}) \boldsymbol{A}^H \boldsymbol{A} diag(\boldsymbol{s}^H) \boldsymbol{X} \boldsymbol{h}} = \frac{\boldsymbol{h}^H \boldsymbol{R}_{XW_1X} \boldsymbol{h}}{\boldsymbol{h}^H \boldsymbol{R}_{XW_2X} \boldsymbol{h}} \quad (4.12)$$

The generalized Rayleigh quotient [291] can be recognized in Eq. 4.12 and can be maximized using an iterative maximization of the eigenvalues:

$$RQ(\boldsymbol{h}) = \frac{\boldsymbol{h}^H \boldsymbol{R}_{XW_1X} \boldsymbol{h}}{\boldsymbol{h}^H \boldsymbol{R}_{XW_2X} \boldsymbol{h}} \quad (4.13)$$

The Rayleigh quotient has the interesting property that its maximal value with respect to $\boldsymbol{h}$ is equivalent to its largest eigenvalue λ and corresponding eigenvector. The proof behind this property is actually relatively short. To find the optimal eigenvector $\boldsymbol{h}$ corresponding to the maximum or minimum of $RQ(\boldsymbol{h})$ in Eq. 4.13, we need to find its derivative with respect to $\boldsymbol{h}$. First, we define the scalar numerator of $RQ(\boldsymbol{h})$ as α, we then know that α is given by:

$$\alpha = \sum_{j=1}^{N} \sum_{i=1}^{N} c_{ij} h_i h_j \quad (4.14)$$

where c_{ij} are the values of the symmetric correlation matrix $\boldsymbol{R}_{XW_1X}$. Differentiating with respect to the k^{th} element of $\boldsymbol{h}$, we get:

$$\frac{d\alpha}{dh_k} = \sum_{j=1}^{N} c_{kj} h_j + \sum_{i=1}^{N} c_{ik} h_i \quad (4.15)$$

for $k = 1, 2, \ldots, N$. This can be written as:

$$\frac{d\alpha}{d\boldsymbol{h}} = \boldsymbol{h}^T \boldsymbol{R}_{XW_1X}^T + \boldsymbol{h}^T \boldsymbol{R}_{XW_1X} = \boldsymbol{h}^T (\boldsymbol{R}_{XW_1X}^T + \boldsymbol{R}_{XW_1X}) \quad (4.16)$$

Since $\boldsymbol{R}_{XW_1X}$ is symmetric, $\boldsymbol{R}^T_{XW_1X} = \boldsymbol{R}_{XW_1X}$, so the derivative of the numerator ends up being:

$$\frac{d\alpha}{d\boldsymbol{h}} = 2\boldsymbol{h}^T\boldsymbol{R}_{XW_1X} \tag{4.17}$$

The denominator can be derived in the exact same way. Thus, we can now simply apply the quotient rule for the derivative of the full Rayleigh quotient, giving us:

$$\frac{d}{d\boldsymbol{h}}RQ(\boldsymbol{h}) = \frac{2\boldsymbol{h}^T\boldsymbol{R}_{XW_1X}(\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_2X}\boldsymbol{h}) - 2\boldsymbol{h}^T\boldsymbol{R}_{XW_2X}(\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_1X}\boldsymbol{h})}{(\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_2X}\boldsymbol{h})^2} \tag{4.18}$$

Setting this expression to zero, gives:

$$\boldsymbol{h}^T\boldsymbol{R}_{XW_1X}(\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_2X}\boldsymbol{h}) = \boldsymbol{h}^T\boldsymbol{R}_{XW_2X}(\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_1X}\boldsymbol{h}) \tag{4.19}$$

$$\boldsymbol{h}^T\boldsymbol{R}_{XW_1X} = \frac{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_1X}\boldsymbol{h}}{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_2X}\boldsymbol{h}}\boldsymbol{h}^T\boldsymbol{R}_{XW_2X} \tag{4.20}$$

$$\boldsymbol{h}^T\boldsymbol{R}_{XW_1X} = RQ(\boldsymbol{h})\boldsymbol{h}^T\boldsymbol{R}_{XW_2X} \tag{4.21}$$

$$\boldsymbol{R}_{XW_1X}\boldsymbol{h} = \lambda\boldsymbol{R}_{XW_2X}\boldsymbol{h} \tag{4.22}$$

with λ being the eigenvalue and $\boldsymbol{h}$ the corresponding eigenvector. In Eq. 4.22 we can recognize a generalized eigenvalue problem.

Thus, by means of solving the generalized eigenvalue problem defined in Eq. 4.22, we can maximize the Rayleigh quotient and find the maximal values of the corresponding indicator and filter. In order to obtain real eigenvalues however, the correlation matrices $\boldsymbol{R}_{XW_1X}$ and $\boldsymbol{R}_{XW_2X}$ need to be Hermitian, and $\boldsymbol{R}_{XW_2X}$ needs to be positive semidefinite. If these conditions are met, the Rayleigh quotient offers an efficient means to calculate iteratively the filter coefficients. Only the largest eigenvalue and corresponding eigenvector need to be computed in each iteration, which can be achieved efficiently by using algorithms such as the power method [292].

The generalized eigenvalue problem to be solved is thus formulated as such:

$$\boldsymbol{R}_{XW_1X}\boldsymbol{h} = \boldsymbol{R}_{XW_2X}\boldsymbol{h}\lambda \tag{4.23}$$

The iterative algorithm used to minimize the prediction error consists of four basic steps:

1. Assume an initial guess for $\boldsymbol{h}$
2. Estimate $\boldsymbol{R}_{XW_1X}$ and $\boldsymbol{R}_{XW_2X}$ based on $\boldsymbol{h}$ and $\boldsymbol{X}$ using Eq. 4.12
3. Solve Eq. 4.23 to find λ_{max} and a new filter $\boldsymbol{h}$ that corresponds to a higher value of the used criterion
4. Return to step 2 using the new $\boldsymbol{h}$ until convergence is reached or the maximum number of iterations

The name of the proposed method is abbreviated to LPE (Linear Prediction of Envelope) filtering in the rest of the chapter.

4.2.3 Sparsity-based blind filters

Now that we defined a filter based on the prediction error of the squared envelope, the scope of the proposed blind filtering methodology is extended to the use of sparsity metrics on the envelope spectrum. Section 4.2.3.1 first looks at potential indicators of sparsity that can be used for the optimization. Afterward, sections 4.2.3.2 to 4.2.3.5 detail the derivation of the suggested sparse envelope spectrum filters.

4.2.3.1 Choice of sparsity measure

Ever since sparsity became a popular research interest, in large part thanks to the rise of compressed sensing, many sparsity measures have been introduced. Each measure typically has its pros and cons. A selection is made in this chapter of four different metrics that have been used in the past to either directly quantify sparsity or are closely related to the sparsity concept:

1. $\frac{l_2}{l_1}$-norm
2. Spectral negentropy
3. Hoyer Index
4. Spectral Flatness

A brief background description is provided for these four metrics.

$\frac{l_2}{l_1}$-norm A widely used sparsity measure is the $\frac{l_2}{l_1}$-norm, which belongs to the family of $\frac{l_p}{l_q}$-norms. More than a decade ago, the inverse $\frac{l_1}{l_2}$-norm (more colloquially known as the Taxicab-Euclidean norm ratio [293]) became much used as a sparsity measure in Non-negative Matrix Factorization [294–298]. In fact, it was already being used as a tool for deconvolution in geophysics [279]. This type of norm has recently been used as well in the context of vibration analysis for deconvolution of time-domain waveforms [299]. The $\frac{l_2}{l_1}$-norm of a vector $\boldsymbol{x}$ is defined as:

$$\frac{l_2}{l_1} = \frac{\sqrt{\sum_{n=1}^{N}|E_x(n)|^2}}{\sum_{n=1}^{N}|E_x(n)|} \tag{4.24}$$

Hoyer Index Hurley et al. [297] made a comprehensive comparison of fifteen commonly used sparsity measures. They evaluated these measures using six criteria a desirable measure should possess as an attribute. The Gini Index came out on top, satisfying all six criteria, with the Hoyer Index a close second, satisfying five of the criteria. However, the calculation of the Gini Index requires sorting the values which is difficult to incorporate in the derivation of the blind filter as will become clear later on. Thus, this chapter opts to include the Hoyer Index as the third sparsity measure of choice for the envelope spectrum. The Hoyer Index was proposed by P. Hoyer [298] in 2004 and is essentially a normalized version of the $\frac{l_2}{l_1}$-norm. The Hoyer Index evaluates to unity if and only if the spectrum contains only a single non-zero component, and takes a value of zero if and only if all spectral components are equal. It is defined as :

$$Hoyer\ Index = \left(\sqrt{N} - \frac{\sum_{n=1}^{N}|x(n)|}{\sqrt{\sum_{n=1}^{N}|x(n)|^2}}\right)\frac{1}{\sqrt{N}-1} \tag{4.25}$$

with x being the sample vector, n the sample number, and N the total number of samples.

Spectral negentropy Another interesting measure to characterize the squared envelope spectrum is the spectral negentropy in the spectral domain. Spectral negentropy proceeds from interpreting the (square of the) instantaneous energy flow in a signal as a probability distribution [276]. Therefore, the definition of negentropy takes into consideration the normalization of the energy flow ϵ_x. The spectral negentropy is the negative of spectral entropy and is defined in the time domain as:

$$\Delta I_\epsilon = -H_\epsilon = \left\langle \frac{\epsilon_x(n)^2}{\langle \epsilon_x(n)^2 \rangle} ln\left(\frac{\epsilon_x(n)^2}{\langle \epsilon_x(n)^2 \rangle} \right) \right\rangle \tag{4.26}$$

In the spectral domain this becomes:

$$\Delta I_E = -H_E = \left\langle \frac{|E_x(\alpha)|^2}{\langle |E_x(\alpha)|^2 \rangle} ln\left(\frac{|E_x(\alpha)|^2}{\langle |E_x(\alpha)|^2 \rangle} \right) \right\rangle \tag{4.27}$$

The negentropy in the time domain quantifies impulsive signal content, much like the kurtogram (Eq. 4.26 is actually a weighted version of kurtosis). The negentropy in the spectral domain on the other hand quantifies the repetitive or cyclostationary signal content. It is possible to combine the two to obtain the averaged spectral negentropy. The weighted average of the spectral negentropy in the two domains reaches its upper bound when we just take the normal mean of the two quantities. The average spectral negentropy is defined as:

$$\Delta I_{1/2} = \frac{1}{2}\Delta I_\epsilon + \frac{1}{2}\Delta I_E \tag{4.28}$$

Spectral flatness One of the measures to characterize a spectrum is the spectral flatness or tonality coefficient. Spectral flatness provides a way to quantify how noise-like a spectrum is. The maximum spectral flatness of 1 indicates that you have a pure white noise spectrum, while a low spectral flatness (near zero) indicates a 'pure tone' which shows up as a discrete peak in the spectrum. This concept can be related to the squared envelope spectrum since fault modulations are expected to show up as discrete peaks in the squared envelope spectrum (SES). Thus minimizing the spectral flatness, or maximizing the

inverse, corresponds to finding the filter that produces the most 'pure tone'-like SES. The spectral flatness (SF) of the SES is calculated by dividing the geometric mean by the arithmetic mean of the SES and is given by:

$$SF = \frac{\sqrt[N]{\prod_{n=0}^{N-1}|E_x(n)|}}{\frac{1}{N}\sum_{n=0}^{N-1}|E_x(n)|} = \frac{e^{\frac{1}{N}\sum_{n=0}^{N-1} ln(|E_x(n)|)}}{\frac{1}{N}\sum_{n=0}^{N-1}|E_x(n)|} \tag{4.29}$$

4.2.3.2 Derivation of filter using $\frac{l_2}{l_1}$-norm

The $\frac{l_2}{l_1}$-norm of the envelope spectrum $\boldsymbol{E_x}$ is defined as:

$$\frac{l_2}{l_1} = \frac{\sqrt{\sum_{n=1}^{N}|E_x(n)|^2}}{\sum_{n=1}^{N}|E_x(n)|} \tag{4.30}$$

Rewriting using vector notation, gives:

$$\frac{l_2}{l_1} = \frac{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}}{\boldsymbol{E_x}^H diag(\frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x}} \tag{4.31}$$

The denominator is obtained using following identity:

$$\sum_{n=1}^{N}|E_x(n)| = \sum_{n=1}^{N}\frac{|E_x(n)|^2}{|E_x(n)|} = \boldsymbol{E_x}^H diag(\frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x} \tag{4.32}$$

Manipulating the numerator of Eq. 4.31 further, it follows that:

$$\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}} = \frac{\boldsymbol{E_x}^H\boldsymbol{E_x}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}} = \boldsymbol{E_x}^H diag(\frac{1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{E_x} \tag{4.33}$$

Eq. 4.31 can thus be rewritten in following form:

$$\frac{l_2}{l_1} = \frac{\boldsymbol{E_x}^H diag(\frac{1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{E_x}}{\boldsymbol{E_x}^H diag(\frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x}} \tag{4.34}$$

Inserting Eq.4.5 in Eq. 4.34 allows to write the ratio in full:

$$\frac{l_2}{l_1} = \frac{\boldsymbol{h^H X^H} diag(\boldsymbol{s}) \boldsymbol{F} diag(\frac{1}{\sqrt{\boldsymbol{E_x}^H \boldsymbol{E_x}}}) \boldsymbol{F^H} diag(\boldsymbol{s^H}) \boldsymbol{Xh}}{\boldsymbol{h^H X^H} diag(\boldsymbol{s}) \boldsymbol{F} diag(\frac{1}{|\boldsymbol{E_x}|}) \boldsymbol{F^H} diag(\boldsymbol{s^H}) \boldsymbol{Xh}} \tag{4.35}$$

The generalized Rayleigh quotient [291] can be recognized in Eq. 4.35. Thus the $\frac{l_2}{l_1}$-norm is rewritten as follows:

$$\frac{l_2}{l_1} = \frac{\boldsymbol{h^H R}_{XW_1X} \boldsymbol{h}}{\boldsymbol{h^H R}_{XW_2X} \boldsymbol{h}} \tag{4.36}$$

with:

$$\boldsymbol{R}_{XW_1X} = \boldsymbol{X^H} diag(\boldsymbol{s}) \boldsymbol{F} diag(\frac{1}{\sqrt{\boldsymbol{E_x}^H \boldsymbol{E_x}}}) \boldsymbol{F^H} diag(\boldsymbol{s^H}) \boldsymbol{X} \tag{4.37}$$

$$\boldsymbol{R}_{XW_2X} = \boldsymbol{X^H} diag(\boldsymbol{s}) \boldsymbol{F} diag(\frac{1}{|\boldsymbol{E_x}|}) \boldsymbol{F^H} diag(\boldsymbol{s^H}) \boldsymbol{X} \tag{4.38}$$

4.2.3.3 Derivation of filter using Hoyer Index

Since the Hoyer Index is essentially a normalized version of the $\frac{l_2}{l_1}$-norm, the derivation is very similar. The Hoyer index of the squared envelope spectrum $\boldsymbol{E_x}$ is defined as:

$$Hoyer\ Index = (\sqrt{N} - \frac{\sum_{n=1}^{N} |E_x(n)|}{\sqrt{\sum_{n=1}^{N} |E_x(n)|^2}}) \frac{1}{\sqrt{N} - 1} \tag{4.39}$$

Rewriting Eq. 4.39 using vector notation results in:

$$Hoyer\ Index = (\sqrt{N} - \frac{\boldsymbol{E_x}^H diag(\frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\frac{1}{\sqrt{N}-1} \tag{4.40}$$

$$= \frac{\sqrt{N}\frac{\boldsymbol{E_x}^H\boldsymbol{E_x}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}} - \boldsymbol{E_x}^H diag(\frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x}}{(\sqrt{N}-1)\frac{\boldsymbol{E_x}^H\boldsymbol{E_x}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}}} \tag{4.41}$$

$$= \frac{\boldsymbol{E_x}^H diag(\frac{\sqrt{N}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{E_x} - \boldsymbol{E_x}^H diag(\frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x}}{\boldsymbol{E_x}^H diag(\frac{\sqrt{N}-1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{E_x}} \tag{4.42}$$

$$= \frac{\boldsymbol{E_x}^H diag(\frac{\sqrt{N}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}} - \frac{1}{|\boldsymbol{E_x}|})\boldsymbol{E_x}}{\boldsymbol{E_x}^H diag(\frac{\sqrt{N}-1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{E_x}} \tag{4.43}$$

After some more rewriting, the generalized Rayleigh quotient emerges again:

$$Hoyer\ Index = \frac{\boldsymbol{h^H X^H} diag(\boldsymbol{s})\boldsymbol{F} diag(\frac{\sqrt{N}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}} - \frac{1}{|\boldsymbol{E_x}|})\boldsymbol{F^H} diag(\boldsymbol{s^H})\boldsymbol{Xh}}{\boldsymbol{h^H X^H} diag(\boldsymbol{s})\boldsymbol{F} diag(\frac{\sqrt{N}-1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{F^H} diag(\boldsymbol{s^H})\boldsymbol{Xh}} \tag{4.44}$$

So for the Hoyer Index we have the following end formula:

$$Hoyer\ Index = \frac{\boldsymbol{h^H R}_{XW_1X}\boldsymbol{h}}{\boldsymbol{h^H R}_{XW_2X}\boldsymbol{h}} \tag{4.45}$$

with:

$$\boldsymbol{R}_{XW_1X} = \boldsymbol{X^H} diag(\boldsymbol{s})\boldsymbol{F} diag(\frac{\sqrt{N}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}} - \frac{1}{|\boldsymbol{E_x}|})\boldsymbol{F^H} diag(\boldsymbol{s^H})\boldsymbol{X} \tag{4.46}$$

$$\boldsymbol{R}_{XW_2X} = \boldsymbol{X^H} diag(\boldsymbol{s})\boldsymbol{F} diag(\frac{\sqrt{N}-1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{F^H} diag(\boldsymbol{s^H})\boldsymbol{X} \tag{4.47}$$

4.2.3.4 Derivation of filter using spectral negentropy

The spectral negentropy ΔI_E is the negative of spectral entropy and is defined for the envelope spectrum as:

$$\Delta I_E = -H_E = \left\langle \frac{|E_x(\alpha)|^2}{\langle |E_x(\alpha)|^2 \rangle} ln\left(\frac{|E_x(\alpha)|^2}{\langle |E_x(\alpha)|^2 \rangle} \right) \right\rangle \tag{4.48}$$

Rewriting again using vector notation:

$$\Delta I_E = \frac{N}{\boldsymbol{E_x}^H \boldsymbol{E_x}} \left\langle |E_x(\alpha)|^2 ln\left(\frac{|E_x(\alpha)|^2}{\langle |E_x(\alpha)|^2 \rangle} \right) \right\rangle \tag{4.49}$$

$$= \frac{\boldsymbol{E_x}^H diag(ln(\frac{|\boldsymbol{E_x}|^2}{\langle |\boldsymbol{E_x}|^2 \rangle})) \boldsymbol{E_x}}{\boldsymbol{E_x}^H \boldsymbol{E_x}} \tag{4.50}$$

This can again be written as a generalized Rayleigh quotient.

$$\Delta I_E = \frac{\boldsymbol{h^H X^H} diag(\boldsymbol{s}) \boldsymbol{F} diag(ln(\frac{|\boldsymbol{E_x}|^2}{\langle |\boldsymbol{E_x}|^2 \rangle})) \boldsymbol{F^H} diag(\boldsymbol{s^H}) \boldsymbol{X h}}{\boldsymbol{h^H X^H} diag(\boldsymbol{s}) \boldsymbol{F F^H} diag(\boldsymbol{s^H}) \boldsymbol{X h}} \tag{4.51}$$

So for the spectral negentropy of the envelope spectrum we get:

$$\Delta I_E = \frac{\boldsymbol{h^H R}_{XW_1X} \boldsymbol{h}}{\boldsymbol{h^H R}_{XW_2X} \boldsymbol{h}} \tag{4.52}$$

with:

$$\boldsymbol{R}_{XW_1X} = \boldsymbol{X^H} diag(\boldsymbol{s}) \boldsymbol{F} diag(ln(\frac{|\boldsymbol{E_x}|^2}{\langle |\boldsymbol{E_x}|^2 \rangle})) \boldsymbol{F^H} diag(\boldsymbol{s^H}) \boldsymbol{X} \tag{4.53}$$

$$\boldsymbol{R}_{XW_2X} = \boldsymbol{X^H} diag(\boldsymbol{s}) \boldsymbol{F F^H} diag(\boldsymbol{s^H}) \boldsymbol{X} \tag{4.54}$$

The spectral negentropy also has the useful property that it is always greater than zero and smaller than $ln(N)$:

$$0 \leq \Delta I_E \leq ln(N) \tag{4.55}$$

The same is valid for the negentropy in the time domain ΔI_e. This is proven in Appendix A 4.6.

For the sake of completeness, also the derivation of the blind filter formulatuion for the spectral negentropy of the squared envelope and the averaged spectral negentropy is given.

Spectral negentropy of squared envelope The formula for the squared envelope spectral negentropy is identical almost to the one of the envelope spectrum and is given by:

$$\Delta I_\epsilon = \frac{\boldsymbol{\epsilon_x}^H diag(ln(\frac{\boldsymbol{\epsilon_x}^2}{\langle \boldsymbol{\epsilon_x}^2 \rangle}))\boldsymbol{\epsilon_x}}{\boldsymbol{\epsilon_x}^H \boldsymbol{\epsilon_x}} \tag{4.56}$$

Since there is no need for a Fourier transform matrix here compared to the envelope spectrum, the formula for the Rayleigh quotient here is consequently slightly simpler :

$$\Delta I_\epsilon = \frac{\boldsymbol{h^H X^H} diag(\boldsymbol{s}) diag(ln(\frac{\boldsymbol{\epsilon_x}^2}{\langle \boldsymbol{\epsilon_x}^2 \rangle})) diag(\boldsymbol{s^H}) \boldsymbol{Xh}}{\boldsymbol{h^H X^H} diag(\boldsymbol{s}) diag(\boldsymbol{s^H}) \boldsymbol{Xh}} \tag{4.57}$$

So for the spectral negentropy of the squared envelope we get:

$$\Delta I_\epsilon = \frac{\boldsymbol{h^H R}_{XW_1X} \boldsymbol{h}}{\boldsymbol{h^H R}_{XW_2X} \boldsymbol{h}} \tag{4.58}$$

with:

$$\boldsymbol{R}_{XW_1X} = \boldsymbol{X^H} diag(\boldsymbol{s}) diag(ln(\frac{\boldsymbol{\epsilon_x}^2}{\langle \boldsymbol{\epsilon_x}^2 \rangle})) diag(\boldsymbol{s^H}) \boldsymbol{X} \tag{4.59}$$

$$\boldsymbol{R}_{XW_2X} = \boldsymbol{X^H} diag(\boldsymbol{s}) diag(\boldsymbol{s^H}) \boldsymbol{X} \tag{4.60}$$

The maximization of this criterion however does not guarantee a an informative envelope spectrum since it is a weighted version of kurtosis. Therefore, by itself, it might not be ideal to analyze non-impulsive CS2 signal content.

Average spectral negentropy The weighted average of the spectral negentropy in the two domains reaches its upper bound when we just

take the normal mean of the two quantities. The average spectral negentropy is thus defined as:

$$\Delta I_{1/2} = \frac{1}{2}\Delta I_\epsilon + \frac{1}{2}\Delta I_E \tag{4.61}$$

Replacing ΔI_ϵ and ΔI_E with their Rayleigh quotients (where we have used the symbol $\boldsymbol{W}$ to represent the weighting matrices), we get:

$$\Delta I_{1/2} = \frac{1}{2}\left(\frac{\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{1\epsilon}\boldsymbol{X}\boldsymbol{h}}{\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2\epsilon}\boldsymbol{X}\boldsymbol{h}} + \frac{\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{1E}\boldsymbol{X}\boldsymbol{h}}{\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2E}\boldsymbol{X}\boldsymbol{h}}\right) \tag{4.62}$$

Some restructuring gives:

$$\Delta I_{1/2} = \frac{1}{2}\left(\frac{\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{1\epsilon}\boldsymbol{X}\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2E}\boldsymbol{X}\boldsymbol{h} + \boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{1E}\boldsymbol{X}\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2\epsilon}\boldsymbol{X}\boldsymbol{h}}{\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2\epsilon}\boldsymbol{X}\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2E}\boldsymbol{X}\boldsymbol{h}}\right) \tag{4.63}$$

$$= \frac{1}{2}\left(\frac{\boldsymbol{h}^H\boldsymbol{X}^H\left[\boldsymbol{W}_{1\epsilon}\boldsymbol{X}\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2E} + \boldsymbol{W}_{1E}\boldsymbol{X}\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2\epsilon}\right]\boldsymbol{X}\boldsymbol{h}}{\boldsymbol{h}^H\boldsymbol{X}^H\left[\boldsymbol{W}_{2\epsilon}\boldsymbol{X}\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{X}^H\boldsymbol{W}_{2E}\right]\boldsymbol{X}\boldsymbol{h}}\right) \tag{4.64}$$

$$= \frac{\boldsymbol{h}^H\boldsymbol{R}_{X\boldsymbol{W}_{1,1/2}X}\boldsymbol{h}}{\boldsymbol{h}^H\boldsymbol{R}_{X\boldsymbol{W}_{2,1/2}X}\boldsymbol{h}} \tag{4.65}$$

This equation can also be expressed in terms of the weighted correlation matrices of the spectral negentropy in the two domains:

$$\Delta I_{1/2} = \frac{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{1\epsilon}X}\boldsymbol{h}}{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2\epsilon}X}\boldsymbol{h}} + \frac{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{1E}X}\boldsymbol{h}}{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2E}X}\boldsymbol{h}} \tag{4.66}$$

$$= \frac{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{1\epsilon}X}\boldsymbol{h}\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2E}X}\boldsymbol{h} + \boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{1E}X}\boldsymbol{h}\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2\epsilon}X}\boldsymbol{h}}{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2\epsilon}X}\boldsymbol{h}\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2E}X}\boldsymbol{h}} \tag{4.67}$$

$$= \frac{\boldsymbol{h}^{\boldsymbol{H}}\left[\boldsymbol{R}_{XW_{1\epsilon}X}\boldsymbol{h}\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2E}X} + \boldsymbol{R}_{XW_{1E}X}\boldsymbol{h}\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2\epsilon}X}\right]\boldsymbol{h}}{\boldsymbol{h}^{\boldsymbol{H}}\left[\boldsymbol{R}_{XW_{2\epsilon}X}\boldsymbol{h}\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2E}X}\right]\boldsymbol{h}} \tag{4.68}$$

$$= \frac{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{1,1/2}X}\boldsymbol{h}}{\boldsymbol{h}^{\boldsymbol{H}}\boldsymbol{R}_{XW_{2,1/2}X}\boldsymbol{h}} \tag{4.69}$$

The average spectral negentropy was not used in further analyses since it was found to have a bad convergence behavior.

4.2.3.5 Derivation of filter using spectral flatness

The spectral flatness of the squared envelope spectrum (SES) is calculated by dividing the geometric mean by the arithmetic mean of the SES and is given by:

$$SF = \frac{\sqrt[N]{\prod_{n=0}^{N-1}|E_x(n)|}}{\frac{1}{N}\sum_{n=0}^{N-1}|E_x(n)|} = \frac{e^{\frac{1}{N}\sum_{n=0}^{N-1} ln(|E_x(n)|)}}{\frac{1}{N}\sum_{n=0}^{N-1}|E_x(n)|} \tag{4.70}$$

Rewriting this using matrix notation gives for the numerator:

$$e^{\frac{1}{N}\sum_{n=0}^{N-1} ln(|E_x(n)|)} = \frac{|E_x(n)|^2 e^{\frac{1}{N}\sum_{n=0}^{N-1} ln(|E_x(n)|)}}{|E_x(n)|^2} \tag{4.71}$$

$$= E_x(n)^H diag(\frac{e^{\frac{1}{N}\sum_{n=0}^{N-1} ln(|E_x(n)|)}}{|E_x(n)|^2})E_x(n) \tag{4.72}$$

$$= \mathbf{E_x}^H diag(\frac{e^{\langle ln(|\mathbf{E_x}|)\rangle}}{|\mathbf{E_x}|^2})\mathbf{E_x} \tag{4.73}$$

and for the denominator:

$$\frac{1}{N}\sum_{n=0}^{N-1}|E_x(n)| = E_x(n)^H diag(\frac{|E_x(n)|}{N|E_x(n)|^2}E_x(n) \tag{4.74}$$

$$= \mathbf{E_x}^H diag(\frac{|\mathbf{E_x}|}{N})\mathbf{E_x} \tag{4.75}$$

The spectral flatness in matrix notation is thus given by:

$$SF = \frac{\mathbf{E_x}^H diag(\frac{e^{\langle ln(|\mathbf{E_x}|)\rangle}}{|\mathbf{E_x}|^2})\mathbf{E_x}}{\mathbf{E_x}^H diag(\frac{|\mathbf{E_x}|}{N})\mathbf{E_x}} \tag{4.76}$$

which can be written as a Rayleigh quotient:

$$SF = \frac{\mathbf{h^H X^H} diag(\mathbf{s})\mathbf{F} diag(\frac{e^{\langle ln(|\mathbf{E_x}|)\rangle}}{|\mathbf{E_x}|^2})\mathbf{F^H} diag(\mathbf{s^H})\mathbf{Xh}}{\mathbf{h^H X^H} diag(\mathbf{s})\mathbf{F} diag(\frac{|\mathbf{E_x}|}{N})\mathbf{F^H} diag(\mathbf{s^H})\mathbf{Xh}} = \frac{\mathbf{h^H R}_{XW_1X}\mathbf{h}}{\mathbf{h^H R}_{XW_2X}\mathbf{h}} \tag{4.77}$$

with:

$$\boldsymbol{R}_{XW_1X} = \mathbf{X^H} diag(\mathbf{s})\mathbf{F} diag(\frac{e^{\langle ln(|\mathbf{E_x}|)\rangle}}{|\mathbf{E_x}|^2})\mathbf{F^H} diag(\mathbf{s^H})\mathbf{X} \tag{4.78}$$

$$\boldsymbol{R}_{XW_2X} = \mathbf{X^H} diag(\mathbf{s})\mathbf{F} diag(\frac{|\mathbf{E_x}|}{N})\mathbf{F^H} diag(\mathbf{s^H})\mathbf{X} \tag{4.79}$$

The actual calculation is done by computing the largest eigenvalue of the inverse of the spectral flatness since this corresponds to finding the most sparse SES.

4.2.4 Overview of derived Rayleigh quotient filters

In total, six different blind filters are further investigated. Table 4.1 presents a concise summary of the weights of the Rayleigh quotient correlation matrices of these blind filters as to provide a quick means for the reader to see the mathematical differences between the filter formulations.

TABLE 4.1: Overview of the Rayleigh quotient weights of the different blind filters.

Indicator	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$
Envelope prediction error	1	$\boldsymbol{A}^H\boldsymbol{A}$
$\frac{l_2}{l_1}$-norm	$\boldsymbol{F}diag(\frac{1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{F^H}$	$\boldsymbol{F}diag(\frac{1}{\lvert\boldsymbol{E_x}\rvert})\boldsymbol{F^H}$
Hoyer Index	$\boldsymbol{F}diag(\frac{\sqrt{N}}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}}-\frac{1}{\lvert\boldsymbol{E_x}\rvert})\boldsymbol{F^H}$	$\boldsymbol{F}diag(\frac{\sqrt{N}-1}{\sqrt{\boldsymbol{E_x}^H\boldsymbol{E_x}}})\boldsymbol{F^H}$
Spectral negentropy ΔI_E	$\boldsymbol{F}diag(ln(\frac{\lvert\boldsymbol{E_x}\rvert^2}{\langle\lvert\boldsymbol{E_x}\rvert^2\rangle}))\boldsymbol{F^H}$	$\boldsymbol{FF^H}$
Spectral negentropy ΔI_ϵ	$diag(ln(\frac{\boldsymbol{\epsilon_x}^2}{\langle\boldsymbol{\epsilon_x}^2\rangle}))$	1
Spectral flatness	$\mathbf{F}diag(\frac{e^{\langle ln(\lvert\mathbf{E_x}\rvert)\rangle}}{\lvert\mathbf{E_x}\rvert^2})\mathbf{F^H}$	$\mathbf{F}diag(\frac{\lvert\mathbf{E_x}\rvert}{N})\mathbf{F^H}$

These weights are to be used in the expression of the Rayleigh quotient to find the corresponding indicator as follows:

$$RQ(\boldsymbol{h}) = \frac{\boldsymbol{h^H X^H}diag(\mathbf{s})\boldsymbol{W_1}diag(\mathbf{s^H})\boldsymbol{Xh}}{\boldsymbol{h^H X^H}diag(\mathbf{s})\boldsymbol{W_2}diag(\mathbf{s^H})\boldsymbol{Xh}} \tag{4.80}$$

4.2.5 Practical considerations

The maximization of the sparsity of the envelope spectrum and the minimization of the envelope prediction error is based on the assumption that the component responsible for increasing the sparsity or predictability, is the potential fault signature. Therefore, it is recommended to first prewhiten the signal such that the deterministic components are removed before solving the maximization problem since these components often introduce high amplitude discrete peaks in the envelope spectrum and can cause erroneous convergence.

The calculation of correlation matrices $\boldsymbol{R}_{XW_1X}$ and $\boldsymbol{R}_{XW_2X}$ includes for some filters a Fourier matrix to calculate the envelope spectrum. However, in the actual calculation this can be replaced by an FFT to decrease the computation time and the memory requirements.

The total duration of the computations will depend mainly on the signal length and the chosen filter length. Since the matrices can become quite large for long signals and filters, the number of multiplications and FFTs that need to be calculated can lead to long computation times. Currently, the filter length is still chosen arbitrarily by visual inspection of the results, but future research will focus on automating this filter length choice.

4.2.6 Extension for non-stationary angular speed cases

A potential extension of the envelope spectrum-based methodology is to make the kernel of the Fourier matrix in the filter correlation matrices dependent on angle. This way, the instantaneous angular speed variations can be taken into account directly in the filter. This extension is not further investigated in this dissertation but it can be easily achieved by means of the Velocity Synchronous Discrete Fourier Transform (VSDFT) [300]. The Fourier matrix then needs to be reformulated as follows:

$$\boldsymbol{F} = \frac{1}{\Theta}\begin{bmatrix} \dot{\theta}_0 & \dot{\theta}_0 & \dots & \dot{\theta}_0 & \dots & \dot{\theta}_0 \\ \dot{\theta}_1 & \dot{\theta}_1 e^{-j\theta_1\Omega_1} & \dots & \dot{\theta}_1 e^{-j\theta_1\Omega_k} & \dots & \dot{\theta}_1 e^{-j\theta_1\Omega_{K-1}} \\ \vdots & \vdots & \ddots & & & \vdots \\ \dot{\theta}_n & \dot{\theta}_n e^{-j\theta_n\Omega_1} & \dots & \dot{\theta}_n e^{-j\theta_n\Omega_k} & \dots & \dot{\theta}_n e^{-j\theta_n\Omega_{K-1}} \\ \vdots & \vdots & & & \ddots & \vdots \\ \dot{\theta}_{L-N-1} & \dot{\theta}_{L-N-1} e^{-j\theta_{L-N-1}\Omega_1} & \dots & \dot{\theta}_{L-N-1} e^{-j\theta_{L-N-1}\Omega_k} & \dots & \dot{\theta}_{L-N-1} e^{-j\theta_{L-N-1}\Omega_{K-1}} \end{bmatrix} \tag{4.81}$$

with $\theta_n = \Sigma_{n=1}^{L}\dot{\theta}(n)$, $\Omega(k)$ the vector of orders for the representation of the order spectrum, and the normalization factor $\Theta = \Sigma_{n=1}^{L}\dot{\theta}(n)$.

An alternative approach would be to resample the signal to the angular domain before passing it to the filter optimization.

4.2.7 Extension for single-input multiple-output (SIMO) systems

Similar to the description in [288], the indicator calculations described here can also be extended for the case when there is more than one acquisition channel that measures the vibration response. The proposed approach allows a fairly straightforward extension to the case where one has Q responses $\boldsymbol{x}_q$. Each response $\boldsymbol{x}_q$ that is filtered by $\boldsymbol{h}_q$,

can be summed together to return an estimation of the signal of interest $\boldsymbol{s}$. The iterative procedure using the Rayleigh quotient can still be used for the SIMO case. The only adaptation necessary is of the correlation matrices and the filter vector.

$\boldsymbol{R}_{XW_1X}$ and $\boldsymbol{R}_{XW_2X}$ need to be expressed as cross-correlation matrices:

$$\boldsymbol{R}_{XWX} = \begin{bmatrix} \ddots & & \boldsymbol{R}_{qW\tilde{q}} \\ & \boldsymbol{R}_{qWq} & \\ \boldsymbol{R}_{\tilde{q}Wq} & & \ddots \end{bmatrix} \tag{4.82}$$

with $\boldsymbol{R}_{qWq}$ the weighted auto-correlation matrix of $\boldsymbol{x}_q$ and the off-diagonal matrices are the weighted cross-correlation matrices of $\boldsymbol{x}_q$ and $\boldsymbol{x}_{\tilde{q}}$. The filter $\boldsymbol{h}$ then becomes:

$$\boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_q \\ \vdots \\ \boldsymbol{h}_Q \end{bmatrix} \tag{4.83}$$

The q^{th} contribution to the signal of interest $\boldsymbol{s}$ is then calculated as:

$$\boldsymbol{s}_q = \boldsymbol{X}_q \boldsymbol{h}_q \tag{4.84}$$

with $\boldsymbol{X}_q$ the Toeplitz matrix of $\boldsymbol{x}_q$ as defined in Eq. 4.2. Thus, the overall signal of interest $\boldsymbol{s}$ is then found by:

$$\boldsymbol{s} = \Sigma_{q=1}^{Q} \boldsymbol{s}_q \tag{4.85}$$

4.3 Simulation analysis

To validate and illustrate the proposed approach, two simulated cases are first considered. To add some point of reference, the performance

of the proposed approach with the different sparsity indicators is compared to time-domain blind deconvolution filtering based on Minimum Entropy Deconvolution (MED), ICS2, and the time-domain spectral negentropy. Two cases with cyclostationary source signals are examined:

- Case 1: Periodic impulses with Gaussian distributed amplitudes
- Case 2: Periodic impulses with Gaussian distributed amplitudes and Gaussian jitter on the impulse period

These two cases are the same first two cases as in [288] to provide additional reference for comparison. The periodic impulse signal s_1 is convolved with the impulse response function (IRF) $g_{1,s}$ and the Gaussian noise with $g_{1,n}$. The generated signals are shown separately in Fig. 4.6. Instead of only analyzing a single simulation though, the additive Gaussian noise is varied in signal-to-noise ratio (SNR) from -80 dB to 0 dB to assess the trending behavior of the filters. The overall variance of the full signal is kept constant. For the time-domain blind filters, a filter length of 40 samples is used, while the sparsity-based spectral filters use a filter length of 20 samples. The maximum number of iterations is set to 50. Preferably the filter lengths are short to decrease computation time and improve the convergence of the filter. At the same time, the filter lengths cannot be too short since they need to be able to effectively suppress unwanted frequency content. Currently these values are thus still chosen arbitrarily, but in the future, effort should go into investigating automated approaches for choosing the filter lengths.

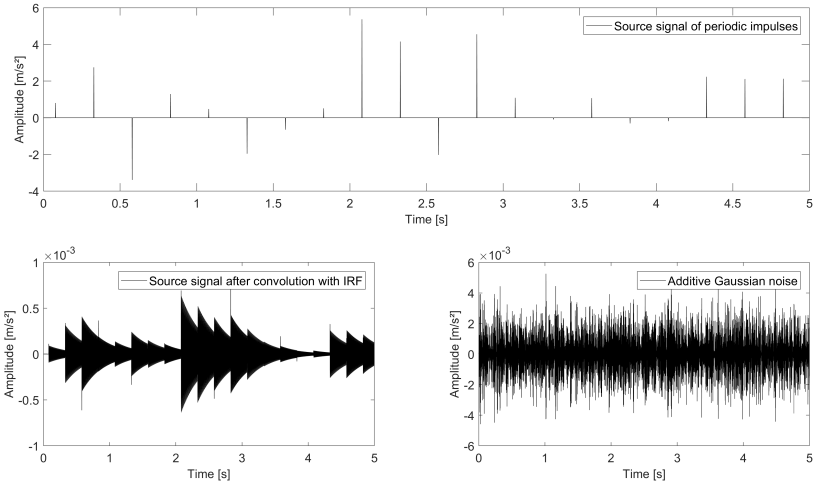

FIGURE 4.6: Input signals used for simulation case 1: (Top) Periodic impulses with Gaussian distributed amplitudes s_1 , (Bottom-left) after convolution with IRF $g_{1,s}$, (Bottom-right) Additive Gaussian noise after convolution with IRF $g_{1,n}$.

Figure 4.7 shows the resulting indicator evolutions from -80 dB to 0 dB SNR for the first case. It can be seen that in this particular case six out of eight filters show a significant increase at approximately -38 dB. Only the LP-envelope filter and the spectral flatness filter have a significant delay in feature value change.

To inspect the results further, Fig. 4.8 displays a color map of the squared envelope spectra for the first case. The envelope spectra of the time-domain filters (ICS2, MED, and time-domain negentropy) exhibit clear harmonics of the fault frequency at 4 Hz for higher SNRs. In contrast, the envelope spectra of the sparsity-based filters and the LP-envelope filter exhibit the first fault harmonic at 4 Hz but also a lower frequency component at 0.4 Hz that is higher in amplitude. This is due to the envelope spectrum of the simulation signal containing a very low frequency modulation due to the employed Gaussian distributed amplitudes of the periodic impulses. Additionally, this low frequency modulation increases the sparsity of the envelope spectrum. The only

exception is the spectral flatness filter that appears to be rather unstable with respect to the SES but does filter out the 4 Hz harmonics.

Lastly, the filtered signals of case 1 are shown in Fig. 4.9 at the SNR of -19 dB, similar to the comparison of signals shown in [288]. The amplitude of the signals has been normalized between -1 and +1. It turns out that the time-domain filters deconvolve the periodic impulse train of the input source signal, while the envelope spectrum sparsity-based filters extract the periodic impulse train after deconvolution with its IRF. This result makes sense considering that the most sparse envelope spectrum is that of a signal with a pure sinusoidal amplitude modulation. A Dirac impulse train however results in an envelope spectrum that exhibits multiple harmonics which decreases the sparsity as compared to a single discrete peak. Therefore the sparsity-based methods filter out the impulse train after convolution with the IRF which leads to a sparsity that is in-between that of a pure sinusoidal modulation and that of an impulse train. The only bad result is produced by the LP-envelope filter in this case, which is unable to filter out any meaningful signal at any SNR below -13 dB.

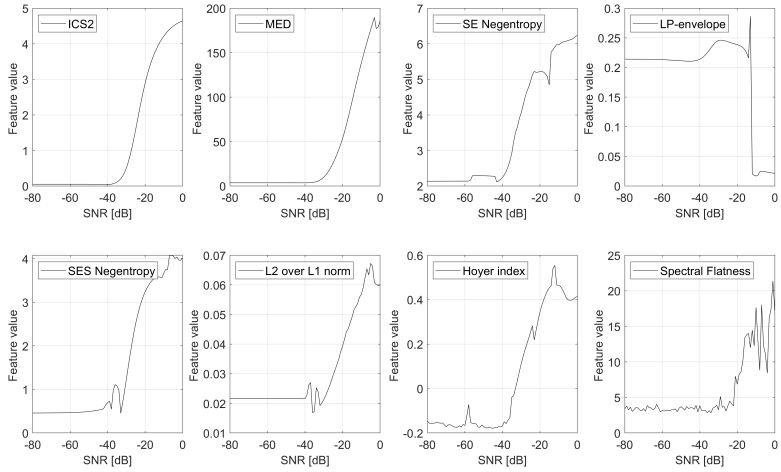

FIGURE 4.7: Trending of the different indicators for simulation case 1.

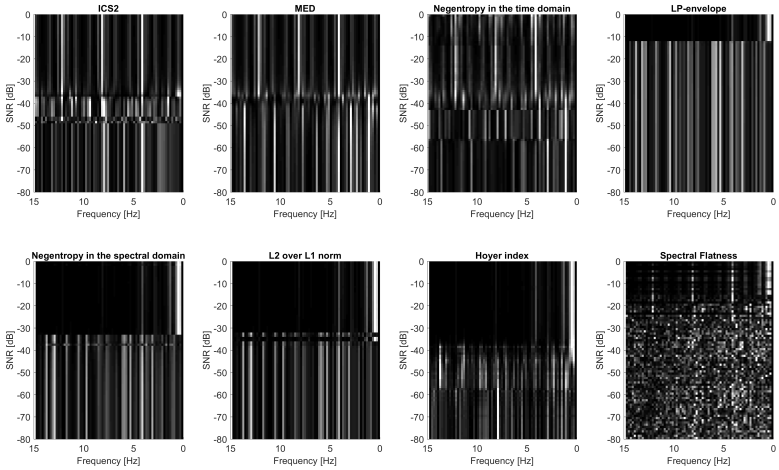

FIGURE 4.8: Trending of the normalized squared envelope spectra for simulation case 1. The fault frequency of interest is 4 Hz. (Black = 0, white = 1)

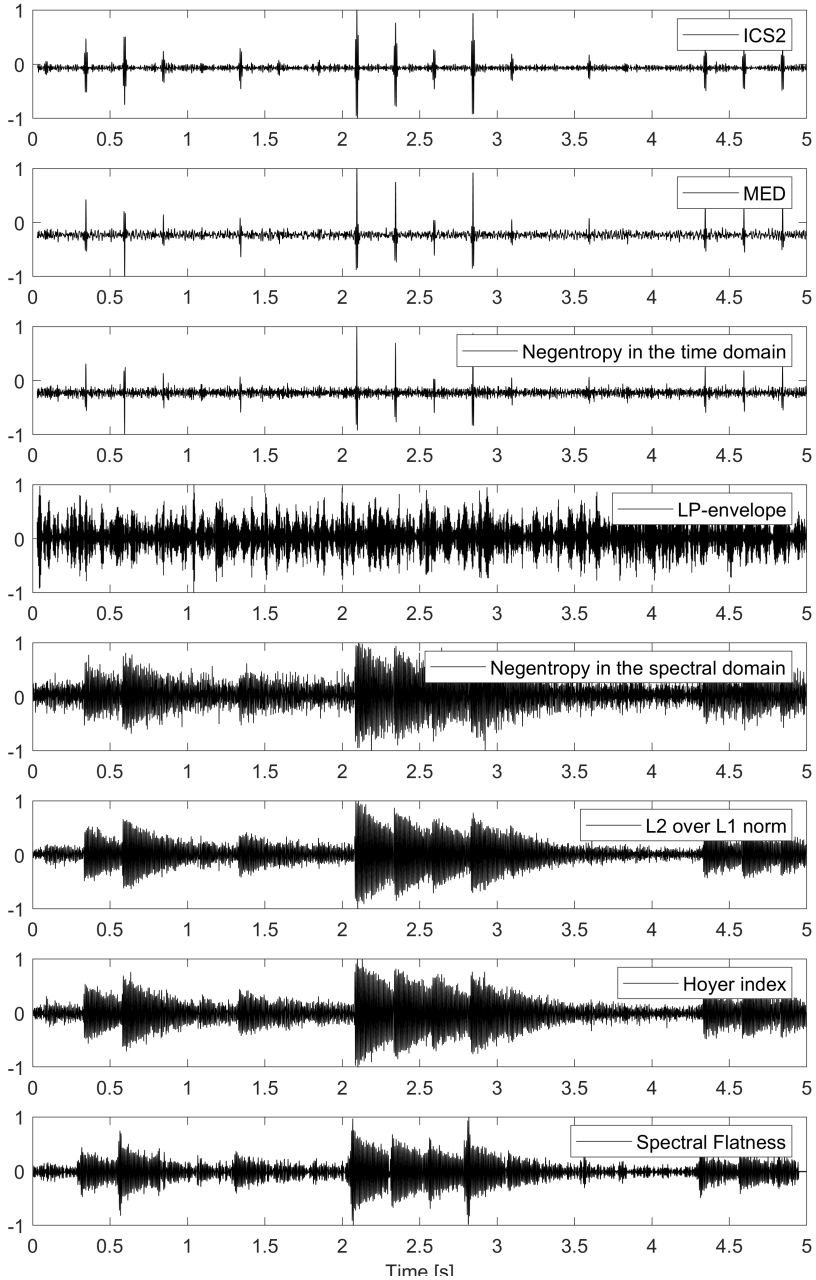

FIGURE 4.9: Time waveforms at -19 dB SNR for simulation case 1. The amplitudes have been normalized between -1 and +1.

The second example investigates the case where the fundamental fault period of the impulses fluctuate due to Gaussian jitter. The results are similar to the first case since the jitter only impacts the filtering potential of the sparsity-based filters slightly. Figures 4.10, 4.12, 4.11 display respectively the indicator trends, the envelope spectra color maps, and the filtered time waveforms at -19 dB. In case the jitter would be chosen to have a larger value, it can be expected that the jitter of the fault period would cause the discrete envelope spectrum peaks to spread out over multiple frequency bins and thus reduce the sparsity of the envelope spectrum.

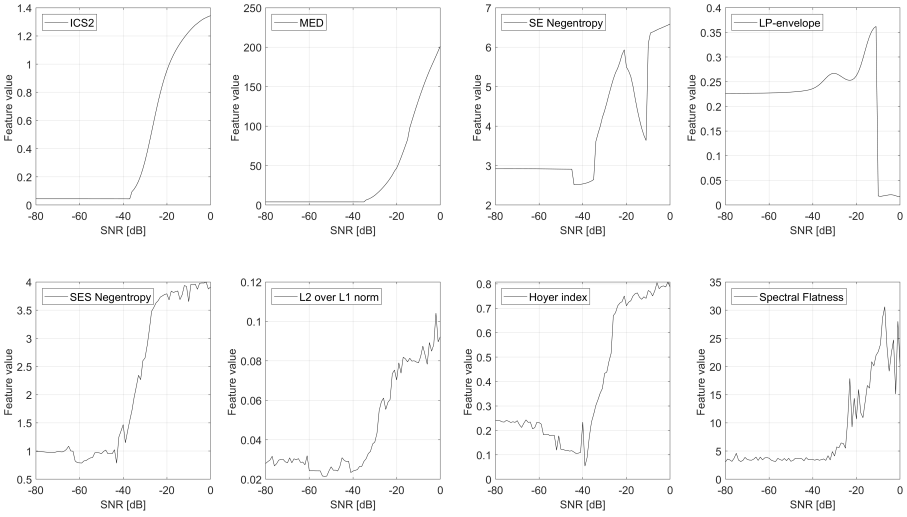

FIGURE 4.10: Trending of the different indicators for simulation case 2.

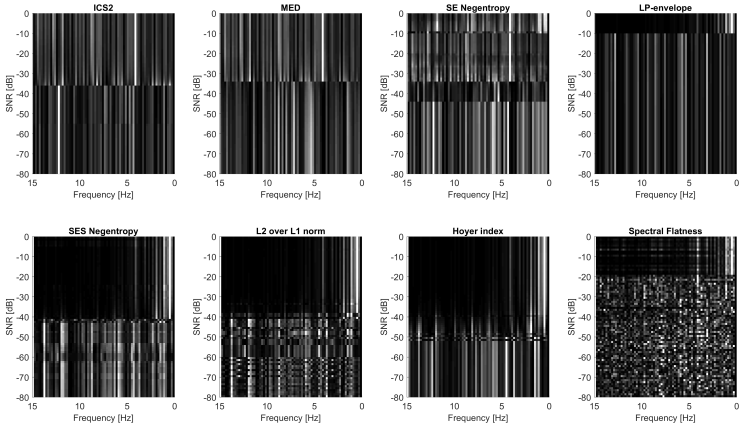

FIGURE 4.11: Trending of the squared envelope spectra for simulation case 2. The fault frequency of interest is 4 Hz. (Black = 0, white = 1)

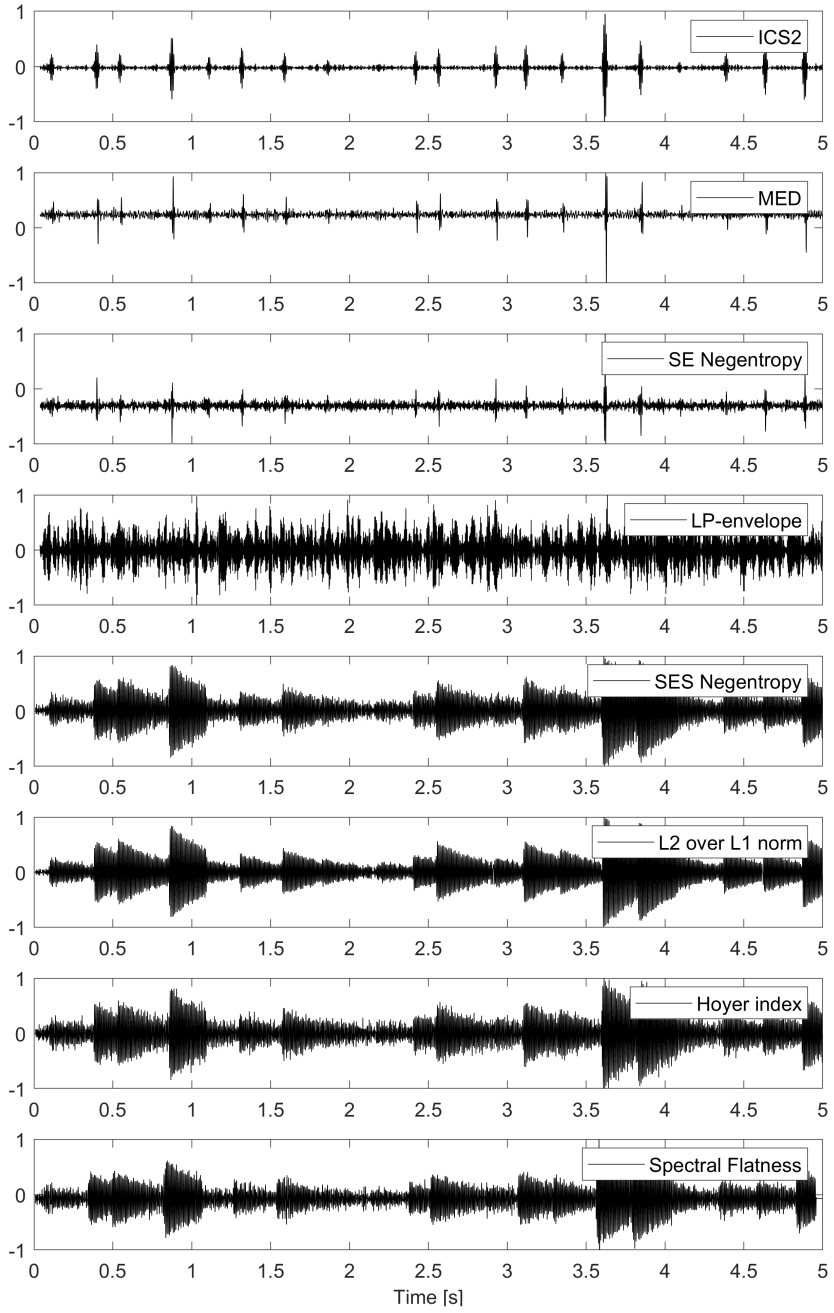

FIGURE 4.12: Time waveforms at -19 dB SNR for simulation case 2. The amplitudes have been normalized between -1 and +1.

4.4 Experimental analysis & results

The proposed approach is applied on vibration signals from the Prognostics Data Repository of NASA as provided by the Center for Intelligent Maintenance Systems (IMS) of the University of Cincinnati [168]. This data set was measured during a run-to-failure experiment in which a bearing experienced an outer race fault. The bearing test rig, as shown in Fig. 4.13, consists of a shaft coupled to an AC motor rotating at 2000 RPM. A radial load of 2700 kg is applied. Four Rexnord ZA-2115 bearings are mounted on the shaft together with high-end accelerometers. In total three run-to-failure experiments were conducted where one second measurements were recorded at a sample rate of 20 kHz every 10 min. The approach described in this chapter is examined on the second data set containing 984 measurement samples and in which Bearing 1 failed due to an outer race defect.

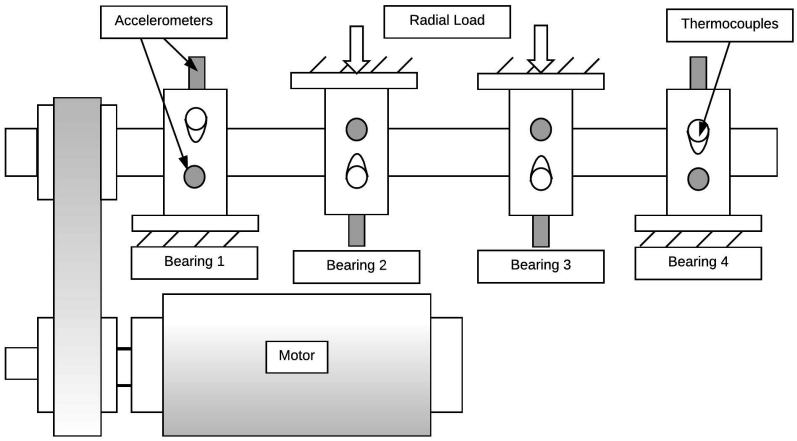

FIGURE 4.13: Bearing test rig [301] of the IMS dataset.

Again, all filters used in the simulation analysis are employed here to provide some reference for comparison of the methods' performance. The characteristic fault frequency of the outer race is approximately 236 Hz and is thus used as input for the ICS2 filter. The filter length is set at 20 samples and the maximum number of iterations is

50. Theoretically the ICS2 filter should be most capable in tracking signal signature changes due to the fault emergence since it tracks the fault frequency modulation directly.

The results shown in Fig. 4.14 seem to confirm this theory. The ICS2 filter manages to already show a rise in ICS2 indicator at the start of day 4. The MED filter and the time-domain negentropy filter perform badly which comes as no surprise since it was already proven in [302] that kurtosis or other indicators tracking non-Gaussianity perform badly for this particular data set. The signal is in fact already non-Gaussian from the start of the test. These two filters therefore underperform in effectively tracking the fault in this case. The LP-envelope filter does not seem to perform that well either based on the relative prediction error evolution. It showcases a slight decrease starting on day 5, but a significant change can only be observed after day 6. The envelope spectrum sparsity-based filters on the other hand suggest the increase of mainly cyclostationarity of the signal due to the fault. Apart from the spectral flatness filter, the sparsity-based filters exhibit a sudden increase in indicator value around the start of day 5. As described in [302], the degradation of the bearing mainly manifests itself by a significant surge in cyclostationarity and not non-Gaussianity. The spectral flatness filter exhibits more outliers after day 4 but does not display any significant trend in feature value.

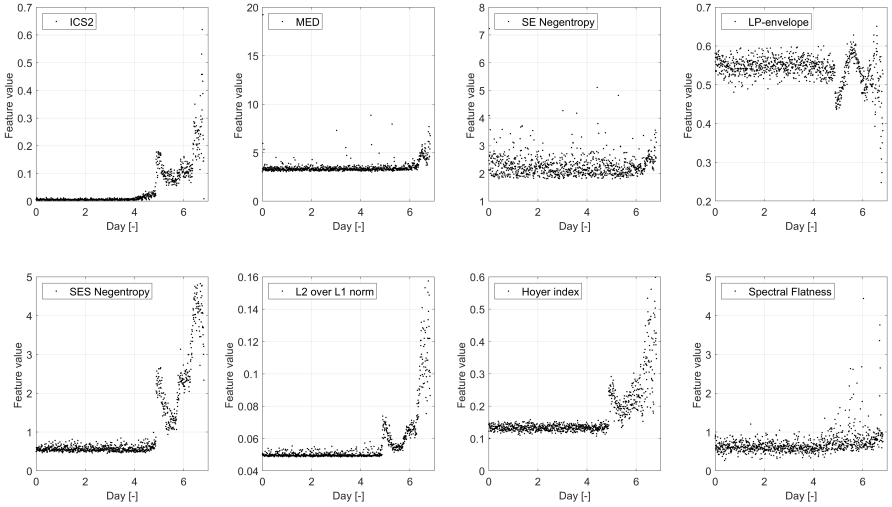

FIGURE 4.14: Evolution of the indicators after filtering with the six different methods on the IMS data set. (Top row, left to right) ICS2 filter, MED filter, Negentropy in the time-domain filter. (Bottom row, left to right) Filters based on respectively the negentropy, the $\frac{l_2}{l_1}$-norm, and Hoyer Index of the envelope spectrum.

The increase in envelope spectrum sparsity means that a dominant discrete peak or set of discrete peaks should show up in the envelope spectrum. Figure 4.15 shows a waterfall colormap plot of the envelope spectra as a function of time from 0 Hz to 500 Hz. The fault frequency at 236 Hz can be detected clearly starting from day 4 for the ICS2 filter, and from day 5 for the sparsity-based filters. Surprisingly, both the LP-envelope and the spectral flatness filter manage to still filter out the 236 Hz fault frequency despite not showcasing any significant indicator increase. This might indicate that their feature definitions are not that well-suited for trending purposes, although more in-depth research needs to be done in order to rule this out. Lastly, the MED and SE negentropy filter perform the worst in enhancing the fault frequency in the SES, which is in correspondence with their feature value evolutions of Fig. 4.14.

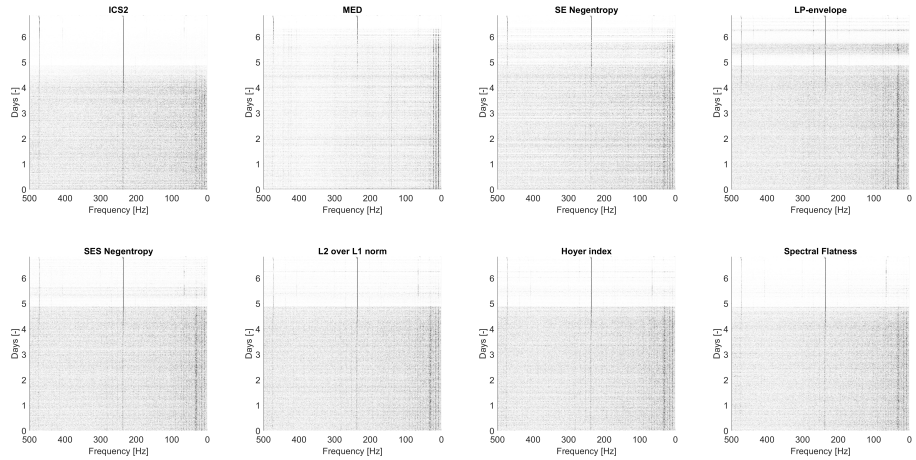

FIGURE 4.15: Waterfall plots of the normalized envelope spectra after filtering with the six different methods on the IMS data set. (Top row, left to right) ICS2 filter, MED filter, Negentropy in the time-domain filter. (Bottom row, left to right) Filters based on respectively the negentropy, the $\frac{l_2}{l_1}$-norm, and Hoyer Index of the envelope spectrum. (Black = 0, white = 1)

4.5 Discussion

The simulation and experimental results showcase the potential for condition monitoring using a blind filtering approach of vibration signals without any prior knowledge of the fault frequencies. The simulation results indicate that filtering based on the sparsity of the envelope spectrum can potentially outperform traditional time indicator-based filtering approaches such as MED. This is especially so when the fault signature introduces cyclostationarity in the signal which is detected effectively by the envelope spectrum. A limitation of kurtosis-based filtering is exemplified for the case when the fault impulses have a high repetition rate and do not fully decay before the next impulse. In this case the kurtosis will only be affected slightly due to the lack of strong non-Gaussian behavior.

Another significant difference in the way the sparsity-based filters function as compared to more conventional blind deconvolution filters such as MED is the fact that the proposed filters purely try to maximize a certain property of the signal. This means that they do not attempt to recover source signals (e.g. repetitive impacts), separate signals from a mixture, or denoise the signal. This is a different blind filtering perspective than typically used in the past. As a consequence, the user needs to keep this property in mind when analysing the results since the filtered response might not accurately represent the signal of interest, rather a distorted version of it that maximizes the used criterion. Nevertheless, the experimental applicability of the suggested approach is apparent since the envelope spectrum is one of the most prevalent tools in vibration-based condition monitoring.

The experimental application illustrates the sensitivity of the proposed approach to cyclostationary sources. While the sparsity-based filters are slightly slower in detecting the fault than the ICS2 filter, they still manage to track the fault accurately without any knowledge of the actual bearing fault frequency. The results suggest as well that looking at the envelope spectrum sparsity of the filtered signal could be a reliable blind tracking measure for the amount of cyclostationarity in a signal. Lastly, the results indicate that more research is necessary to investigate which sparsity indicator performs best in filtering out and tracking cyclostationary faults.

The proposed approach employs a powerful new way of looking at the combination of blind filtering and the envelope time signal and spectrum. As is the case for almost every method however, there are some drawbacks or hindrances that can negatively influence results. The main difficulty in using this type of filters resides in the choice for the filter length. This length can impact the results and the computation time significantly. Obviously longer filter lengths increase the computation time, but they can also lead to slower convergence or non-convergence of the filter coefficients. The latter can be especially troublesome if the indicator values oscillate strongly due to this non-convergence. Therefore it is recommended to keep the filter length short, while maintaining the capability of the filter in suppressing unwanted frequency content.

Some additional side-notes about this sparsity-based approach can be made. While the goal of the method is to maximize the sparsity of the envelope spectrum after filtering, it is assumed that the fault content is localized in a frequency band or a limited set of frequency bands. This means that even though a more sparse envelope spectrum is obtained after filtering, it still contains the modulation sidebands since it is assumed they modulate the same frequency region as the one the fault occupies, and thus they cannot be filtered out by simple frequency filtering. Nevertheless, it is possible that the filtering can alter the modulation sidebands as compared to the source signal of the fault. This assumption also implies that the performance of the method can suffer significantly if there are other modulating or impulsive signals occupying the same frequency range as the signal of interest, such as electromagnetic interference (EMI) signals (from e.g. variable frequency drives) or vibrations due to cavitation in pumps. Some caution is thus required when interpreting the results of the method.

It is in any case recommended to track the sparsity of the envelope spectrum over time such that a reference is available for when the sparsity increases due to an emerging second-order cyclostationary component. This way, if an increase in sparsity is observed and a frequency peak arises (or multiple peaks), the operator knows that it is related to a cyclostationary source, which is often related to a potential mechanical fault such as gear or bearing damage. To fully determine the fault type and source however, more in-depth analysis is still required.

Despite this obstacle, this chapter hopes to pave the way for further research in novel blind filtering approaches for vibration monitoring. The Rayleigh quotient iteration framework detailed in this chapter is the main tool responsible for the ease of implementation of the proposed filters. Yet this does not mean other approaches are not viable. A future improvement could be to reduce the sensitivity to the filter length and improve the convergence behavior.

4.6 Conclusion

This chapter investigates novel blind filtering techniques that utilize the predictability of the envelope and the sparsity of the envelope spectrum to track the presence of faults with a second-order cyclostationary signature in vibration signals. The envelope prediction error and four sparsity measures on the envelope spectrum are used to derive filter formulations that are applicable in a methodology based on Rayleigh quotient iteration. This approach allows for fast computation and straightforward implementation of the proposed filters.

Both the simulation and experimental results indicate the potential of the proposed blind filters in tracking the emergence of cyclostationary sources. Three out of four suggested sparsity measures perform in a similar manner with regards to their rise in absolute values and to their corresponding filtering behavior. When the fault signal does not exhibit strong impulsive content, the described approach outperforms conventional blind deconvolution methods such as MED and its variants.

Future work can focus on improving the stability of the results so that filter length plays less of an influential role in the computation time and convergence behavior of the algorithm.

Appendix A

It can be proven that the spectral negentropy is always greater than zero and smaller than $ln(N)$:

$$0 \leq \Delta I_E \leq ln(N) \tag{4.86}$$

Starting from the Shannon (or information) entropy H, the following is valid:

$$H = -\sum_i p_i ln(p_i) \geq 0 \; and \; \sum_i p_i = 1 \tag{4.87}$$

with p_i a discrete set of probabilities.

Thus for $p_i = \frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2}$ the following is true:

$$\sum_i^N \frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2} ln(\frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2}) \leq 0 \tag{4.88}$$

$$\frac{1}{N}\sum_i^N \frac{|\boldsymbol{E_i}|^2}{\frac{1}{N}\sum_i^N |\boldsymbol{E_i}|^2} ln(\frac{\frac{1}{N}|\boldsymbol{E_i}|^2}{\frac{1}{N}\sum_i^N |\boldsymbol{E_i}|^2}) \leq 0 \tag{4.89}$$

$$\left\langle \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) + \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{1}{N}) \right\rangle \leq 0 \tag{4.90}$$

$$\left\langle \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \right\rangle + \frac{\langle |\boldsymbol{E_i}|^2 \rangle}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{1}{N}) \leq 0 \tag{4.91}$$

$$\left\langle \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \right\rangle - ln(N) \leq 0 \tag{4.92}$$

$$\left\langle \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \right\rangle \leq ln(N) \tag{4.93}$$

$$\tag{4.94}$$

It is also known from Kullback-Leibler divergence (or relative entropy) that:

$$D_{KL}(p||q) = \sum_i p_i ln(\frac{p_i}{q_i}) \geq 0 \; and \; \sum_i p_i = 1, \sum_i q_i = 1 \tag{4.95}$$

Thus for $p_i = \frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2}$ and taking a uniform distribution for q with $q_i = \frac{1}{N}$ the following is true:

$$\sum_i^N \frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2} ln(\frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2 \frac{1}{N}}) \geq 0 \tag{4.96}$$

$$\frac{N}{N} \sum_i^N \frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \geq 0 \tag{4.97}$$

$$\frac{1}{N} \sum_i^N \frac{|\boldsymbol{E_i}|^2}{\sum_i^N |\boldsymbol{E_i}|^2 \frac{1}{N}} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \geq 0 \tag{4.98}$$

$$\left\langle \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \right\rangle \geq 0 \tag{4.99}$$

Therefore, with $\Delta I_E = \left\langle \frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle} ln(\frac{|\boldsymbol{E_i}|^2}{\langle |\boldsymbol{E_i}|^2 \rangle}) \right\rangle$:

$$0 \leq \Delta I_E \leq ln(N) \tag{4.100}$$

The same is thus also valid for the negentropy in the time domain ΔI_ϵ.

Chapter 5

Fault detection & diagnosis

5.1 Introduction

The previous chapters discussed methodologies that focus on preprocessing vibration signals such that they become easier to utilize for fault analysis. The reduction in influence of speed fluctuation was discussed in these chapters, next to separation into deterministic and stochastic content, and spectral filtering. This chapter details the one remaining and typically last step in a full vibration analysis procedure, namely that of the fault identification. This consists of translating the preprocessed signals into information that is understandable for a human end-user. Normally, this is accomplished by calculating scalar condition indicators on the preprocessed signals. A concise overview is given in this chapter of the possible options to compute such condition or health indicators.

Ideally, an end-user needs to inspect only a few summarized, high-level graphs in order to get a quick idea about the health of a machine. However, due to the complexity of machinery nowadays, this is often a difficult feat to achieve. Potentially, the condition of every machine subcomponent is tracked over time leading to a plethora of health indicators. For this reason, the usage of machine learning techniques is employed to reduce the number of indicators to a manageable set of health indicators. This chapter therefore illustrates the machine learning approach developed to tackle this issue.

The structure of this chapter is slightly different from the previous chapters in the sense that there is no large body of content introducing and investigating a new signal processing methodology. Instead, the state-of-the-art in time and frequency domain indicators for vibration-based condition monitoring is discussed in sections 5.2 & 5.3. Nevertheless, a minor new development that was derived during this PhD work is briefly presented in sections 5.2.1 & 5.3. The majority of the performed investigative work is presented in section 5.5 about the hybrid physics and model-based automation approach used in major experimental data analysis cases.

5.2 Time domain approaches

Current practice in condition monitoring systems often revolves around tracking time-domain statistical indicators [303–305]. The advantage of using simple scalar time-domain indicators is that no prior knowledge about the characteristic fault frequencies is required and thus the number and complexity of the components is not taken into account. This simplifies the analysis procedure significantly, with the trade-off being that it does not allow pinpointing which component is exhibiting the anomalous behavior. Not having any knowledge about the exact location of a fault is a downside, but in an early analysis stage it is often easier and quicker to have a straightforward overview of which measurement sensor is picking up faulty behavior. Usually these time-domain condition indicators are trended over time such that the potential fault evolution can be followed up on.

Examples of commonly used time-domain indicators are given in Table 5.1 [302, 306–314]. While many more indicators can be found in the literature, it is not the intent to give a full exhaustive overview here. These time indicators can all be used to characterize trends in measured vibration signals.

TABLE 5.1: Overview of commonly used time-domain condition indicators in vibration analysis.

Indicator	Expression
RMS	$\sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)}$
Crest factor	$\frac{\lvert x_{peak}\rvert}{\sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)}}$
Kurtosis	$\frac{\frac{1}{n}\sum_{n=1}^{N}(x_i-\overline{x})^4}{(\frac{1}{n}\sum_{n=1}^{N}(x_i-\overline{x})^2)^2} - 3$
Moors kurtosis	$\frac{(E_7-E_5)+(E_3-E_1)}{E_6-E_2}$
Peak-to-peak	$x_{max} - x_{min}$
Peak energy index	$\sqrt{\frac{1}{N_p}\sum_{n=1}^{N} x_p^2(n)}$
Form/Shape factor	$\frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)}}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$
Impulse factor	$\frac{x_{max}}{\left(\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert\right)^2}$
Margin factor	$\frac{x_{max}}{\left(\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert\right)^2}$
Skewness	$\frac{\frac{1}{N}\sum_{n=1}^{N}(x(n)-\overline{x})^3}{\left[\frac{1}{N-1}\sum_{n=1}^{N}(x(n)-\overline{x})^2\right]}$
Higher-order moments	$\frac{1}{N-1}\sum_{n=1}^{N}[x(n)]^k$
FMo	$\frac{x_{max}-x_{min}}{\sum_{n=1}^{N} A(n)}$
FM4	$\frac{\frac{1}{N}\sum_{n=1}^{N}\left(d(n)-\overline{d(n)}\right)^4}{\left[\frac{1}{N}\sum_{n=1}^{N}\left(d(n)-\overline{d(n)}\right)^2\right]^2}$
NA4	$\frac{\frac{1}{N}\sum_{n=1}^{N}\left(r(n)-\overline{r(n)}\right)^4}{\left[\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{n=1}^{N}\left(r(n,m)-\overline{r(m)}\right)^2\right]^2}$
NB4	$\frac{\frac{1}{N}\sum_{n=1}^{N}\left(x_{env}(n)-\overline{x_{env}(n)}\right)^4}{\left[\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{n=1}^{N}\left(x_{env}(n,m)-\overline{x_{env}(m)}\right)^2\right]^2}$
Smoothness index	$\frac{\left(\prod_{n=1}^{N} S(n)\right)^{1/N}}{\frac{1}{N}\sum_{n=1}^{N} S(n)}$
Shannon entropy	$-\sum_{n=1}^{N} p(x_n) log(p(x_n))$
Renyi entropy	$\frac{1}{1-\alpha} log(\sum_{n=1}^{N} p(x_n)^\alpha)$
Sample entropy	See [315]
Mean absolute difference	$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\lvert x_i-x_j\rvert}{N(N-1)}$

With $x(n)$ the sampled vibration signal, E_i the i^{th} octile, N_p the number of peaks above a chosen threshold, $d(n)$ the difference signal [308], $r(n)$ the residual signal [308, 313], M the time record number in the run ensemble [308, 313], x_{env} the envelope signal.

5.2.1 Frequency filtering

One preprocessing technique that is often added right before the indicator calculation, is frequency filtering. In order to add sensitivity of the statistical indicators to frequency-localized phenomena such as resonance amplification of bearing or gear faults, vibration signals can be filtered using a set of band-pass filters. While many different frequency regions can be defined as pass-bands, an interesting filtering approach was introduced by Antoni by means of the kurtogram [262]. The kurtogram filterbank concept was already introduced in Chapter 4 and shown in Fig. 4.1. The dyadic binary-ternary frequency grid for the filterbank used by the kurtogram can however be reused for other indicators aside from kurtosis. This idea was explored during the PhD work and an example of the work can be seen in Fig. 5.1. This figure shows four different indicators calculated with a binary-ternary filterbank decomposition up to level 7, on a signal consisting of weak repetitive transients hidden in stationary white Gaussian noise. The spectral content of the transients is in the frequency band from 0.15 Hz to 0.19 Hz. The four chosen indicators highlight these impulsive transients nicely, but only thanks to the additional frequency filtering. In theory, every single indicator of each frequency band in these 2D maps can be trended over time, yet again reiterating the need for a comprehensive automated indicator assessment methodology.

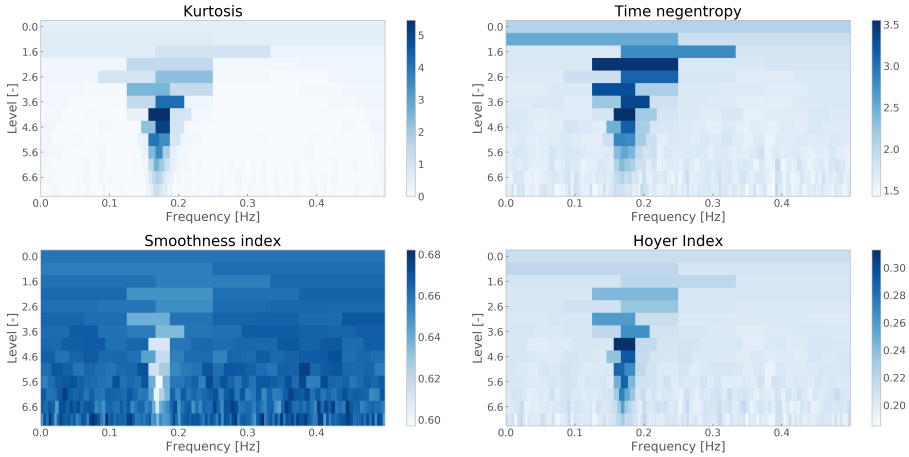

FIGURE 5.1: Different indicators calculated with the binary-ternary filter tree decomposition concept (based on the kurtogram [262]).

5.2.2 Regarding the abundance of features

All the indicators mentioned in this chapter can be calculated on both raw measurements or preprocessed vibration signals. This optional choice for additional preprocessing is one of the main reasons why monitoring a complex machine brings about immense numbers of condition indicators.

As a quick example, assume that a machine operator or end-user does not know which preprocessing methodology is optimal to detect faults in his machine (which is not such an unrealistic scenario). Therefore, to keep his options open, he decides to calculate a set of ten indicators on each preprocessed signal and on the raw data. If his software allows to do four different preprocessing techniques, this already leads to 50 different condition indicators. If certain techniques can be combined with one another, then the number of indicators increases even further. In case the end-user operates a fleet of machines instead of just one, this number also gets multiplied with the number of machines in the fleet.

It is thus easy to see that it becomes unfeasible very quickly for a human operator to inspect every single indicator trend. This phenomenon is the main reason and motivation for the methodology presented in section 5.5.

5.3 Frequency domain approaches

This section describes diagnostic approaches that are primarily based on the tracking of amplitudes in the frequency domain. The main reason why the use of spectral methods is so popular in condition monitoring is that it not only allows for fault detection, but also for fault diagnosis. A rotating mechanical subsystem normally has a set of known characteristic frequencies associated with it. Typical examples of machine components that allow for fairly straightforward calculation of their characteristic frequencies are shafts, gears, and bearings. These frequencies provide the necessary means to an operator for pinpointing potential mechanical defects of the machine components. For example when a bearing has an inner race defect such as micropitting, this defect shows up at the corresponding ball pass frequency of the inner race, often abbreviated as BPFI. In general, bearing vibrations are described with four fundamental characteristic frequencies, each one related to one of their mechanical parts, i.e. inner race, outer race, roller cage, and roller. The expressions necessary to calculate these four frequencies are given in Table 5.2. Gears have a different set of characteristic frequencies that can be calculated. These are given in Table 5.3 and are related to the gear meshing process.

TABLE 5.2: Characteristic bearing frequencies.

Name	Expression
FTF	$\frac{\omega}{2}(1 - \frac{D_r}{D_p}\cos(\phi)$
BPFI	$\frac{N_r}{2}\omega(1 + \frac{D_r}{D_p}\cos(\phi)$
BPFO	$\frac{N_r}{2}\omega(1 - \frac{D_r}{D_p}\cos(\phi)$
BSF	$\frac{D_p}{2D_r}\omega\big(1 - (\frac{D_r}{D_p})^2[\cos(\phi)]^2\big)$

With FTF the Fundamental Train Frequency, BPFI the Ball Pass Frequency of the Inner ring, BPFO the Ball Pass Frequency of the Outer ring, BSF the Ball Spin Frequency, ω the rotation speed, D_r the diameter of the roller, D_p the pitch diameter, N_r the number of rollers, and ϕ the contact angle.

TABLE 5.3: Characteristic gear frequencies.

Name	Expression
Gear rotational frequency f_{rg}	ω_g
Pinion rotational frequency f_{rp}	ω_p
Mesh frequency f_m	$\omega_g N_g$ or $\omega_p N_p$
Assembly phase passage frequency f_a	$\frac{f_m}{N_a}$
Tooth repeat or hunting tooth frequency f_{tr}	$\frac{f_m N_a}{N_g N_p}$

With N_p the number of teeth on the pinion, ω_p the pinion rotation speed, ω_g the gear rotation speed, N_g the number of gear teeth, and N_a the product of common prime factors of N_g & N_p.

While both gears and bearings have well-defined fault frequencies, the detection of these fault frequencies often happens in different ways due to the inherent nature of the mechanical processes behind gear and bearing fault phenomena. The main reason for this difference in detection approaches can be attributed to the slip phenomenon that occurs during the rotation of a bearing. While the roller cage of a bearing guides the rotation of the rollers, it does not physically lock them at the rotation speed of the shaft the bearing is mounted on. The

rollers are actually allowed to experience slip between themselves and the race surfaces. Therefore, the rotation speed of these rollers is slightly asynchronous with the rotation speed. When a fault such as pitting then occurs on one of the races, this slip introduces jitter on the fundamental period of the impulses caused by the pitting. This jitter causes the discrete spectral peaks associated with the frequency signature of the fault to smear, leading to a significant decrease in amplitude in the spectrum. Gears on the other hand are locked at the rotation speed due to the shaft mounting and the gear meshing. Gear signals are therefore considered to be purely periodic or deterministic. In contrast, bearing signals are normally characterized as being second-order cyclostationary, meaning they have a periodic autocorrelation. This difference in signal nature lead to the development of different diagnostic methodologies.

For gears, approaches for fault detection usually focus more on tracking the amplitudes of harmonics and sidebands in the spectrum or the cepstrum, while for bearings more cyclostationarity-based methods tend to be employed. Nonetheless, also for gear faults it is recommended to look at the cyclostationary signature of a signal since distributed gear faults can also significantly impact the modulation of the deterministic gear signals. An example of how sidebands evolve with the type of the fault, is presented in Fig. 5.2. A local fault on one of the gear teeth will introduce low-level sidebands in the spectrum, whereas distributed gear damage exhibits higher level sidebands.

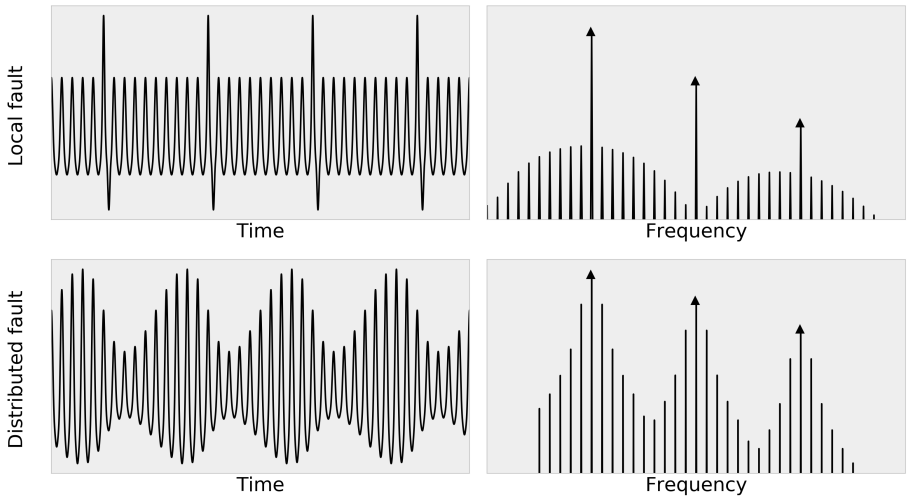

FIGURE 5.2: Comparison of time and frequency domain signatures of local and distributed faults in gears.

The presence of modulation sidebands is also the main reason why the usage of cepstrum-based techniques is popular for gear diagnostics. Since the cepstrum groups together equally spaced harmonics, it provides a very effective means to track the average amplitude of sidebands. By tracking the amplitude of the first rahmonics of sideband or harmonic families, the overall condition of the gear can be trended in a simple manner, instead of having to track multiple peaks in the spectrum simultaneously. Another benefit of looking at the cepstrum rahmonics is the accuracy of the sideband frequency estimation thanks to the averaging effect.

5.3.1 Cyclostationarity

Nowadays, probably the most popular tools for the detection of bearing faults are based on the theory of cyclostationarity. Generally speaking, the cyclostationary properties of a signal are interesting from a condition monitoring perspective since they tell us something about the modulations present in a signal. Knowledge of the modulations in a signal is valuable considering that mechanical faults in a rotating system typically introduce modulations. This includes bearing faults.

Cyclostationary signals with different cyclic frequencies also allow for signal separation, similar to what is presented in Chapter 3. In the literature, most research focuses on the detection and analysis of *second-order* cyclostationary signals, since this class of signals encompasses the majority of fault signatures encountered in mechanical rotating systems. Briefly put, a *first-order* cyclostationary signal has a periodic mean value (e.g. a sine wave plus noise), whereas a *second-order* cyclostationary signal has a periodic autocorrelation (e.g. amplitude-modulated white noise). In other words, a second-order cyclostationary signal has a periodic energy flow with repetitive bursts of energy. An example of the latter signal and its corresponding amplitude spectrum is presented in Fig. 5.3 (the axes are not specified here since the figure is solely meant as an example). It should be noted that cyclostationarity is not limited to the first and second-order statistics of a signal, there are also higher moments (and cumulants) that can describe signals. A full exploration of cyclostationarity is not the focus of this chapter, but for the interested readers, there is a large body of literature available describing in much more detail the cyclostationary theory (e.g. higher-order cyclostationarity) [316–321].

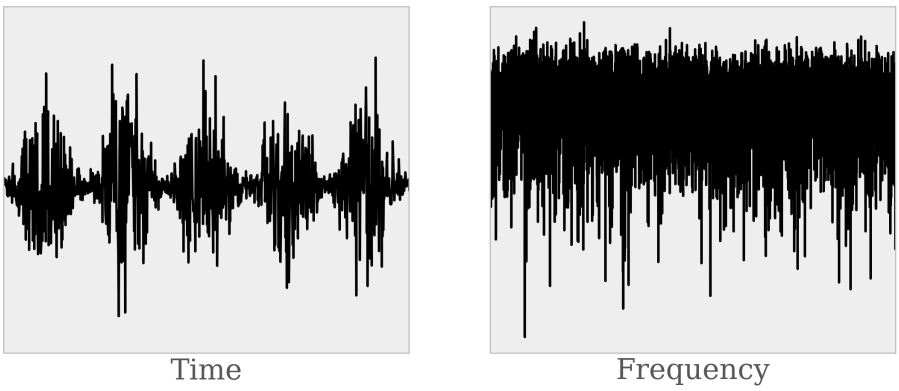

FIGURE 5.3: Example of a second-order cyclostationary signal (i.c. amplitude-modulated white noise) (Left) in the time domain, (Right) in the frequency domain.

Looking at Fig. 5.3, it can be observed that while the white noise clearly has a visually distinguishable sinusoidal modulation, this is

not observable in the amplitude spectrum at all. In this case, the carrier is white noise so the sidebands are effectively smeared out over the entire spectrum. If we would calculate the autocorrelation of the spectrum however, we would observe peaks at the corresponding cyclic frequency of the modulation, meaning the frequency bins of the white noise are correlated with each other despite being visually indistinguishable. This phenomenon is detailed further in the paragraph about spectral correlation in section 5.3.1.1.

5.3.1.1 Cyclostationarity analysis methods

Envelope analysis Probably by far the most popular approach for the analysis of such second-order cyclostationary signals (and correspondingly for bearing diagnostics) is envelope analysis. The envelope of a signal is considered to be any function that 'encloses' the energy variation in the signal. By taking the modulus of the analytic version of the signal, obtained through the Hilbert transform, the envelope time waveform can be found. Usually the envelope is squared (effectively done by multiplying the analytic signal with its conjugate) before taking the Fourier transform to analyse its envelope spectrum [159]. Figure 5.4 shows the amplitude envelope obtained through the Hilbert transform of the example signal of Fig. 5.3 and the corresponding squared envelope spectrum (SES). The cyclic frequency of the modulation is easily detected, indicating the efficacy of the SES.

It should be noted that when the signal of interest is a narrowband signal, e.g. when bearing fault impulses are amplified by housing resonances, the signal is usually band-pass filtered at the resonance frequency band prior to amplitude demodulation.

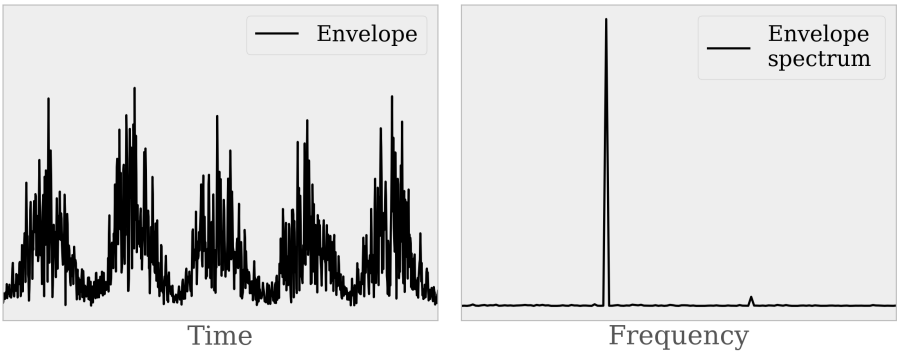

FIGURE 5.4: Envelope analysis of the example signal in Fig. 5.3: (left) Envelope time waveform, (right) Envelope spectrum.

Spectral correlation While envelope analysis has long proven its worth, it provides only part of the available cyclostationary information in a signal. Cyclostationarity actually generalizes envelope analysis in the sense that it can remove the dependency on the frequency band choice for the band-pass filtering. This is where the spectral correlation density comes to the fore and a brief introduction of this quantity is necessary.

The spectral correlation density $SC_x^\alpha(f)$ of a signal $x(t)$ is defined as:

$$SC_x^\alpha(f) = \lim_{\Delta f \to 0, T \to \infty} \frac{1}{T\Delta f} \int_T x_{\Delta f}(t; f + \alpha/2)\overline{x_{\Delta f}(t; f - \alpha/2)} e^{-j2\pi\alpha t} \mathrm{d}t \tag{5.1}$$

where α represents the cyclic frequency and f the carrier frequency. From this definition it can be seen that $SC_x^\alpha(f)$ is a density of correlation between two spectral components spaced apart by α around the central carrier frequency of f [319]. If a signal exhibits pure second-order cyclostationarity, the $SC_x^\alpha(f)$ is expected to have non-zero values parallel to the carrier frequency axis. A good example of such a pure second-order cyclostationary signal is the signal of Fig. 5.3 (with

no periodic carrier). The instantaneous autocorrelation and the spectral correlation of this signal are shown in Fig. 5.5 & 5.6. The cyclic modulation of the white noise is clearly supported over the entire carrier frequency range.

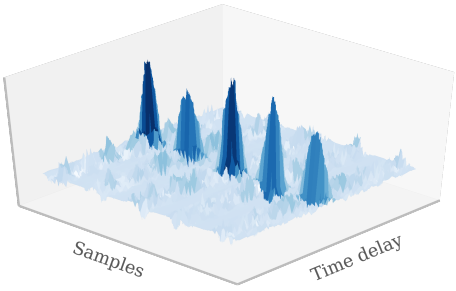

FIGURE 5.5: Instantaneous autocorrelation as a function of time delay/lag and time samples.

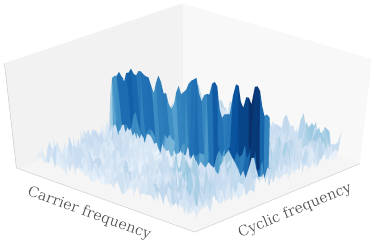

FIGURE 5.6: Spectral correlation, obtained after a two-dimensional FFT over the instantaneous autocorrelation.

In practice, the most used version of the spectral correlation density is likely the energy-normalized one, known as the spectral coherence. It is defined as follows:

$$\gamma_x^\alpha(f) = \frac{SC_x^\alpha(f)}{\sqrt{SC_x^0(f+\alpha/2)SC_x^0(f-\alpha/2)}} \tag{5.2}$$

The spectral coherence $\gamma_x^\alpha(f)$ ranges from 0 to 1 and forms a very useful detection method for second-order cyclostationary signatures [287, 322]. It can also be seen as the spectral correlation of the whitened signal.

Since the spectral coherence $\gamma_x^\alpha(f)$ is a two-dimensional map, it is often simplified by taking the average of the amplitudes over a carrier frequency band. This is the so-called “enhanced envelope spectrum” (EES):

$$S_x^{EES}(\alpha) = \int_{f_1}^{f_2} |\gamma_x(\alpha, f)|\, df \tag{5.3}$$

This spectrum allows for easy tracking of the amplitudes of modulating frequencies in a signal and is thus often an effective tool in condition monitoring schemes for bearing fault signals.

5.3.1.2 Harmonogram

Based on the idea of extending the filterbank approach originally proposed as the kurtogram in [262], another new frequency-based approach was developed to help in finding the optimal resonance frequency band to demodulate a bearing fault. Similar to how the filterbank can be employed for different statistical indicators (as presented in section 5.2.1), it is now used instead to find the frequency band that maximizes the amplitude of a fault frequency of interest in the squared envelope spectrum. Similar to the kurtogram, the signal is filtered following a binary-ternary tree decomposition. Next, the squared envelope spectrum is then calculated of the filtered analytic signals and normalized between 0 and 1. Given the fault frequency α of the bearing, a simple feature $\lambda(f, \Delta f)$ is then computed on each normalized SES as follows:

$$\lambda(f, \Delta f) = \sum_{i=1}^{I} SES_{f,\Delta f}(i\alpha) - \beta(i\alpha) \tag{5.4}$$

with I the number of SES peaks that should be taken into account for the calculation of the feature, and β a threshold calculated by taking a moving average over the SES. Thus, the method tries to maximize the envelope spectrum peaks while avoiding maximizing a noise peak through the use of a threshold. Since the method employs the envelope spectrum harmonics, the method is coined 'harmonogram'. The usage of this method is more envisioned to be an a-posteriori method for fault analysis, rather than a means that can be used for trending. The latter would imply calculating a harmonogram for every potential fault frequency in a system, which would lead to long computation times and many indicators to track.

The proposed method's performance is illustrated on a simulated outer race bearing fault signal. This bearing fault signal is polluted with additive white Gaussian noise at an SNR of 5 dB. The sample rate is 1 Hz and a signal of $1e^5$ samples is simulated. Also harmonics are added at multiples of 0.02 Hz. A summary of the parameters for the outer race bearing fault is given:

- BPFO = 0.025 Hz

- $f_{resonance} = 0.2$ Hz
- damping $\zeta = 0.1$
- random fault period jitter of 2%

Figure 5.7 shows part of the simulated outer race bearing fault signal and the resulting signal with the noise and the harmonics, as well as the amplitude spectrum of the simulated signal.

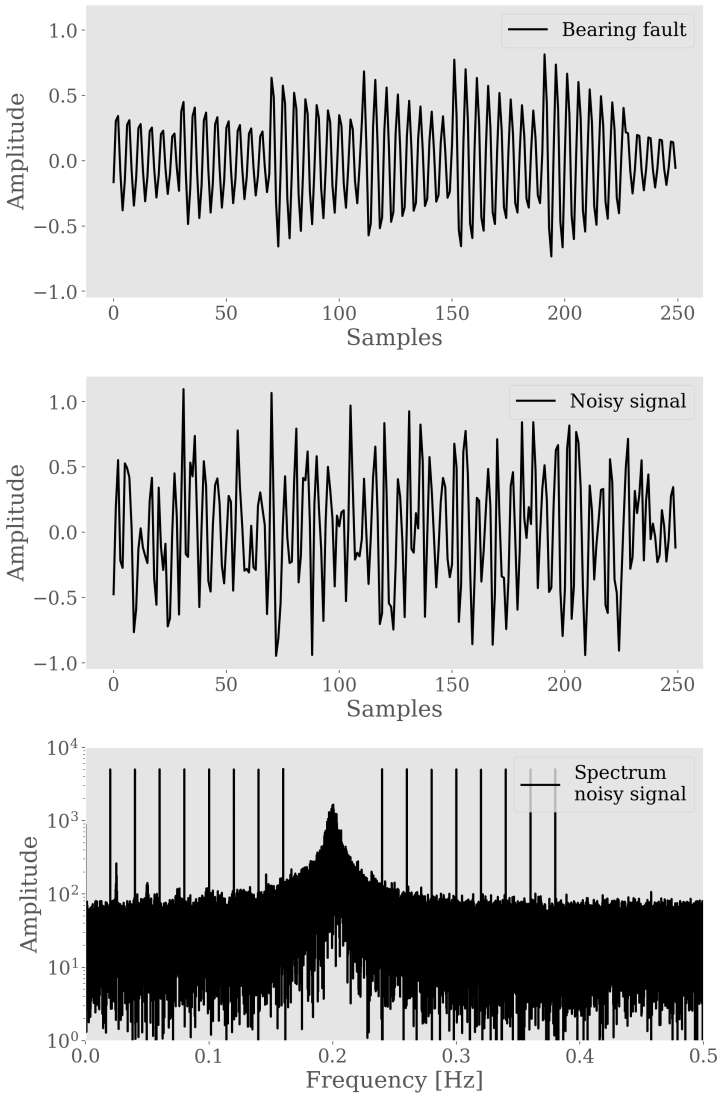

FIGURE 5.7: (Top) the simulated outer race bearing fault signal, (Middle) the full signal with added white Gaussian noise and harmonics, (Bottom) the spectrum of the simulated signal.

Next, the harmonogram and the kurtogram (for comparison) are

computed up to level 5 and shown in Fig. 5.8. It is clear that the harmonogram succeeds in detecting the bearing fault resonance, whereas the kurtogram is skewed due the harmonics and indicates an erroneous resonance band in this case. This is corroborated in Fig. 5.9 where the envelope spectra of the filtered demodulated signals is shown for both approaches. The harmonogram SES has a distinctive peak at the fault frequency of 0.25 Hz (represented by the red dotted line), while the fault frequency in the kurtogram SES is entirely masked by the harmonics.

It is worth noting that the performance of the kurtogram could probably be improved by removing the harmonics prior to its application by means of a discrete-random separation method. However, the example presented here is for illustration purposes and indicates the potential advantage of an approach like harmonogram for cases when the deterministic component removal fails to properly separate the harmonics.

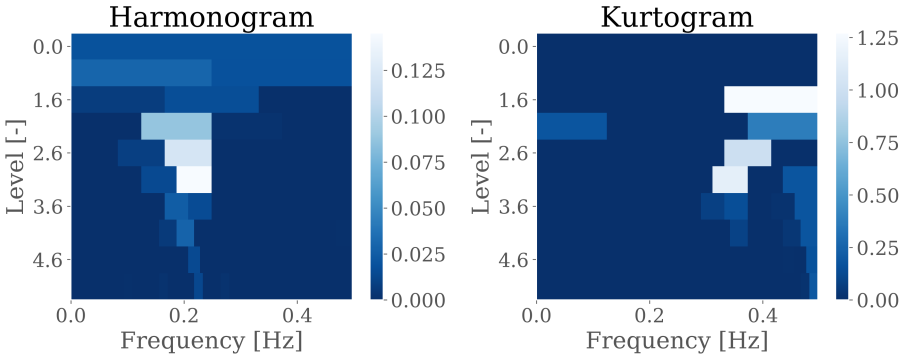

FIGURE 5.8: The harmonogram on the left indicates the resonance frequency band of the simulated fault at 0.2 Hz, while the kurtogram on the right is skewed and fails to indicate the resonance band.

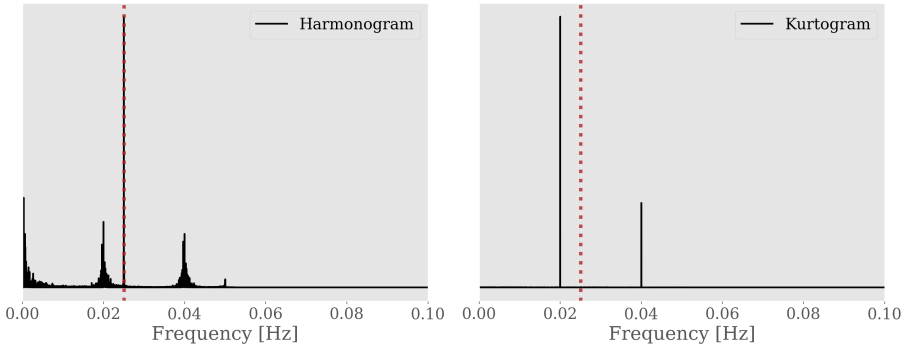

FIGURE 5.9: Comparison of the demodulated envelope spectra for the harmonogram (left) and the kurtogram (right). The red dotted line indicates the BPFO fault frequency at 0.025 Hz, while the 0.02 Hz peak corresponds to the harmonic frequencies.

5.4 Regarding the domain classification

Some explanation is necessary as to why this chapter chooses to separate the concept of vibration indicators into the two separate classes of time and frequency domain indicators. There are a couple of reasons why the choice for this classification is made this way. First off, in the majority of publications in the literature, you will find the same or similar domain-based categories for condition indicators. Often a set of indicators or a new indicator is introduced with the focus on whether it is defined in the time or frequency domain. Even MATLAB uses this exact same categorization in their Predictive Maintenance Toolbox [323]. Another reason to use these two categories is because of the intuitiveness. Often concepts get explained in the domain that makes for the simplest explanation. For example, kurtosis or impulsiveness is typically described from a time domain perspective since it is much more instinctive to comprehend as compared to kurtosis in the frequency domain.

Nonetheless, it is possible to also bring up counterarguments to this classification into time and frequency domain. The main objection

to it is probably that most indicators are quite easily interchangeable between the time and frequency domain since the Fourier transform is a linear and invertible transform. In fact, very recently another classification for condition indicators was proposed that does not utilize the time and frequency domain partitioning. Instead it proposes to categorize condition indicators into two groups: one for characterizing non-stationary behavior and one for non-Gaussian behavior [302]. It turns out that most condition indicators are quantifying one or both of these signal characteristics, regardless of which domain they are calculated in. Examples of non-Gaussian indicators include kurtosis, crest factor, or skewness. Indicators of non-stationarity are typically related to the concept of cyclostationarity, e.g. the degree of cyclostationarity. Taking all this into consideration, it would therefore also make sense to adopt this alternative classification. However, to stay in line with some of the previously done work, the former classification of time and frequency domain indicators is used in this dissertation.

5.5 Data-driven automation

Now that we have established an overview of the most commonly used methodologies for fault detection, we take a look at how we can start to automate the inspection of their output results. Thanks to the recent advances of machine learning, it has now become more feasible to not only identify relevant condition indicators, but also relate them to the operating regime of a machine for each measurement. It turns out that most of the commonly used condition indicators are sensitive to changes in the operating parameters and thus can lead to erroneous conclusions if they are not compensated for these changes.

This section about automation is the result of a collaboration with the Artificial Intelligence department at VUB, and in particular with *Timothy Verstraeten* and *Ann Nowé*. They did all the necessary steps with regards to training the ML models and the anomaly detection. Our joint research efforts examined the potential for using machine learning methods to reduce the influence of strongly varying operating regimes. It illustrates how condition indicators can be combined with machine learning (ML) techniques to allow for the normalization

of indicators, making them independent of the operating regime, and how it can provide a streamlined methodology for efficient early fault analyses. Ideally, the provided indicators measure a certain property in the vibration signal that is linked to a specific failure type in order to provide the diagnostic ML models with physics-based and thus relevant features. The meaning of physics-based used here, should not be confused with physics-based modeling. The ML models are not linked with any physics-based model, but use indicators that are based on tracking certain physical phenomena in the monitored machine.

Currently, this research has mainly focused on automatic feature reduction and automatic alarm thresholding. However, in future work, further steps towards complete automation of the vibration signal processing chain will be taken (e.g. automatically setting the input parameters for the different methods).

5.5.1 Expected behavior modeling

As shown in the previous sections, various condition indicators can be constructed from vibration data to represent the machine conditions. Analysing these indicator values can help us to find anomalies in the data, which may indicate failure. For example, when impulses are found in a subset of the indicators in a particular time frame, this can be reported to the machine operator, such that an investigation of specific components can be carried out.

However, manually investigating a large set of features is tedious work. Moreover, it is subjective in the sense that smaller trends may be overlooked, or complex trends are not considered to be indicative for an anomaly, even if it deviates from the normal trends. Therefore, we opt for using ML techniques and statistics to objectively and automatically annotate anomalies over the set of indicators.

In order to automatically detect anomalies, we first must define what *normal* machine behavior is, such that deviations from this behavior can be detected. A potential way to achieve this is by modeling the machine behavior by training a ML model using indicators of historic vibration data under healthy conditions.

5.5.1.1 Operating condition independence.

The health indicators can be made independent from the operating conditions by performing k-means clustering on the operational data (e.g. SCADA data). K-means is a machine learning technique that finds k centroids, such that the squared distances between each data point and any centroid is minimized [324]. Afterwards, each data point is assigned to its closest centroid, effectively clustering the data into partitions. To define the operating condition, we use operational parameters that are known to influence the machine's vibrational behavior, and then perform k-means on a large set of historic data. This clustering step makes sure that each condition indicator is objectively binned into operating regimes, enforcing independence for later modeling.

5.5.1.2 Healthy condition model.

Per operating regime, we bin the data and can use techniques such as linear Bayesian Ridge Regression [325] to map the operational parameters (e.g. rotation speed and power production) to a particular feature. Bayesian Ridge Regression is a probabilistic approach to regression with regularization. Essentially, it fits the linear parameters (i.e. slopes and intercept) and inherent noise to the observed data, while maintaining the uncertainty over the parameters. Formally, the following model is used.

$$\begin{aligned} f(x) &= [x, 1]w^T \\ y(x) &= f(x) + \epsilon \\ \epsilon &\sim \mathcal{N}(0, \sigma_n I) \\ w &\sim \mathcal{N}(0, \sigma_f I) \end{aligned} \tag{5.5}$$

where x, y are operational parameters and associated observed statistical feature values, and σ_f, σ_n are respectively the regularizer and observational noise. The frequently used vague inverse-gamma priors [326] are chosen to model the uncertainty over σ_n and σ_f, such that the posterior estimations depend mainly on the observed data. For modeling the machine's healthy behavior, we are mainly interested in posterior estimations of w (linear parameters) and σ_n (observational

noise). Note that such a model is trained for each operating condition and statistical feature separately.

The choice of this model has several advantages. First, once independence w.r.t. the operating conditions is induced, the features show either constant or linear behavior, which makes the linearity of the model sufficient for our purposes. Second, the regularizer is often introduced when dealing with over-determined systems [327]. This provides more robustness against unexpected phenomena (e.g. outliers) that are challenging for a traditional linear model to deal with. Finally, a Bayesian approach maintains the uncertainties of the parameters in the face of small data sets. When only a few effective data points are available, the linear weight parameters are heavily influenced by noise, which makes the trend uncertain (i.e. many trends can describe the observed data). Quantifying all types of uncertainty properly is necessary for our anomaly detection mechanism, which analyzes deviations from a healthy linear trend in terms of the noise present in the model. As detecting a deviation from a normal linear trend is meaningless when this trend is highly uncertain, this noise should be properly quantified by taking both the observational noise and parameter uncertainty into account [328].

5.5.2 Anomaly detection

The aim of the proposed approach is to predict failure by detecting anomalies in the expected machine behavior. This behavior is modeled using time frames of the operational data and statistical features where the machine is known to be healthy.

We define an anomaly to be any observed data point that deviates significantly from the expected behavior of the machine. Therefore, for each operating condition and health indicator, an expected behavioral model is fitted. An alarm is raised when an indicator value is β number of standard deviations away from the behavioral model, where we define β to be the alarm level. In accordance with Eq. 5.5, for a linear model $f(\cdot)$ with noise parameter σ_n, associated with a specific operating condition and health feature, the alarm level β is defined as follows:

$$\beta_t = \left\lfloor \frac{|y_t - \mathbb{E}[f(x_t)]|}{\sqrt{\mathbb{V}[f(x_t)]} + \sigma_n} \right\rfloor \tag{5.6}$$

where x_t, y_t are in our case respectively the observed operational parameters and constructed feature at time t, and $\mathbb{E}, \mathbb{V}$ are the expectation and variance operators. The expected value and variance of f can be directly computed from the posterior weight parameters w in Equation 5.5. Note that the denominator includes both the observational noise and the uncertainty of the model parameters, which is important as mentioned previously.

To provide intuitive anomaly indicators, we set three alarm levels: green ($\beta < 2$), yellow ($2 \leq \beta < 4$) and red ($\beta > 4$), which respectively have values 0, 0.5 and 1. These thresholds can be chosen based on qualitative investigations of the data. Additionally, we remove outliers in the alarm time series by using a sliding windows and keeping the median alarm level.

As a final step in our approach, we reduce the workload of the early failure analysis by compressing the large set of anomaly alarms to a smaller *fused* set per indicator type. For statistical time-domain indicators, this step can improve the robustness of the alarms since they are often expected to show similar properties of the potential failures.

Then, the maximum over all measurement channels per indicator type is taken. This provides a fast warning that an anomaly is present on any measurement channel. The fused alarms are values between 0 and 1, which can be perceived as a severity indicator. For example, if a fused alarm is close to 1, lots of features agree on the alarm level, which makes the occurrence of an anomaly more likely.

5.5.3 Experimental application example

To give an example of how we use the described automation approach, an experimental example is presented in which we process data of four different wind turbines. *Due to confidentiality reasons, the amount of details is kept very concise.*

Vibration sensors are distributed over the wind turbine drivetrains. Data is sampled above 25kHz and acquired for approximately 10 seconds at intermittent times each day. Eight vibration measurement channels are processed using the tools of different complexity discussed in the previous section. In this example, only six statistical time-domain indicators are calculated on the raw sensor data: RMS,

kurtosis, peak-to-peak, crest factor, peak energy index, and Moors kurtosis. In the future we plan to expand this analysis to include both the time and frequency domain indicators discussed in sections 5.2 & 5.3. In total 60 different processing pipelines, of the type discussed in the previous section (cepstrum liftering and frequency filtering), are computed to provide enhanced failure signatures. Subsequently, statistical indicators are derived for each of these pipelines. To allow data-trending, intermittent data samples for a period of 5 years are processed.

We target the detection of a bearing issue known to be present in the slow rotating stage of one of the four wind turbines. The turbines are in each other's vicinity in the farm and believed to be loaded in a similar way. As such indicator trends are compared to illustrate deviations in behavior. Figure 5.10 shows the comparison between the Moors kurtosis of the four turbines after the same processing pipeline. Similarly, Fig. 5.11 shows the results for the crest factor after processing.

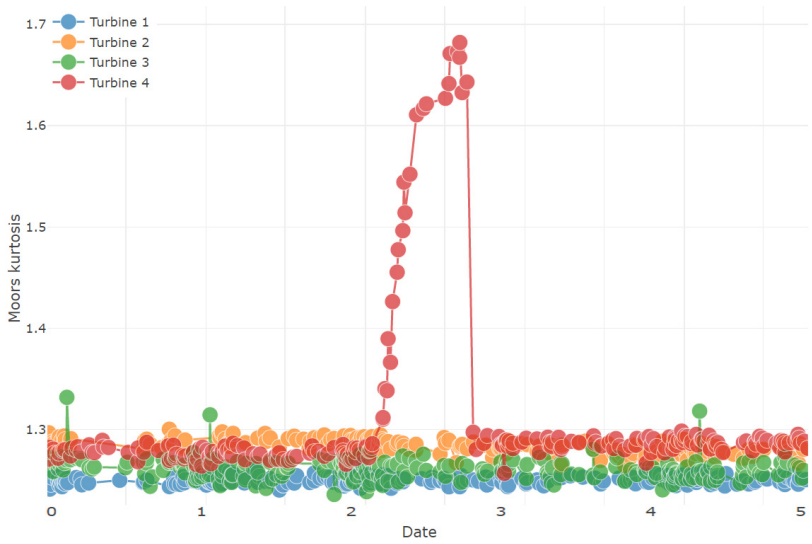

FIGURE 5.10: Comparison of Moors Kurtosis-based indicators for the four wind turbines over a period of 5 years.

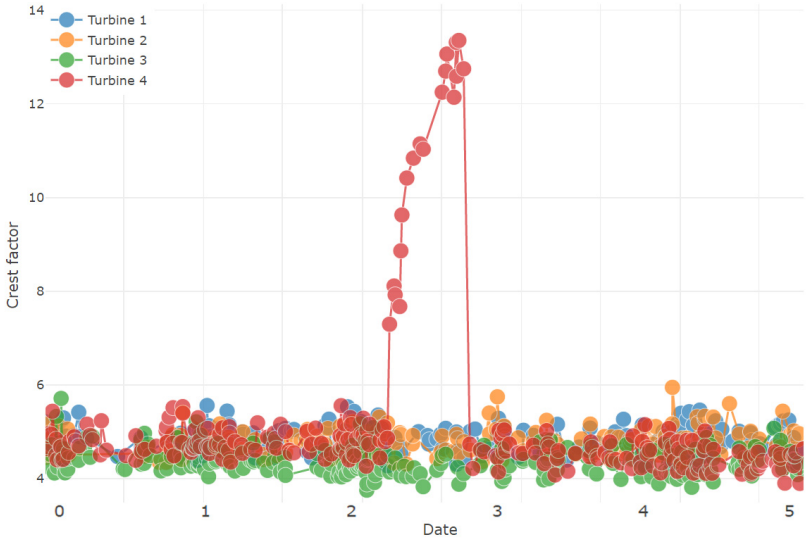

FIGURE 5.11: Comparison of crest factor-based indicators for the four wind turbines over a period of 5 years.

The fault condition in the fourth turbine is clearly visible in all indicators after processing based on the comparison. Detection is achieved approximately 6 months prior to repair. Additionally, the moment of repair is clearly identifiable. Using the neighbor turbine as a reference shows potential. However, to improve this comparison it is better to remove the DC values of the indicators during healthy conditions.

To illustrate the impact of using different processing pipelines, Fig. 5.12 displays the RMS after using two different filters (and thus pipelines) for the fourth turbine. To maximize the detection potential, it is thus recommended to look simultaneously at multiple pipeline outcomes.

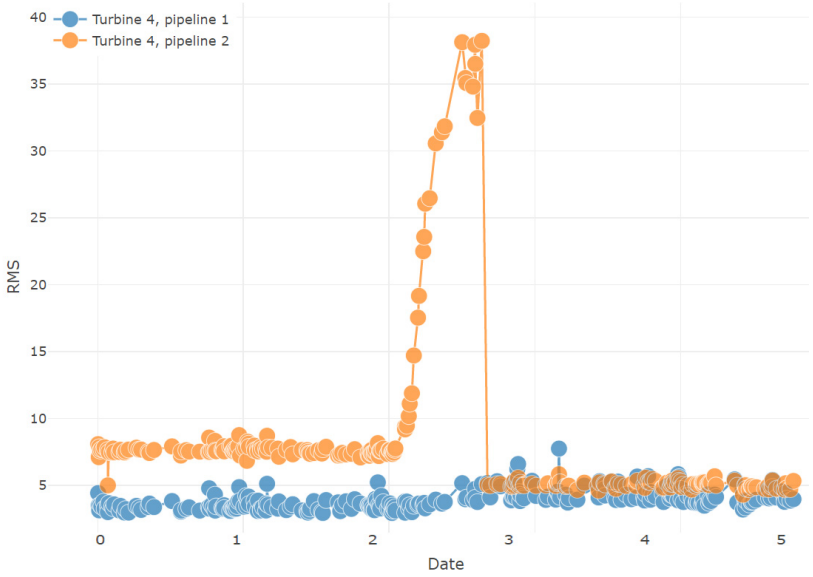

FIGURE 5.12: Comparison of RMS-based indicators with different processing pipelines.

The indicators calculated for the four turbines are also used as inputs for the anomaly detection based on the fused indicators. To define the operational regime of the wind turbine, we use the active power production and rotational speed. Based on the popular silhouette score measure [329], the number of operational regimes (clusters) is set to 4. The statistical indicators are modeled using the Bayesian expected behavior model. Fused alarms are computed by taking the average alarm level over all features per statistical indicator type (e.g. crest factor). This fusion is carried out over all features, which results into 6 fused alarms (one per indicator type) and thus significantly reduces the workload of the analysis.

The effects of the failure on the behavior of the statistical features can be observed in the fusion, reaching an alarm severity of about 70% for the fused moors kurtosis indicator seen in Fig. 5.13 and 100% for the crest factor in Fig. 5.14. Thresholds can be put in place by the operator to alert for potential failures automatically. It can be seen that the alarm severity approximately starts to increase significantly in the

third year of measurements, which corresponds with the raw statistical indicators. After the repair of the bearing, the fused indicators are still slightly raised above the DC alarm level. This can be attributed to the fact the ML model was not retrained after the repair. Such a retraining is necessary when machine components are replaced.

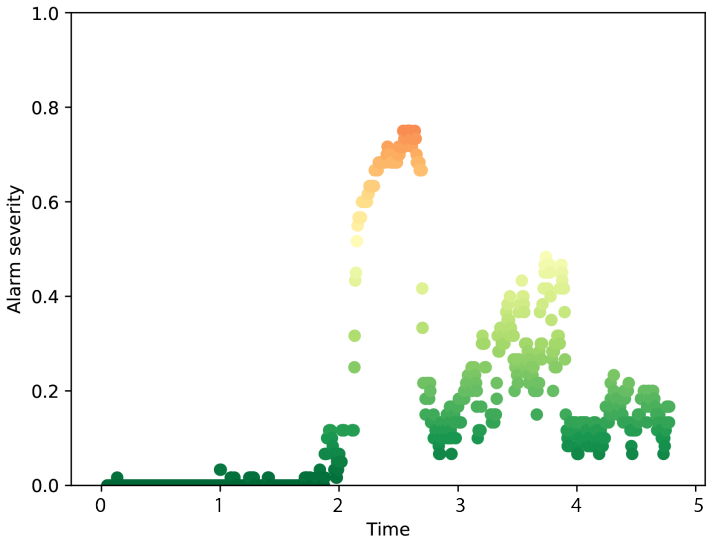

FIGURE 5.13: Fusion of moors kurtosis indicators.

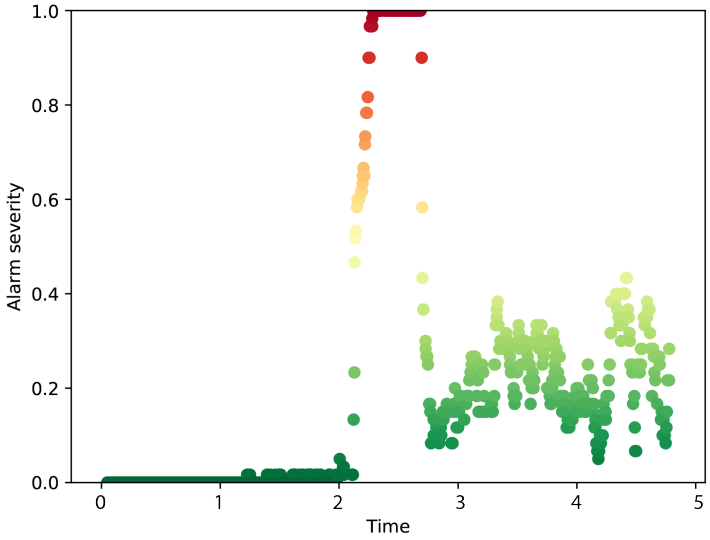

FIGURE 5.14: Fusion of crest factor indicators.

5.6 Conclusion

This chapter presented a state-of-the-art overview of vibration analysis methods for fault detection. The general classes of time domain and frequency domain indicators are discussed together with potential additional processing steps such as frequency filtering. The concepts of envelope analysis and cyclostationarity are summarized since they play an essential role nowadays in vibration-based condition monitoring of rotating machinery. Also, an alternative approach to the kurtogram, called the harmonogram, is proposed if prior knowledge about the characteristic frequency is available. It employs the squared envelope spectrum of the filtered signal in each filterband of the decomposition tree to calculate a fault feature on the harmonics of interest. Despite being a fairly straightforward extension of the kurtogram concept, it can provide a useful tool for finding an optimal demodulation frequency band for a specific fault.

Lastly, a data-driven automated indicator processing methodology, developed in collaboration with the AI department of VUB, is presented. It allows reducing the vast number of condition indicators coming from complex systems into a workable set of alarm severities. Currently, the main work that would still need to be done by an operator manually, involves choosing the right input parameters for the vibration processing methods and inspecting the summarized alarm severity figures. Future work will focus on further reducing the number of input parameter choices for the different methods.

Chapter 6

Experimental case studies

6.1 Introduction

So far, this dissertation discussed individual vibration signal processing methods. Brief illustrations of these methods on simulated and experimental data were given, highlighting their performance. This chapter expands the applied analysis in this dissertation with two more investigations of experimental data sets. These examinations employ multiple of the different techniques mentioned in the previous chapters. The two data examples in this chapter serve as an additional means for giving a better insight into the different ways the discussed methods can be utilized in a practical setting. Showcasing the methods by doing a more in-depth analysis of experimental data also provides a good summary of the application potential.

The two data sets in this chapter are both of wind turbine gearboxes. The first one is the well-documented data set from the National Renewable Energy Laboratory (NREL) in the context of a wind turbine gearbox condition monitoring round robin study. An investigation of this data set was conducted using multiple processing methods, including order tracking, cepstrum editing, the kurtogram, and envelope analysis. This study is described in section 6.2. The second data set comes from an operational multi-megawatt wind turbine that experienced a bearing failure. Due to confidentiality reasons, most of the technical details of the gearbox and measurement setup are omitted. Some of the other techniques discussed in the dissertation are illustrated on this data set, including the multi-harmonic demodulation, the statistical indicator tracking using a filterbank approach, and

the cyclic spectral coherence. This second study is described in section 6.3.

6.2 Case study 1: NREL round robin data set

In order to illustrate the benefit of a multi-step processing approach, experimental data of a wind turbine from the National Renewable Energy Laboratory (NREL) is examined and the results are compared to those summarized in the general report by Sheng [259]. In this report, a number of participating research partners published their findings [330–335] which form a valuable background for this study. The NREL dynamometer test facility (DTF) was used for the data collection.

6.2.1 Description of experimental setup

Turbine As can be seen on Fig. 6.1(a), the test turbine is a three-bladed, stall-controlled, upwind turbine. It has a rated power of 750kW and the generator normally operates at 1800 rpm or 1200 rpm nominal. The complete drivetrain is installed in the NREL DTF as shown on Fig. 6.1(b). It is hard fixed to the floor and misses the hub, rotor, yaw bearing and yaw drives.

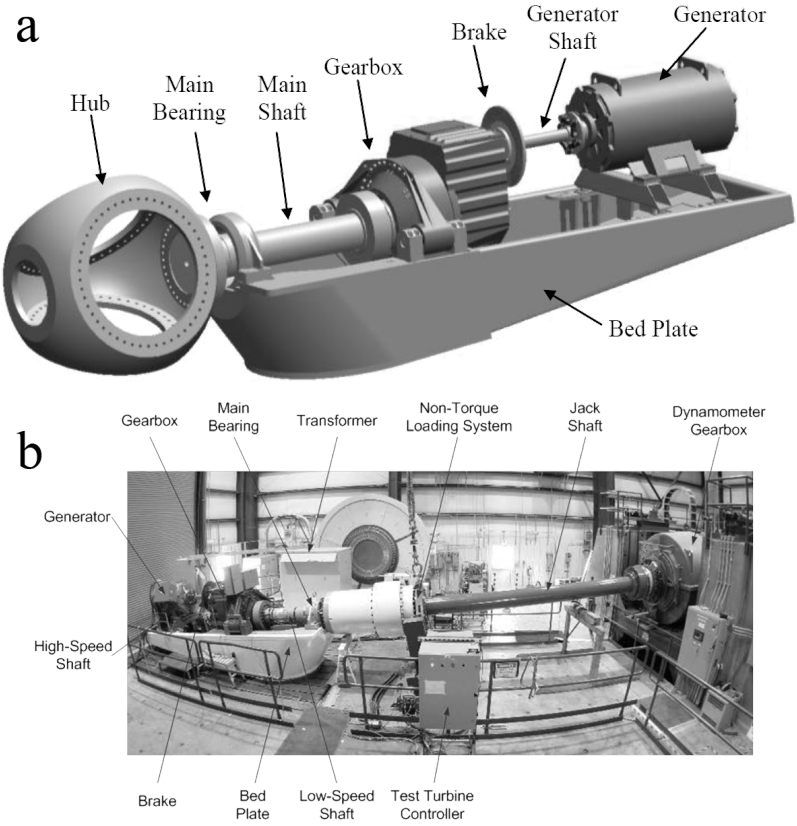

FIGURE 6.1: (a) Drive train configuration of the test turbine. (b) NREL dynamometer test stand with the turbine installed.

Gearbox Two gearboxes of the same design are used for collecting the data, one in "healthy" state and the other one in "damaged" state. An exploded view of the gearbox is shown in Fig.6.2(a). The accelerometer vibration measurements along with high-speed shaft RPM data was made available. The "healthy" gearbox is only tested in the dynamometer while the "damaged" gearbox first did a run-in in the dynamometer and was later sent to a wind farm for field testing. Two loss-of-oil events damaged its internal bearings and gears. Afterward, it was again installed in the dynamometer facility and retested in a

controlled environment. The gearboxes consist of one low speed (LS) planetary stage and two parallel stages. The used nomenclature is the same as the one used by Sheng [259] and can be seen in Fig. 6.2(b).

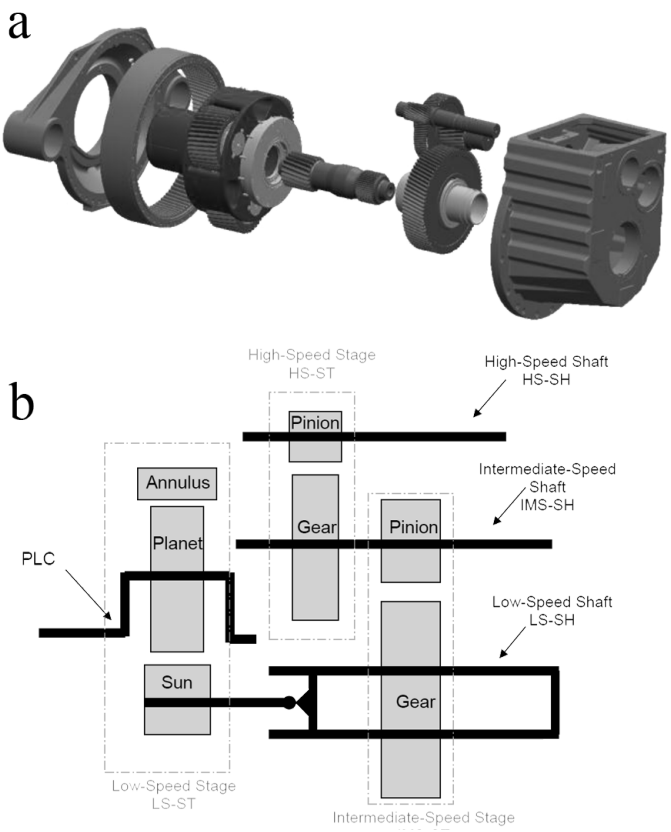

FIGURE 6.2: (a) View of internal components of the test gearbox. (b) Internal nomenclature and abbreviations of the test gearbox.

Bearings Various bearing types are used in the gearboxes corresponding to the loading conditions and life requirements. Two full-complement cylindrical roller bearings (fcCRB) support the planet carrier and two cylindrical roller bearings (CRB) support the planet gears. The parallel shafts are each supported by a CRB on the upwind side and by two tapered roller bearings (TRB) on the downwind side of the assembly.

A list of all bearing locations, manufacturers, part numbers and types can be found in Table 6.1 and Fig. 6.3 illustrates the locations and used names of the different bearings. If the component is positioned upwind, this is denoted with an 'A' and if it is downwind, with a 'B' or a 'C'. A full list of the gear and bearing characteristic frequencies is not given but can be determined based on Table 6.1.

TABLE 6.1: Overview of used bearing types, numbers and locations.

Location	**Location designation**	**Type**	**Provider**	**Part number**
Planet carrier	PLC-A	fcCRB	INA	SL181892E
	PLC-B	fcCRB	INA	SL 18 1880 72/K10
Planet	PL-A	CRB	FAG	NJ2232E.M1
	PL-B	CRB	FAG	NJ2232E.M1
Low-speed shaft	LS-SH-A	fcCRB	INA	SL181856E
	LS-SH-B	TRB	SKF	32948*
	LS-SH-C	TRB	SKF	32948*
Intermediate-speed shaft	IMS-SH-A	CRB	FAG	NU2220E.M1
	IMS-SH-B	TRB	SKF	32032 X
	IMS-SH-C	TRB	SKF	32032 X
High-speed shaft	HS-SH-A	CRB	FAG	NU2220E.M1
	HS-SH-B	TRB	SKF	32222 J2
	HS-SH-C	TRB	SKF	32222 J2

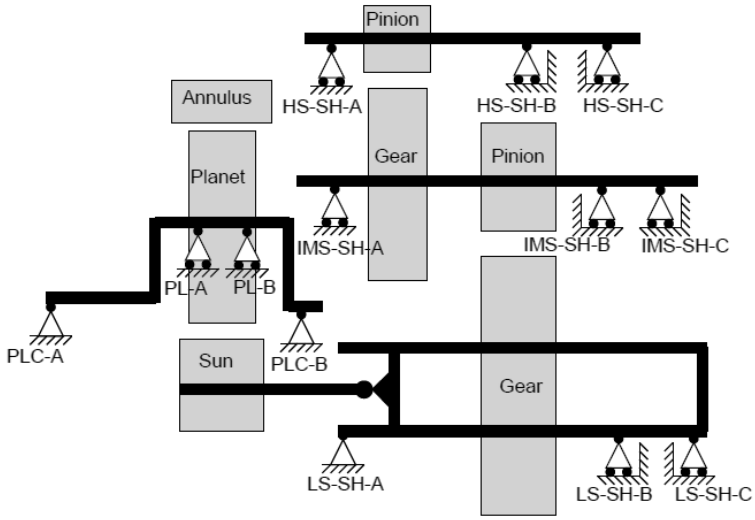

FIGURE 6.3: Locations and nomenclature of the bearings in the test gearbox.

6.2.1.1 Measurement settings

The accelerometers are mounted on the outside of the gearbox and data is sampled at 40 kHz per channel. In total data sets of eight accelerometers (Model: IMI 626B02) are made available and the exact locations can be seen in Fig. 6.4. Lastly, Table 6.2 gives a summary of the signal names and their corresponding locations of the different vibration sensors.

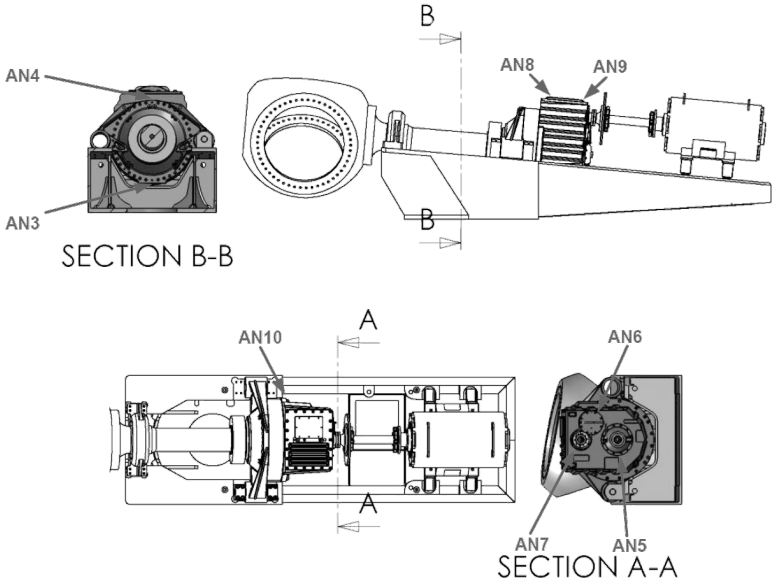

FIGURE 6.4: Overview of the accelerometer locations.

TABLE 6.2: List of the used sensors and their corresponding placement descriptions.

Sensor label	**Location description**
AN3	Ring gear radial 6 o'clock
AN4	Ring gear radial 12 o'clock
AN5	LS-SH radial
AN6	IMS-SH radial
AN7	HS-SH radial
AN8	HS-SH upwind bearing radial
AN9	HS-SH downwind bearing radial
AN10	Carrier downwind radial

6.2.2 Analysis procedure

In order to assess the performance of the developed methods for bearing fault detection, the provided signals at 1800 rpm from the NREL test setup are investigated using a combination of the following techniques:

1. Speed estimation and angular resampling
2. Automated cepstral editing procedure (ACEP)
3. Bandpass filtering based on kurtosis
4. Squared envelope analysis

According to the report of [259], the research partners of the *Wind Turbine Gearbox Condition Monitoring Round Robin Study* agreed to a total of seven damages that were considered to be detectable through vibration analysis. Table 6.3 shows a summary of the detectable gearbox damage.

TABLE 6.3: Actual gearbox damage that should be detectable through vibration analysis.

Damage #	Component	Mode
1	HS-ST gear set	Scuffing
2	HS-SH downwind bearings	Overheating
3	IMS-ST gear set	Fretting corrosion, scuffing, polishing wear
4	IMS-SH upwind bearing	Assembly damage, scuffing, dents
5	IMS-SH downwind bearings	Assembly damage, dents
6	Annulus/ring gear, or sun pinion	Scuffing and polishing, fretting corrosion
7	Planet carrier upwind bearing	Fretting corrosion

This analysis focuses on bearing fault detection, so no analysis is done for the gear faults, meaning that only the four bearing faults are investigated. For each bearing, the characteristic frequencies corresponding to the damages are shown in Table 6.4.

TABLE 6.4: Damaged bearings and their corresponding theoretical characteristic frequencies. *(BPFO: Ball Pass Frequency Outer race, BPFI: Ball Pass Frequency Inner race, FTF: Fundamental Train Frequency)*

Bearing label	**Fault type**	**Detectable characteristic frequency [Hz]**
HS-SH downwind bearings (*HSS-B& C*)	BPFI FTF	345,29Hz 12,75 Hz
IMS-SH downwind bearings (*ISS-B& C*)	BPFO	105,75 Hz
Planet carrier upwind bearing (*PLC-A*)	BPFO	8,81 Hz
IMS-SH upwind bearing (*ISS-A*)	BPFI	73,7 Hz

6.2.3 Analysis results

The bearing faults present in the system are spatially distributed over the gearbox, meaning that some sensors will be more likely to detect possible faults than others due to being in closer proximity to the source. To illustrate this phenomenon and the used processing steps, the signal measured by sensor AN7 is analyzed. This sensor is located near the high-speed shaft bearings, so it can be expected to contain prominent signal content originating from the BPFI and FTF faults described in Table 6.4.

6.2.3.1 High-speed downwind bearings

The first step in the processing scheme is to resample the healthy and the damaged data so the speed variation present in the measurements is neutralized. Using the provided tacho signal for the healthy signals and the RPM measurements for the damaged signals, all the signals are resampled to the angular domain to provide a clearer view of the frequency components present in the data. Figure 6.5 shows a zoomed picture of the 18^{th} high-speed shaft harmonic, around 540 Hz, before and after angular resampling of the damaged signal. It can be seen that the smearing of frequency peaks in the spectrum due to

speed variation, is greatly reduced after resampling. Only one peak remains at a multiple of the fundamental shaft speed frequency.

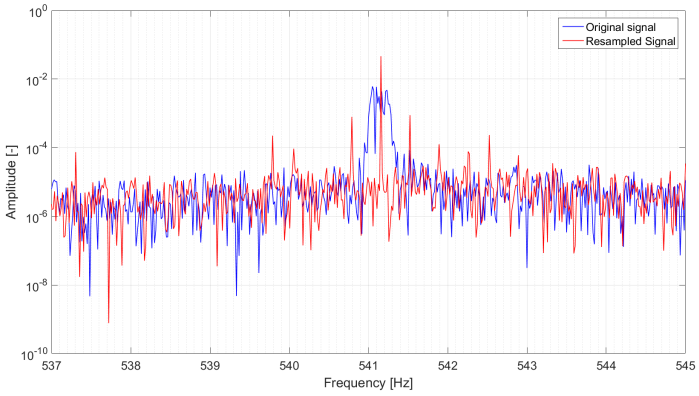

FIGURE 6.5: Zoom of the 18th harmonic of the 30 Hz shaft speed, measured by the AN7 sensor for the damaged gearbox, before and after angular resampling.

Next, automatic cepstral editing is performed on the signal in order to remove the deterministic content of the resampled signal as much as possible. The ACEP method described in Chapter 3 is employed and the spectra of the damaged data of the AN7 sensor before and after ACEP is shown in Fig. 6.6a. The damaged data exhibits very prominent shaft speed harmonics and due to the difficulty in properly displaying the total amount of reduction of harmonic content in a spectrum, a graph of the amplitude reductions of the first 100 shaft speed harmonic is presented in Fig. 6.6b. It can be seen that the ACEP method is able to significantly reduce the shaft speed frequency peaks by an average amount of about 13 dB.

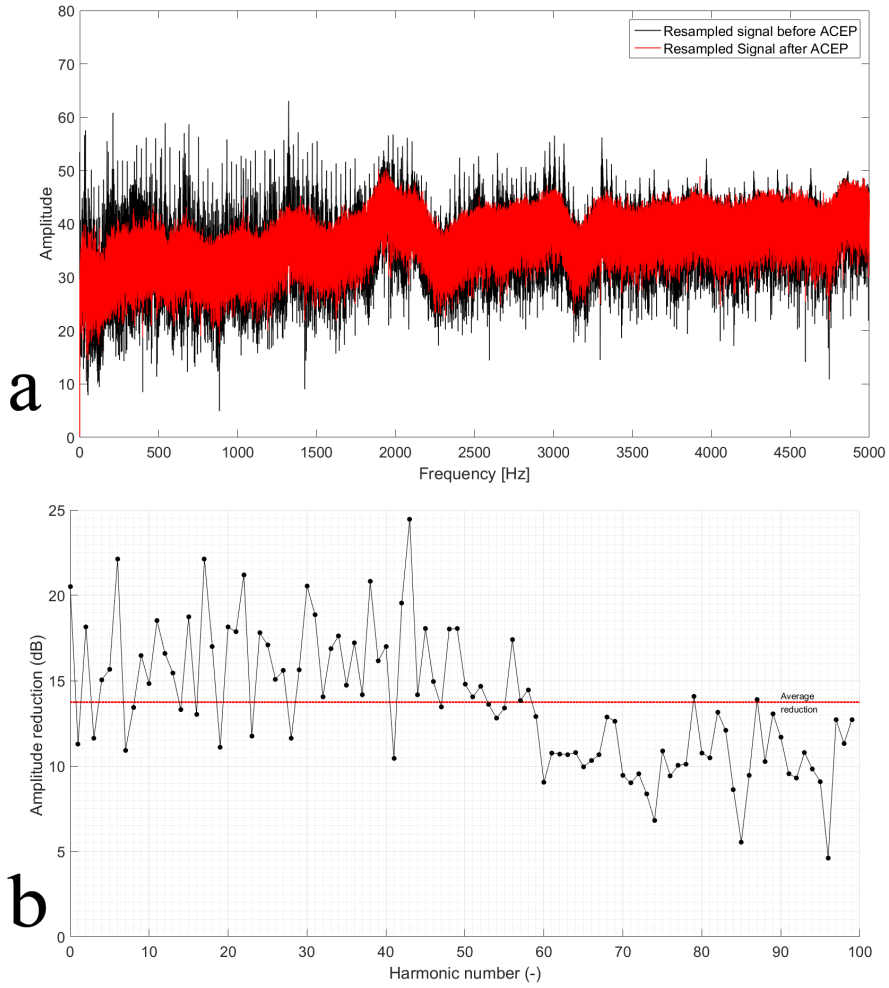

FIGURE 6.6: (a) Amplitude spectrum of the AN7 sensor signal before and after ACEP. (b) Graph showing the amplitude reductions, in dB, of the first 100 shaft speed harmonics.

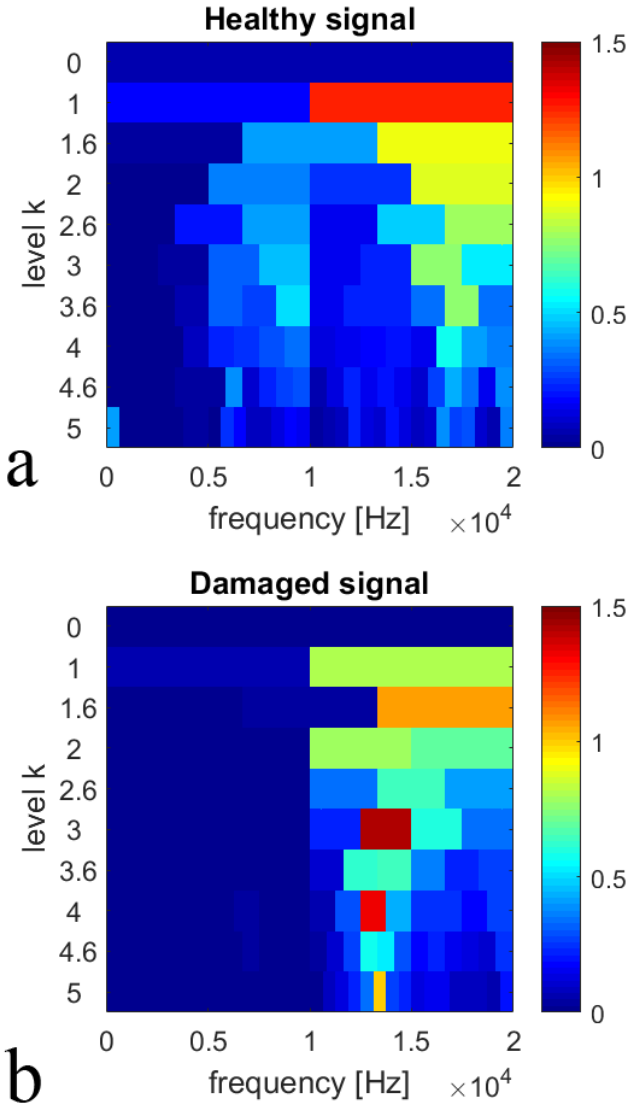

FIGURE 6.7: (a) Kurtogram of the healthy AN7 sensor data. (b) Kurtogram of the damaged AN7 sensor data.

The residual signals after ACEP can now used as input for the kurtogram to determine an optimal demodulation frequency band. The

kurtograms of both the healthy and damaged data are generated and examined for differences. Figure 6.7 displays the calculated kurtograms and here it can be seen that there is an increase in kurtosis around 13,2 kHz. First however, the normalized envelope spectra of the healthy and damaged data are examined without the use of the kurtogram, thus without bandpass filtering. The result of this can be seen in Fig. 6.8a. While the ACEP technique reduces the deterministic influence of the shaft speed, a prominent peak can be seen at the high-speed shaft frequency ($1\times HSS$) in the envelope spectrum. This is expected since ACEP only removes additive components and not modulating ones. Typically, impulses at the BPFI fault frequency are also modulated with the shaft speed, thus they can also serve as a sign of increased bearing wear. Multiple peaks at BPFI harmonics of 345 Hz can be observed for the healthy as for the damaged data. The fundamental 345 Hz peak is close to the 2^{nd} harmonic of the sun gear mesh frequency and the 46^{nd} harmonic of the intermediate shaft speed frequency, but is not an exact multiple of one of these frequencies. Since this component is so prominently present after cepstral editing, it is concluded that the peak in the healthy and damaged data corresponds to the inner race defect frequency or BPFI of the high-speed downwind bearings, which is theoretically around 345.29 Hz for a shaft speed of 30 Hz. This is supported further by the presence of harmonics of the BPFI and side bands at 30 Hz in the envelope spectrum. Based on these envelope spectra, it appears though that the inner race damage was already present in the baseline healthy data.

While the gear faults are not the focus of this analysis, it is noted that in Fig. 6.8a, the envelope spectrum displays a high value at 661 Hz for the damaged data. This corresponds to the gear mesh tooth passing frequency (*GMF*) of the intermediate-speed shaft (ISS) with the high-speed shaft. The damage report verified that the ISS gear exhibited corrosion, scuffing and wear.

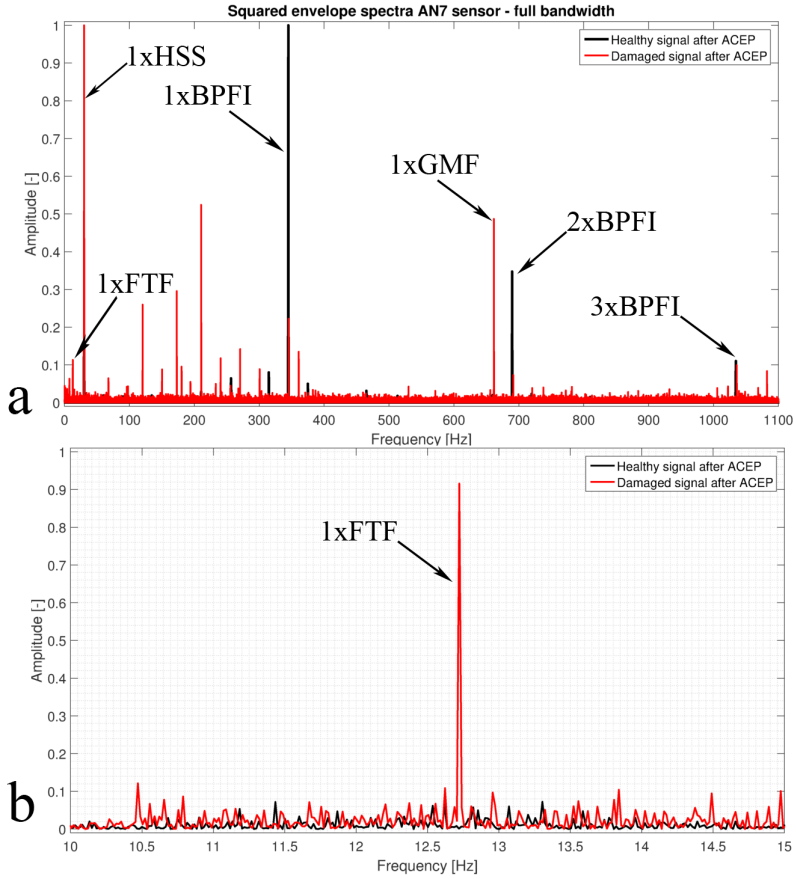

FIGURE 6.8: (a) Envelope spectrum generated without the use of filters/kurtogram, indicating the presence of the BPFI & FTF of the high-speed downwind bearings (HSS-B& C). (b) Zoom of the envelope spectrum around the FTF frequency for the healthy and damaged signals after applying ACEP and kurtogram.

The next analysis step makes use of bandpass filtering, based on the info portrayed by the kurtograms in Fig. 6.7. Figure 6.8b shows a zoom of the envelope spectrum of the residual signal after bandpass filtering. Examination of the spectrum reveals a high peak around 12.71 Hz, which corresponds to the fundamental train frequency (FTF)

of the high-speed downwind bearings. Figure 6.8b shows a clear increase of this frequency in the damaged data compared to the healthy data. Finally, Fig. 6.8a also shows that the 30 Hz HSS modulation has increased significantly for the damaged data set. This can indicate a possible high-speed shaft imbalance or misalignment or it can be associated with the BPFI modulation of the HSS-B&C bearings. Since both the presence of an imbalance and of an inner race fault was later corroborated by the damage report, this component is most likely the combination of both faults.

In order to demonstrate the effect of the ACEP procedure on the envelope spectrum, the damaged data is investigated before and after applying cepstral editing. Figure. 6.9a shows the envelope spectrum of the damaged data of the AN7 sensor without any filtering before and after cepstral editing. The spectrum before ACEP exhibits far more harmonic peaks, making it difficult to assess possible bearing faults. Additionally, these harmonics mask the presence of bearing faults, as is shown in Fig. 6.9b. The BPFI of the HSS downwind bearings cannot be detected here before cepstral editing.

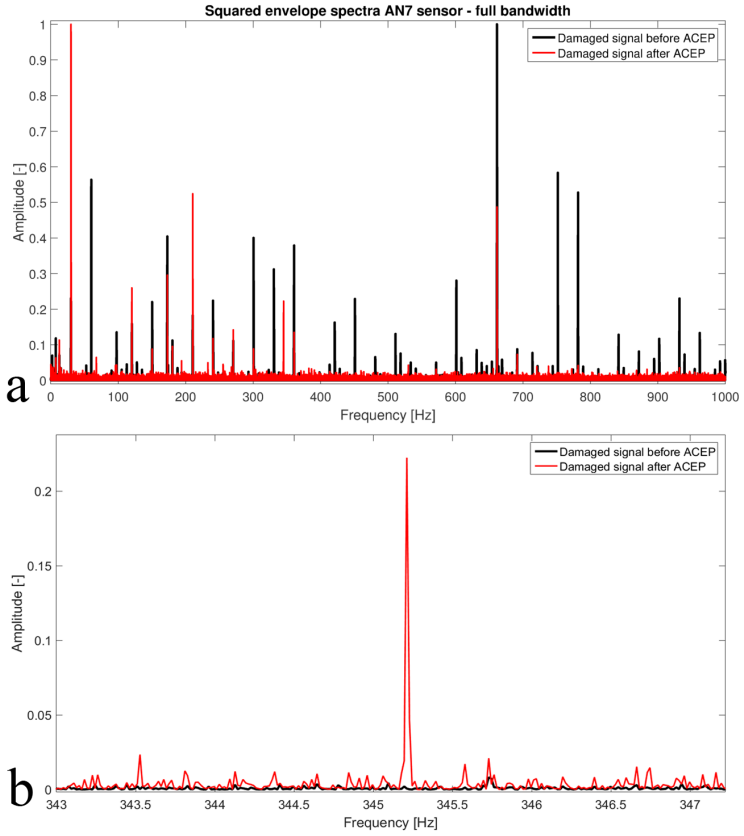

FIGURE 6.9: (a) Envelope spectrum of damaged data before (*black line*) and after ACEP (*red line*). (b) Zoom of envelope spectrum around 345 Hz, the BPFI frequency of the high-speed downwind bearings, before and after ACEP.

6.2.3.2 Intermediate-speed shaft downwind bearings

The analysis procedure followed for the other bearings is similar to the one explained for the high-speed bearings. To reduce the number of figures, only the resulting envelope spectra are shown. As such, the analysis of the AN6 sensor, which is located close to the intermediate-speed shaft (ISS), is investigated for possible ISS bearing faults. After angular resampling and cepstral editing, the full-bandwidth envelope spectrum for the damaged case, shown in Fig. 6.10, exhibits clear peaks

at even harmonics of the BPFO(105.25 Hz) of the ISS-B&C bearings. It does not however offer clear indications of the fundamental defect frequency itself. The kurtogram is used for determining the bandpass filter cut-off frequencies and is shown in Fig. 6.11. Using a bandpass filter centered at 16.4 kHz with a bandwidth of approximately 1 kHz produces the envelope spectrum of Fig. 6.12. This bandpass filter corresponds to the yellow rectangle in the kurtogram figure. In this envelope spectrum, the fundamental fault frequency is clearly revealed. However for the greater part of the full frequency band of the signal, mainly the even harmonics of the BPFO were discernable. This could imply that there are multiple point defects on the bearings, introducing cancellation and reinforcement of components which can modify the envelope signature and thus also the envelope spectrum [260]. The damage report confirms this and describes multiple dents, assembly damage and plastic deformation of the ISS downwind bearings. A zoom of the second BPFO harmonic at 210.4 Hz can be seen in Fig. 6.13 and illustrates the increase in amplitude of this frequency compared to the healthy case.

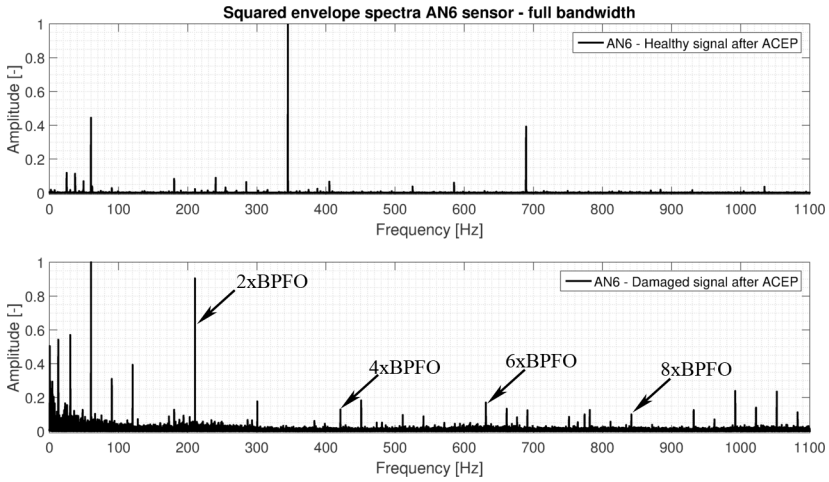

FIGURE 6.10: Envelope spectrum of AN6 acceleration signal after resampling and ACEP for the healthy case (*above*) and damaged case (*below*).

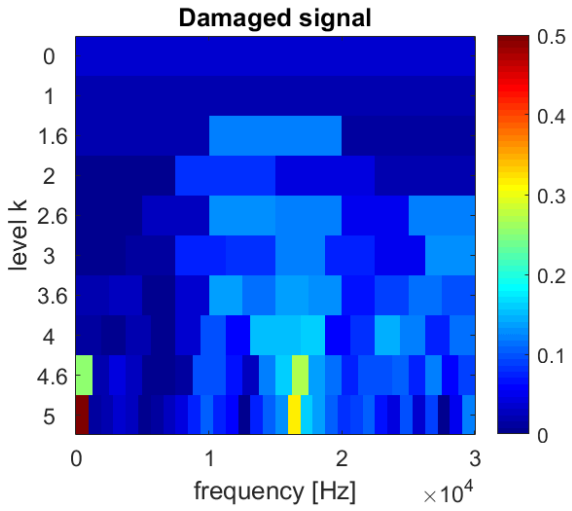

FIGURE 6.11: Kurtogram of the AN6 signal after automated cepstrum editing.

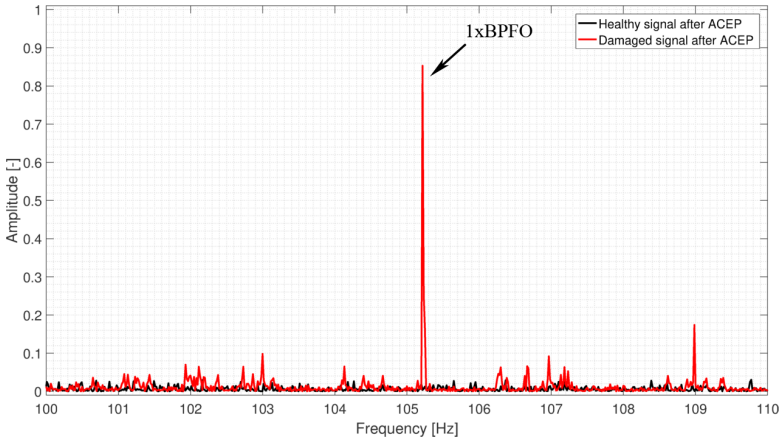

FIGURE 6.12: Zoom of envelope spectrum of AN6 acceleration signal around the BPFO fundamental of the ISS-B&C bearings for the healthy case (*black*) and damaged case (*red*).

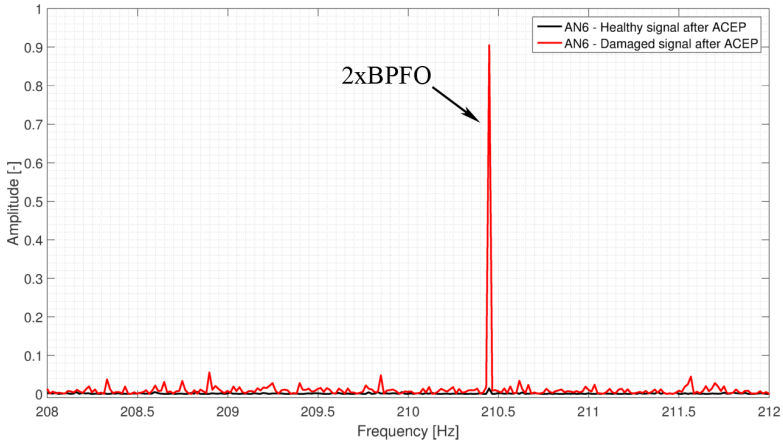

FIGURE 6.13: Zoom of envelope spectrum of AN6 acceleration signal around the second BPFO harmonic of the ISS-B&C bearings for the healthy case (*black*) and damaged case (*red*).

6.2.3.3 Intermediate-speed shaft upwind bearing

Damage was also reported for the ISS upwind bearing and evidence of this damage was found by using a band-pass filter with a center frequency of 10 kHz and bandwidth 1 kHz. The envelope spectrum in Fig. 6.14 exhibits a high peak at 72.94 Hz, which is close to the theoretical BPFI of 73.7 Hz for the ISS upwind bearing. The slight deviation of the theoretical value is likely due to slip of the roller elements. While this component is quite prominently present in the AN6 and AN7 data set, it is easily mistaken for the second harmonic of the gear mesh tooth passing frequency of the planet gears, which is 73.05 Hz. However, due to its presence after ACEP and the small frequency difference and due to it being only in the damaged data, it can be concluded that it is, in fact, the BPFI of the ISS upwind bearing.

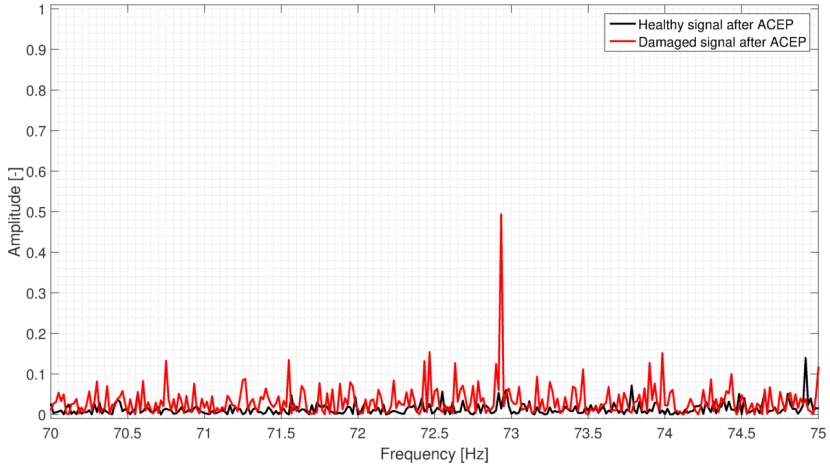

FIGURE 6.14: Zoom of envelope spectrum around 72.94 Hz, the BPFI of the ISS upwind bearing, after bandpass filtering with center frequency 10 kHz and bandwidth 1 kHz.

6.2.3.4 Planet carrier upwind bearing

A similar approach as for the ISS-B & C bearings is used for detecting the 8,8 Hz outer race defect frequency of the planet carrier upwind bearing (PLC-A). A zoom of the full-bandwidth envelope spectrum is shown in Fig. 6.15. While the damage report does state that there was fretting corrosion present on the PLC upwind bearing, the corresponding frequency peaks are small and difficult to detect. The partners who participated in the wind turbine gearbox condition monitoring round robin study confirmed this observation as well. Almost no one of the partners was successful in detecting this damage blindly and from Fig. 6.15a it can be understood why.

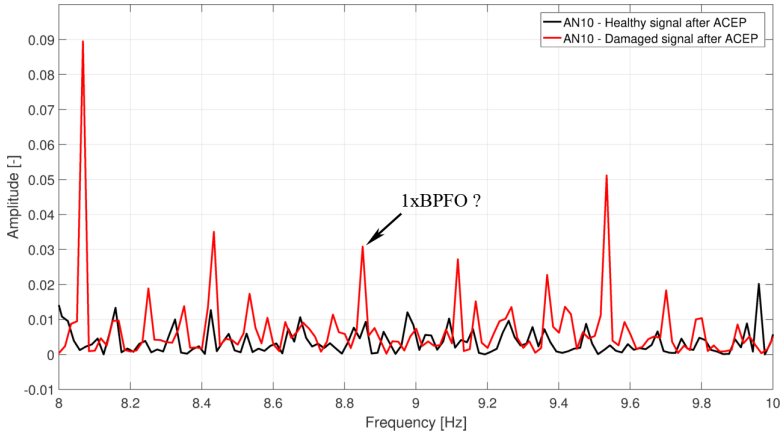

FIGURE 6.15: Zoom of the envelope spectrum around 8.85 Hz, the BPFO of the PLC upwind bearing.

6.2.4 Discussion case study 1

In order to concisely summarize the findings of this case study, the results of the bearing fault vibration analysis are presented again in table 6.5. For four out of five faults, there are clear indications of the fault in the envelope spectra. Only the planet carrier upwind bearing showed no significant frequency peak no matter which demodulation band was chosen. This corresponds to the findings of other research partners that reported similar results. The planet carrier bearing fault was considered to be difficult to detect through standard vibration analysis, given the used measurement setup.

TABLE 6.5: Overview of the results of the bearing fault vibration analysis

Bearing label	**Fault type**	**Detection confidence level**	**Observations**
HS-SH downwind bearings (*HSS-B* & *C*)	BPFI FTF	High High	Fundamental frequencies and harmonics are clearly visible, even for full bandwidth demodulation
IMS-SH downwind bearings (*ISS-B* & *C*)	BPFO	High	Full bandwidth shows even BPFO harmonics clearly, kurtogram is necessary for finding fundamental frequency
Planet carrier upwind bearing (*PLC-A*)	BPFO	Low	Envelope spectrum peak corresponding to BPFO is too small to make any claims about fault presence
IMS-SH upwind bearing (*ISS-A*)	BPFI	Medium	After bandpass filtering, the fundamental frequency is visible

While the mentioned bearing frequencies are deemed detectable through vibration analysis, the fault frequencies originating from bearings mounted on the lower speed shafts are in general deemed to be more difficult to detect compared to those on higher speed shafts. For example, the inner race fault frequency and the fundamental train frequency of the high-speed shaft downwind bearings are quite simple to detect, since these frequencies are already clearly visible in the full-bandwidth envelope spectrum after cepstrum editing. In contrast, the detection of the outer race fault frequency of the intermediate-speed shaft downwind bearings is a bit less straightforward due to the lack of a fundamental frequency peak in the full-bandwidth envelope. The even harmonics however show high envelope spectrum peaks. After finding the right narrow frequency band, the fundamental frequency does also show up in the analysis. Band-pass filtering is also required for the detection of the inner race fault of the intermediate-speed shaft upwind bearing. The planet carrier upwind bearing, which rotates at the lowest speed of all the mentioned bearings, was reported to have fretting corrosion, but did not exhibit any significant signs of actual damage in the vibration signals.

Although it is not the focus of this analysis, high values for some of the gear fault frequencies are observed in the generated envelope spectra. A general remark –which other participants in the NREL round robin study stated as well– is that due the presence of multiple progressed faults of gears and bearings, the analysis is encumbered by the multitude of modulating components present in the envelope spectra. Most of the described faults produce second-order cyclostationary components which make it difficult to analyze them separately since most signal separation techniques, including the cepstrum editing procedure, only filter out the deterministic components and leave the modulating cyclostationary components intact. Thorough inspection of the envelope spectra is thus necessary to attribute the fault frequencies to the correct envelope peaks.

With regards to the performance of the cepstrum editing procedure in the deterministic component removal, it is found that the ACEP technique is well-suited for the application in this case study. The wind turbine contains a complex multi-stage gearbox, consisting of multiple shafts, which would make the removal of masking discrete frequency content by time synchronous averaging methods quite time-consuming since every shaft's influence would have to be removed separately. Even then, modulation sidebands would still remain present in the vibration signal. Using linear prediction filtering based on autoregressive models was also tested (not shown here) but failed to reduce the harmonics sufficiently in this particular case. The complex vibration signal consists of multiple deterministic signal components at various frequencies as well as noise. It was found that a simple autoregressive model fails to fit the deterministic signal part well enough to effectively prewhiten the signal for bearing fault detection. Attempts to optimize the model order as a function of kurtosis or using model order criteria such as the Akaike Information Criterion (AIC) [336] did not improve the results either. While not tested on this data, the SANC and DRS methods would have to take care with their choice of a proper signal delay. The correlation length of the bearing fault signals has to be smaller than the delay, while the correlation length of the discrete content should be longer. There is no guarantee this is the case, making the results potentially unreliable. The cepstrum editing procedure does not have these drawbacks or dependencies. No choice of delay

or model order has to be made and the method only requires one step without having to iterate over all the present shafts or gears. The main parameter that can be changed in the CEP method is the number of samples to put to zero in the liftering step. This amount of samples depends on the sample rate and the length of the signal and could potentially be calculated automatically as well. For the examined data it was found that varying this parameter did not change the results as long as the amount of samples was not chosen to be excessively large.

Overall it is found that the proposed combination of techniques increases the detection rate and sensitivity to faults while remaining robust to non-impulsive background noise and easy to implement. However, care should be taken in the presence of impulsive noise or other impulsive disturbances. Since the use of the kurtogram relies on the use of kurtosis as a means of determining the appropriate bandpass filter, this technique can be affected by impulsive disturbances and thus present misleading results.

While the CEP method does remove discrete frequency content, it still does not remove any other masking cyclostationary content. The inspection of the envelope spectra still needs to take into account the possible presence of other cyclostationary sources, as was observed during the analysis of the NREL data due to the multitude of faults.

6.3 Case study 2: wind turbine HSS bearing failure

This case study investigates the evolution of an outer race fault of a high-speed shaft bearing in a multi-megawatt wind turbine. The vibrations were measured using nine accelerometers distributed over the drivetrain. In total, about one year and a half of vibration measurements is analyzed. Approximately one measurement of ten seconds is performed every two to three days. Due to confidentiality restrictions, the amount of detailed information is limited.

The fact stipulated in Chapter 5 that the number of health indicators rapidly increases with the number of preprocessing techniques, is

illustrated for this case for a subset of calculated statistical time domain indicators. Besides this analysis, the performance of the multi-harmonic demodulation method is again highlighted for speed estimation. Lastly, it is shown that the cyclic spectral coherence in itself can also be used for determining a suitable demodulation carrier frequency band.

6.3.1 Speed estimation

To illustrate the structure of the vibration data, Fig. 6.16 shows the spectrogram over the full bandwidth with a normalized sample rate of 1 Hz, together with a zoom from 0 to 0.1 Hz. Looking at the zoom, it can be noticed that the harmonic structure is quite complex, with many harmonic orders and some being asynchronous to the rotation speed of the high-speed shaft.

Even though there is an encoder signal available, the exercise of vibration-based speed estimation is still carried out in order to get a better insight into the robustness of the novel multi-harmonic demodulation method on such data sets. By means of comparison, also the multi-order probabilistic approach is tested on this data. Figure 6.17 shows an example of the speed estimation of one 10-second signal (the actual rotation frequencies are offset by a certain amount in this figure). While the estimated speed by the MOPA method suffers from time and frequency resolution errors that were found difficult to remedy, the speed estimated by MHD is almost identical to the one of the encoder.

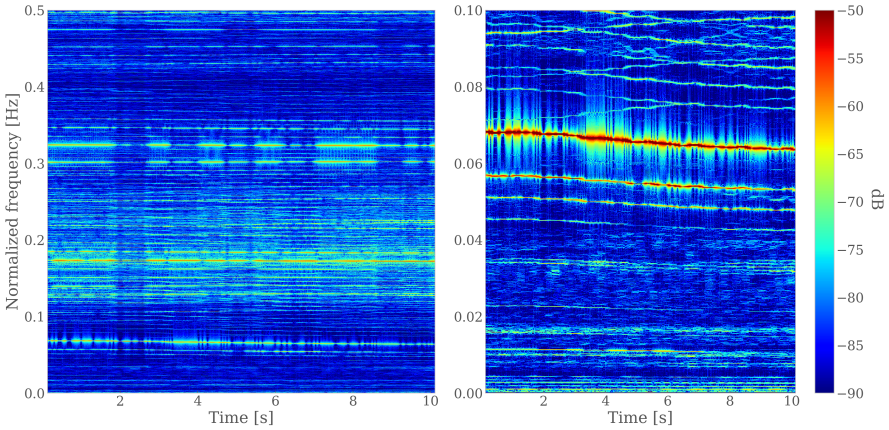

FIGURE 6.16: (Left) Spectrogram of the measured high-speed stage signal, (Right) Zoom of the spectrogram between 0 and 0.1 Hz.

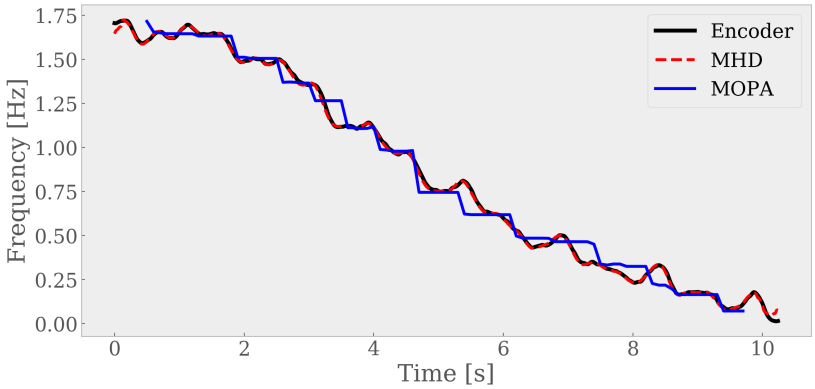

FIGURE 6.17: Comparison of the estimated speeds from the vibration signal with the MHD and MOPA method as compared to the encoder speed.

To showcase the maximum likelihood weighting of the MHD method, Fig. 6.18 displays the time-averaged harmonic weight evolution for the first 30 high-speed shaft harmonics used in the MHD method. It turns

out that the 24th, 20th, and 18th harmonics contribute the most to the speed estimation overall. The weights also converge quickly.

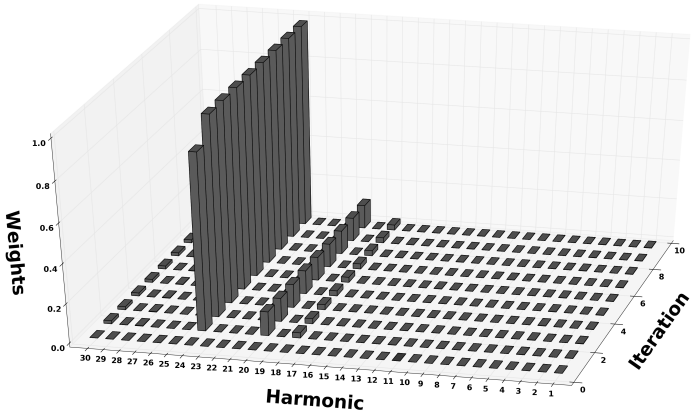

FIGURE 6.18: Evolution of the harmonic weights per iteration of the MHD method.

To get a sense of the overall performance of both methods, they are tested on all available measurements. The mean absolute deviation (MAD) is computed for the estimated speed of each data set. Figure 6.19 shows violin and box plots of the MAD distributions for both the MHD and MOPA method for all data sets. The accuracy of the MHD method is consistently better than the one of the MOPA method in this case. The MAD for the MHD method never goes above 0.1 % absolute error on the average speed estimation. In general, it can be concluded that the MHD method would be a good substitute for the encoder in case it would malfunction or be unavailable due to other circumstances.

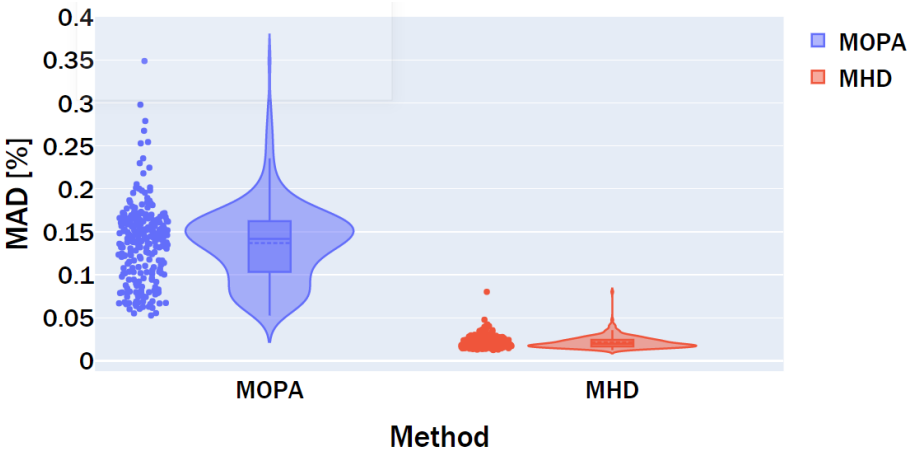

FIGURE 6.19: Violin and box plot of the distribution of mean absolute deviations of the MHD estimated speed compared to the encoder speed for all data sets.

6.3.2 Statistical filterbank trending

This section details how the approach described in section 5.2.1 can be employed for this case study. Several time domain indicators are calculated on the vibration signals after frequency filtering in different frequency bands. Additionally, the indicators are binned in five different active power production regimes, with 1 being the lowest power production regime and 5 being the highest. To keep the number of figures limited, only two example indicators are shown, namely the time negentropy and the Hoyer index.

Time negentropy Since it is difficult to show multiple dimensions at the same time, a parallel coordinates plot format is chosen to simultaneously show all computed indicators on 1 figure. Figure 6.20 shows the time negentropy values together with its independent variables, being the date, filterband and power bin. Every line thus represents one combination of the parameters. It can be observed from the graph that there is a clear increase starting around time 0.5 (dates have been anonymized for confidentiality reasons) and finishing around 0.9,

when the bearing got repaired. A selection is made in Fig. 6.21 of the top time negentropy values. Looking at this graph in Fig. 6.21, the highest values seem to occur in the lower frequency bands in the later stages of the damage evolution, while being fairly independent of the power production.

However, if we make a selection where we only include the initial increase in indicator values, as shown in Fig. 6.22, we can see that actually the higher frequency bands detect the rise in indicator value first. This seems to be more in line with the typical expectation that bearing faults manifest themselves initially in higher frequency resonance bands.

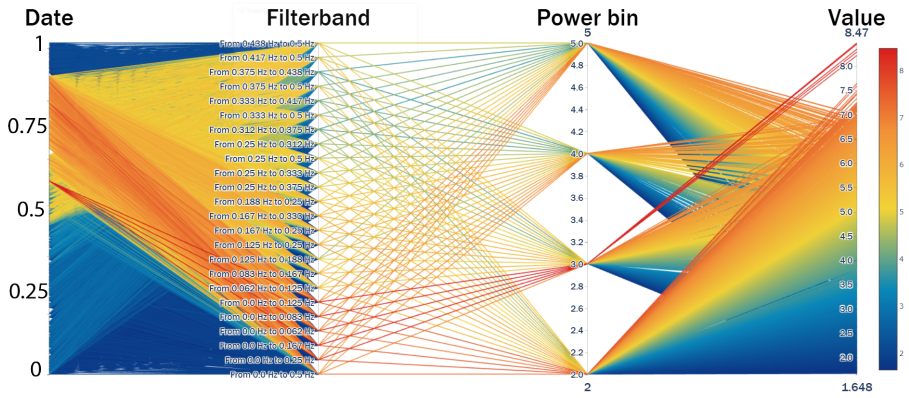

FIGURE 6.20: Full parallel coordinates plot of vibration signal time negentropy, depending on the time, filterband, and power production regime.

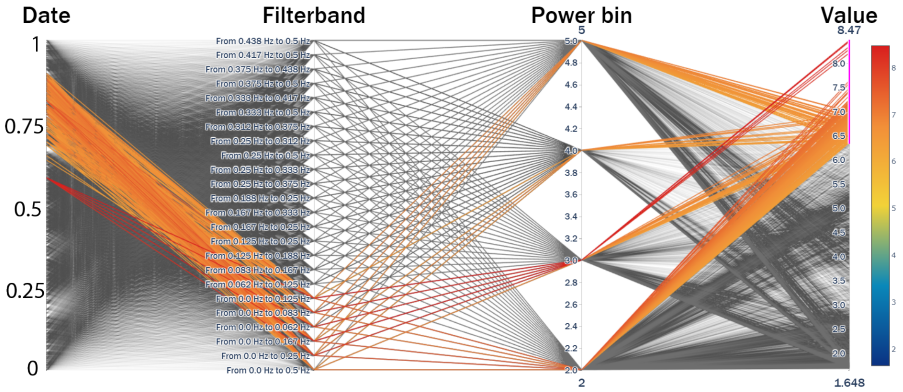

FIGURE 6.21: Selection of the top time negentropy values in the parallel coordinates plot. Note that the highest values seem independent of the power production, but related more to the low frequency filter bands.

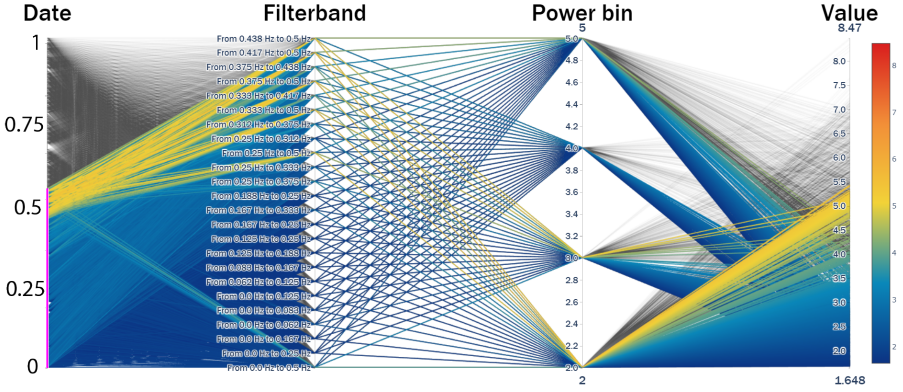

FIGURE 6.22: Selection of the first nine months of data to see which frequency filter bands indicate the first increase in time negentropy values. In this case the higher frequency bands exhibit the best sensitivity to the growing defect.

Lastly, a comparison is shown between an indicator evolution before and after automated cepstrum editing for a specific frequency

band (0 Hz - 0.125 Hz) and power bin regime in Fig. 6.23. For this case, the cepstrum editing seems to help improve both the stability of the indicator prior to damage initiation and the sensitivity of the indicator to the damage growth. While this is not always the case, it proves that cepstrum editing can also provide an added value to the trending of statistical indicators.

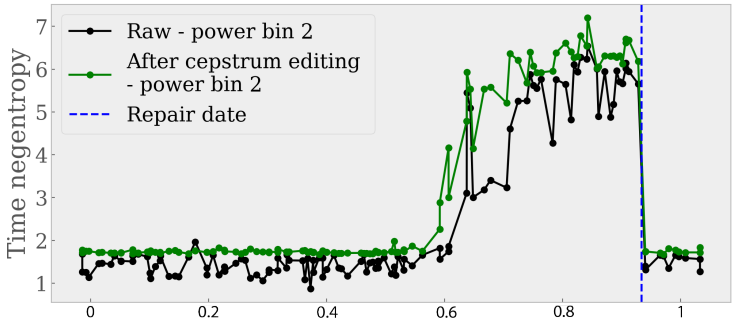

FIGURE 6.23: Time negentropy evolution in the lower frequency band [0 Hz - 0.125 Hz] and in power bin regime 2. Note that the cepstrum editing can improve the indicator sensitivity to faults.

Hoyer index The same analysis as for time negentropy is briefly repeated for the Hoyer index. Figure 6.24 shows the full parallel coordinates figure, which at first glance looks very similar to the one time negentropy. Selecting again only the top values for the Hoyer index, as displayed in Fig. 6.25, it can be seen that now only the higher frequency bands and the lower power production regimes show high Hoyer index values. The difference between the power bins in Hoyer index evolution over time is shown in Fig. 6.26. The second power bin showcases a slightly increased sensitivity to the change in the signal due to the damage initiation. In general, it was concluded that for this bearing fault case it was optimal to define high frequency band filters (in the upper half of the Nyquist frequency bandwidth).

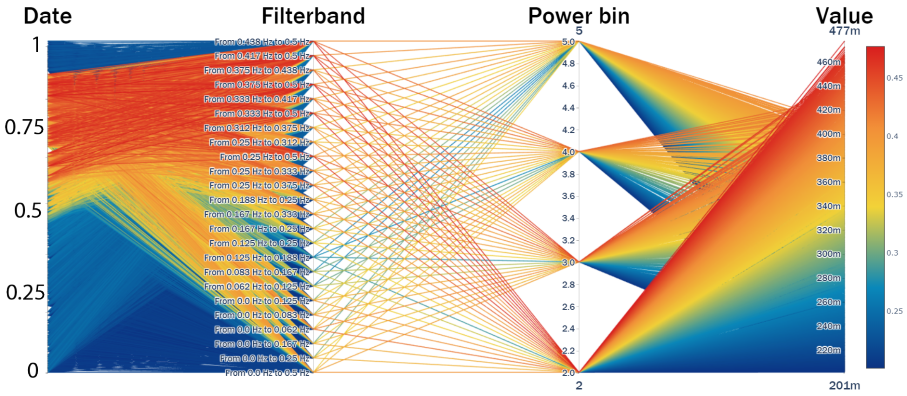

FIGURE 6.24: Full parallel coordinates plot of vibration signal hoyer index, depending on the time, filterband, and power production regime.

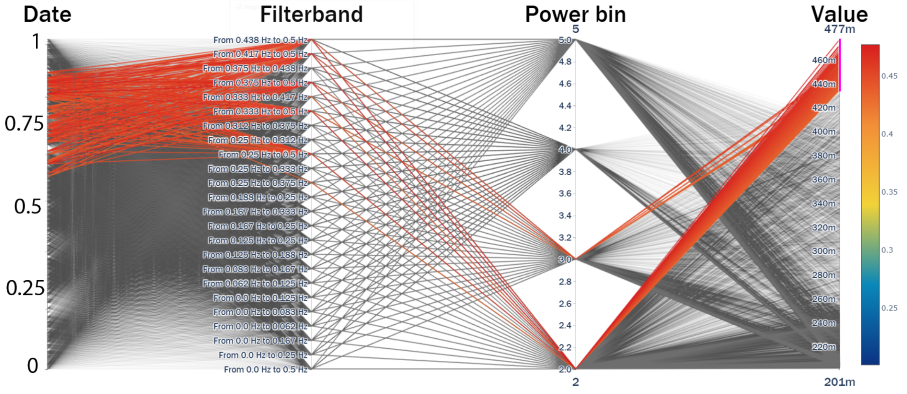

FIGURE 6.25: Selection of the top Hoyer Index values in the parallel coordinates plot. Note that the highest values seem related more to the low power production regimes and high frequency filter bands.

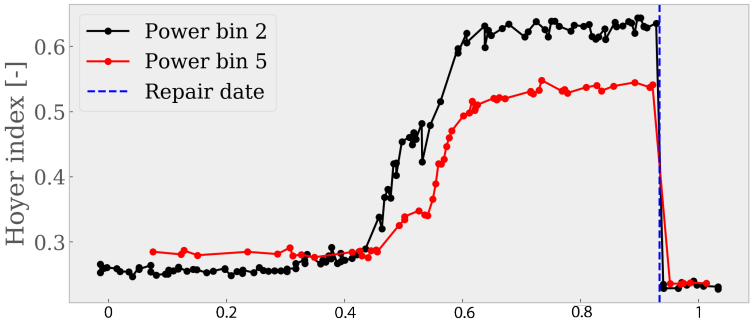

FIGURE 6.26: Hoyer Index evolution in the higher frequency band of [0.375 Hz - 0.5 Hz] and in power bin regimes 2 and 5. Note that in this case a lower active power production seems to influence the feature sensitivity positively.

6.3.3 Envelope analysis

Lastly, to confirm the presence of a bearing fault, the cyclostationary properties of the data is inspected. The cyclic spectral coherence is calculated for each data set and then used to obtain the enhanced envelope spectra. Figure 6.27 shows the evolution in time of the cyclic order band around the BPFO of the fault. The BPFO modulation order can be clearly detected in the two-dimensional map right up till repair, although there does seem to be slight variations in the modulation frequency. Most likely this can be attributed to the slip in the bearing and the changing mechanical properties of the bearing due to the damage progression.

To check which carrier frequency band amplifies the fault the most, the carrier spectra at the BPFO cyclic frequency are extracted from the 2D coherence maps. Figure 6.28 displays the carrier spectra of the BPFO over time. Similar to the observations from the statistical indicator analysis, we can see that the higher frequency bands – from approximately 0.3 Hz to 0.5 Hz – modulate the signal earlier in time and thus prove to be better to detect the damage early on. Another finding that is similar to the statistical analysis, is the strong rise in modulation later on in the damage progression process (starting around 0.6) for low frequency bands (below 0.1 Hz). This is mainly attributable

to the low frequency modulation harmonics of the fault that emerge above the noise floor and indicate a severe progression of the fault. Typically, these harmonics are too low in amplitude and are masked by the noise or other signal components.

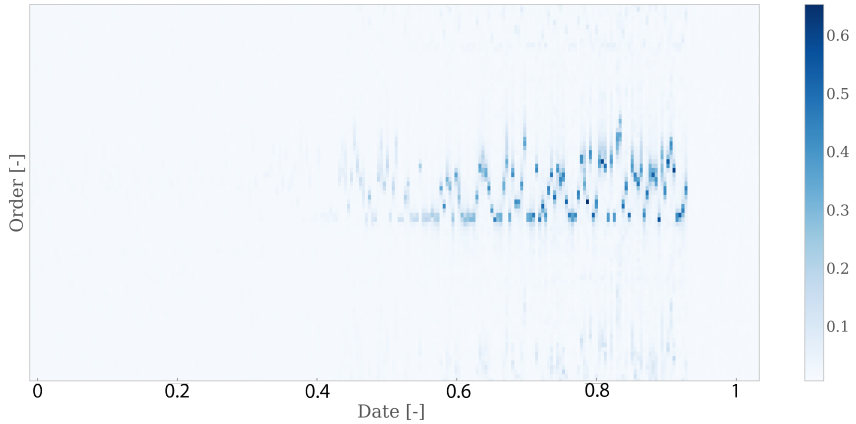

FIGURE 6.27: Two-dimensional color map of the enhanced envelopes over time, zoomed in on the fault order.

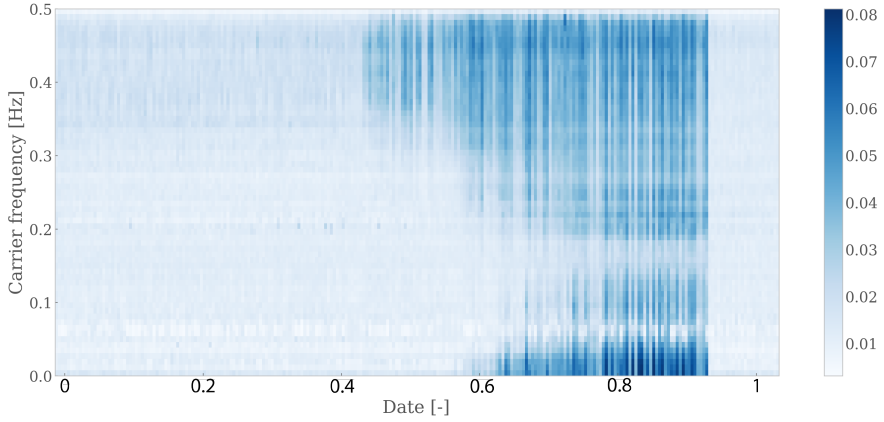

FIGURE 6.28: Two-dimensional map of the carrier spectra over time at the cyclic frequency of the outer race fault.

6.4 Conclusion

This chapter investigates two experimental case studies. Both studies involve the vibrations measured on a wind turbine gearbox housing. The first case study investigates thoroughly the NREL condition monitoring round robin data set using the techniques discussed in this dissertation, with a special emphasis on the automated cepstrum editing procedure. Apart from the planet carrier upwind bearing fault which has the lowest characteristic frequency, all faults are detected using a combination of order tracking, cepstrum editing, band-pass filtering, and envelope analysis. Such a multi-step approach proves vital in the proper and reliable detection of faults in complex machinery.

The second case study involves the investigation of an outer race bearing fault in the high-speed stage of a multi-megawatt wind turbine. Thanks to the availability of an encoder, the analysis also compares the novel MHD method, proposed in Chapter 2, with the MOPA method for the purpose of speed estimation. On average, it is found that the MHD produces significantly more accurate results than the MOPA method for this case. Next, a concise manner to examine the multidimensional statistical filterbank trending results is presented by means of parallel coordinates figures. Also the cyclostationarity of the signal is employed to corroborate the statistical analysis and track the fault frequency specifically. Both the statistical and cyclostationary analysis indicate that the bearing fault first manifests itself in the higher frequency region, about two orders of magnitude higher in frequency than the fault repetition frequency itself. Later in the fault progression, the low frequency modulations become dominant, indicating severe damage.

Chapter 7

Conclusions

7.1 Main contributions

The main objective of this dissertation is the investigation of innovative signal processing methods for vibration-based condition monitoring. Special attention is given to the applicability of the methods in industrial environments and to their automation potential. The dissertation discusses the different processing stages in a multi-step analysis of vibration signals and proposes novel methods in each step that can improve or contribute to the analysis. The chapters are chronologically organized based on this multi-step analysis procedure. The envisioned analysis scheme in this dissertation consists of the following steps:

i Estimation and compensation of the rotation speed fluctuation — Chapter 2

ii Signal separation into deterministic and stochastic components— Chapter 3

iii Signal filtering to increase the potential fault signal-to-noise ratio — Chapter 4

iv Fault detection and automation of the analysis output interpretation — Chapter 5

7.1.1 Instantaneous rotation speed estimation

A vital element in each vibration-based monitoring scheme of non-stationary machinery is the accurate estimation of the instantaneous

rotation speed. In the last two decades, a great deal of research was therefore dedicated to finding an optimal solution for this issue of non-stationarity. Chapter 2 discusses the state-of-the-art in rotation speed estimation techniques and proposes an entirely novel demodulation-based speed estimation method. The proposed method and several of the state-of-the-art methods are examined and compared on three experimental data sets, measured on a wind turbine gearbox, aircraft engine gearbox and a navy ship generator. Due to the very different natures of the three data sets, it is difficult to single out a clear winner from the eight investigated methods. However, the proposed multi-harmonic demodulation method proves to be the most accurate on the wind turbine data set, and runner-up on the aircraft engine data. The ship generator data highlights the need for an approach that is capable of dealing with (very fast) shutdowns and run-ups.

7.1.1.1 Multi-order probabilistic approach

One of the speed estimation approaches that produces promising results is the multi-order probabilistic approach. Using the method is straightforward since only approximate knowledge of the rotation speed range is required. This means that no detailed information about a high signal-to-noise harmonic is necessary. It can therefore cope easier with temporary shutdowns, as illustrated in the experimental analysis of the ship generator data in chapter 2. Even when there is no information available about the kinematics of the machine, it is often possible to extract this information in a fairly accurate manner from a spectrogram and/or an order spectrum. This method therefore proves to be a valuable and versatile asset for industrial analysis cases.

7.1.1.2 Multi-harmonic demodulation

The proposed demodulation-based method is coined the multi-harmonic demodulation method due to its simultaneous usage of multiple harmonics in the demodulation step. Thanks to the use of many harmonics instead of just one, the variance of the speed estimation is significantly reduced. This is a considerable upside compared to single harmonic demodulation since very often the latter would produce

sudden jumps in its estimation due to the instantaneous amplitude moving close to zero.

The reliability of this method is additionally improved thanks to the incorporation of a time-varying maximum-likelihood weighting of the speed-synchronous harmonics. This improvement significantly enhances the ease of use and the accuracy of the method for complex experimental applications. If kinematic ratios of the machine of interest are available, these can all simply be used as input for the method, together with their higher harmonic orders. The maximum-likelihood weighting automatically emphasizes the optimal set of harmonics over time. The only downside of the method is the need for a rough prior speed estimation. While often such a rough speed estimation can be easily obtained, e.g. through edge tracking in the spectrogram, this does make the method in essence a two-step approach.

7.1.2 Automated cepstrum editing

Chapter 3 investigates the separation of deterministic/discrete signal content from stochastic/random signal content. The main focus was on the performance and versatility of cepstrum editing in this separation. An automated cepstrum editing procedure is proposed that detects peaks in a denoised version of the cepstrum and constructs a corresponding notch lifter. A thorough comparison of this approach is made with the so-called cepstrum prewhitening technique. This technique is very easy to implement and to test, since it essentially equalizes the entire amplitude spectrum of a signal. It therefore relies on the phase of a signal to contain all the relevant information necessary to detect a fault. This is often the case when the signal has an impulsive fault signature. The equalization of the amplitude spectrum also implies that strong non-fault related resonances and harmonics get suppressed relatively to the incipient resonance of the fault. For low amplitude faults this can prove essential with regard to early detection. This phenomenon is also observed in the comparison of both techniques. In general, it was found that the appropriate choice of the right cepstrum editing method is not always so clear-cut with down- and upsides for both techniques. Nonetheless, editing the cepstrum provides an uncomplicated, yet potent manner for signal separation compared to some of the other techniques available in literature.

7.1.3 Blind filtering using generalized Rayleigh quotients

Chapter 4 details the development of a novel framework and perspective on blind filtering of vibration signals for fault detection. The need for such a novel approach was actually identified by looking at the available literature about blind deconvolution filters and making the observation that a significant number of techniques is based on the use of kurtosis (or a derivative) as an optimization metric.

The proposed approach employs the generalized Rayleigh quotient to allow for easy integration of envelope-based and envelope spectrum-based optimization metrics into the iterative filter updating procedure. While in the past the main approach for deriving a blind (deconvolution) filter typically involved complicated derivations to obtain the filter coefficient updating formulas, this Rayleigh quotient-based approach enables the use of more complicated condition indicators for blind filtering whilst not increasing the difficulty in finding the correct filter updating expressions. The computation time of the proposed method is also relatively low since only the largest eigenvalues of the generalized eigenvalue problem need to be computed in each iteration.

In total, six indicators are used for the definition of six new filters. One is based on the predictability of the squared envelope, one on the weighted impulsiveness (time negentropy) of the time domain signal, and four on the sparsity of the envelope spectrum. A simulation analysis shows that the defined filters have the potential to detect bearing faults without any prior knowledge about the characteristic fault frequencies. It also demonstrates that the indicators can be used as trending metrics for the health of a bearing, although more in-depth verification for this usage should be performed. The experimental analysis showcases the difference in performance between the different filters, with some filters performing better than others. Further research is required to determine the optimal sparsity indicator and to further optimize the filtering procedure such that it becomes less dependent on the choice of the filter length.

7.1.4 Harmonogram

A general overview of fault detection approaches is given in chapter 5. A minor extension to filterbank approaches, named the harmonogram, is proposed. It combines the idea of the kurtogram with an envelope spectrum analysis by defining a fault frequency-based indicator on each envelope spectrum of each filtered signal. Extensive testing of the method is planned for future work, but a simulation analysis demonstrates the potential efficacy of the method to easily find a demodulation band using cyclostationarity. Next to the harmonogram, chapter 5 also discusses statistical time domain indicators, frequency filtering, and the cyclic spectral correlation.

7.1.5 Dealing with large condition monitoring data sets

The last section of chapter 5 details how to deal with large numbers of condition indicators coming from complex machinery in an automated manner. The first step in the suggested approach is to model the expected behavior of the condition indicators. This entails making the condition indicators independent of the operating regime, i.c. this is done using k-means clustering on operational data parameters (e.g. SCADA data). Then a healthy condition model per condition indicator and operating regime is trained using linear Bayesian ridge regression. The final step is then to detect any anomalies in the condition indicators that have been made independent of the operating regime. Whenever an indicator veers a certain number of standard deviations from the expected behavior, an alarm is raised. After fusing the normalized indicators together, the anomalies can be given easy to interpret colors such that an an end-user can easily inspect the figures for potential alarms. This automation approach is illustrated in chapter 5 on a wind turbine bearing fault case and proves capable of significantly simplifying the analysis process.

7.1.6 Experimental case studies

To conclude the dissertation, chapter 6 illustrates the presented methods in in-depth analyses of two experimental wind turbine gearbox data sets. The first case study details a full processing flow involving

order tracking, automated cepstrum editing, band-pass filtering with the kurtogram, and envelope analysis. The second case study demonstrates the proposed multi-harmonic demodulation method of chapter 2 and analyzes its reliability for a long term measurement campaign. Also the proposed statistical filterbank trending approach of chapter 5 is showcased by visualizing the multi-dimensional data in an compressed manner. Lastly, the cyclic spectral coherence is computed for this data set in order to highlight that the carrier spectra at known cyclic frequencies can also be employed to find optimal demodulation frequency bands for specific faults.

7.2 Recommendations for future research

This dissertation showed the need for including many different processing steps in a vibration-based condition monitoring scheme. This section presents several suggestions to further improve and extend the present work.

- One of the main findings of the analysis of the ship generator data in chapter 2 indicated that there is still a need for a more adequate method for the speed estimation of very fast run-ups and run-downs. Additionally, methods based on demodulation or on edge tracking in the spectrogram should be enabled to deal with temporary shutdowns in such a way that they are capable of resuming in an accurate fashion.

- The multi-order probabilistic approach is already a very accurate and versatile tool to estimate the speed. However, one of the main input parameters of the method that often needs manual tweaking, is the window length used for the spectrogram computation. Future research will investigate how this window size choice can be automated such that essentially the only required input will be the approximate speed range.

- In general, the user experience and findings with the multi-harmonic demodulation method of chapter 2 have been very positive thanks to the high accuracy and robustness of the method to interfering sources. It therefore begs the question whether the output of the

method could be used as input for condition monitoring based on its estimated instantaneous angular speed. Recent research efforts have been looking at employing high-resolution rotary encoders for condition monitoring instead of vibration sensors. It would thus be interesting to investigate if a sufficiently accurate speed can be estimated with the MHD method to use it as well for condition monitoring.

- Chapter 3 investigated cepstrum editing for discrete-random signal separation. Despite producing good results, it is clear that further improvement is still possible since neither the cepstrum prewhitening nor the automated cepstrum editing procedure always produce the optimal results. Investigating how the editing step can be improved will probably be essential in this research.

- Chapter 4 proposed a new approach to develop blind vibration filters using complex optimization metrics. While six different filters have already been derived, future research should focus on finding new optimization metrics and better ways to make the filtering procedure more independent from the choice of filter length.

- The presented automation approach of chapter 5 is one possible way to implement an automated condition monitoring scheme. However, many more techniques are possible to use for this purpose. Therefore, the presented approach should be further expanded and validated on more experimental data sets.

- Finally, the techniques introduced in this dissertation mainly focused in their experimental applications on the goal of bearing fault detection. The performance of the presented techniques and of the multi-step procedure should be examined for gear fault detection.

List of publications

International peer-reviewed journal articles

I Peeters, C., Leclère, Q., Antoni, J., Lindahl, P., Donnal, J., Leeb, S., & Helsen, J. (2019). **Review and comparison of tacholess instantaneous speed estimation methods on experimental vibration data**. *Mechanical Systems and Signal Processing*, 129(1), 407-436.

II Peeters, C., Guillaume, P., & Helsen, J. (2018). **Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy**. *Renewable Energy*, 116(Part B), 74-87.

III Peeters, C., Guillaume, P., & Helsen, J. (2017). **A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection**. *Mechanical Systems and Signal Processing*, 91, 354-381. [91].

IV Gioia, N., Peeters, C., Guillaume, P., Helsen, J. (2019). **Field data-driven Noise, Vibration and Harshness behavior identification of wind turbine drivetrain**. *Energies: Design, Fabrication and Performance of Wind Turbines*, 518040

V Peeters, C., Antoni, J., Helsen, J. (2019). **Blind filters based on envelope spectrum sparsity indicators for bearing and gear vibration-based condition monitoring**. *Mechanical Systems and Signal Processing*, MSSP19-940, *Under Review*

Peer-reviewed Articles in Journal of Physics

I Helsen, J., Peeters, C., Gioia, N., & jordaens, P. J. (2017). **Integrated condition monitoring of a fleet of offshore wind turbines with focus on acceleration streaming processing**. *Journal of Physics: Conference Series*, 842(1), 1-8..

II Peeters, C., Leclère, Q., Antoni, J., Guillaume, P., & Helsen, J. (2017). **Vibration-based angular speed estimation for multi-stage wind turbine gearboxes**. *Journal of Physics: Conference Series*, 842(1), 1-9..

III Peeters, C., Gioia, N., & Helsen, J. (2018). **Stochastic simulation assessment of an automated vibration-based condition monitoring framework for wind turbine gearbox faults**. In *The Science of Making Torque from Wind* (3 ed., Vol. 1037). (Journal of Physics: Conference Series). IOP Publishing.

IV Gioia, N., Daems, P-J., Peeters, C., El-kafafy, M., Guillaume, P., & Helsen, J. (2018). **Long term operational modal analysis for rotating machines**. In *The Science of Making Torque from Wind*. 3 ed. IOP Publishing (5 ed., Vol. 1037, pp. 103-112). (Journal of Physics: Conference Series).

Peer-reviewed international conference articles

I Peeters, C., Verstraeten, T., Nowe, A., & Helsen, J. (2019). **Wind turbine planetary gear fault identification using statistical condition indicators and machine learning**. In *OMAE Proceedings 2019*, 38th International Conference on Ocean, Offshore & Arctic Engineering, Glasgow, UK.

II Peeters, C., Leclère, Q., Antoni, J., Guillaume, P., & Helsen, J. (2017). **Instantaneous shaft speed estimation for wind turbine using a multi-order probabilistic approach in a continuous streaming data processing framework**. In *International conference on Structural Engineering Dynamics (ICEDyn)*, Ericeira, Portugal.

III Peeters, C., Guillaume, P., & Helsen, J. (2016). **Vibration-based bearing fault detection analysis algorithms and diagnostics results on experimental wind turbine gearbox data**. In *Proceedings of the Prognostics Health Monitoring Conference* (pp. 660-671). PHME 2016, Bilbao, Spain.

IV Peeters, C., Helsen, J., & Guillaume, P. (2016). **Vibration data preprocessing techniques for rolling element bearing fault detection**. In *Proceedings of the international conference on sound and vibration* (pp. 307-314), Athens, Greece.

V Helsen, J., Peeters, C., Doro, P., Ververs, E., & jordaens, P. J. (2017). **Wind farm operation and maintenance optimization using big data**. In *Proceedings - 3rd IEEE International Conference on Big Data Computing Service and Applications*, BigDataService 2017 (Vol. 10, pp. 179-184), IEEE, San Francisco, USA.

VI Gioia, N., Daems, P. J., Peeters, C., El-Kafafy, M., Guillaume, P., & Helsen, J. (2019). **Influence of the harmonics on the modal behavior of wind turbine drivetrains**. In *Rotating Machinery, Vibro-Acoustics & Laser Vibrometry*, Volume 7 (pp. 231-238). Springer, IMAC, Orlando, USA.

VII Soares, M. N., Gyselinck, J., Mollet, Y., Peeters, C., Gioia, N., & Helsen, J. (2018). **Vibration-Based Rotor-Side-Converter Open-Switch-Fault Detection in DFIGs for Wind Turbines**. In: *IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 1-6). IEEE, Chicago, USA.

VIII Gioia, N., Peeters, C., Daems, P. J., Guillaume, P., & Helsen, J. (2018). **Rotating machines system identification by means of big data analysis**. In: *European Workshop on Structural Health Monitoring (EWSHM)*, Manchester, UK.

Other international conference articles

I Helsen, J., Peeters, C., Verstraeten, T., Verbeke, J., Gioia, N., & Nowé, A. (2018) **Fleet-wide condition monitoring combining vibration signal processing and machine learning rolled out in a cloud-computing environment**. In *International Conference on Noise and Vibration Engineering*. ISMA, Leuven, Belgium.

II Peeters, C., Antoni, J., Gioia, N., Guillaume, P., & Helsen, J. (2018) **A novel multi-harmonic demodulation technique for instantaneous speed estimation**. In *International Conference on Noise and Vibration Engineering*. ISMA, Leuven, Belgium.

III Peeters, C., Guillaume, P., & Helsen, J. (2016). **Signal pre-processing using cepstral editing for vibration-based bearing fault detection**. In *International Conference on Noise and Vibration Engineering* (pp. 2489-2502). ISMA 2016, Leuven, Belgium.

IV Verstraeten, T., Marulanda, F. G., Peeters, C., Daems, P. J., Nowé, A., & Helsen, J. (2019). **Edge computing for advanced vibration signal processing**. In: *International Conference of Surveillance, Vibration and Noise*, Lyon, France.

V Peeters, C., Leclère, Q., Antoni, J., & Helsen, J. (2017). **Instantaneous angular speed estimation for multi-stage wind turbine gearboxes**. In *International conference of Surveillance 9*, Fez, Morocco.

VI Peeters, C., Antoni, J., & Helsen, J. (2019). **Blind vibration filtering using envelope spectrum indicators for bearing and gear fault detection without knowledge of machine kinematics**. In *HAL Archives* (pp. 1-8), International Conference of Surveillance, Vibration and Noise, Lyon, France.

International conferences with abstract-only proceedings

I Peeters, C., Guillaume, P., Helsen, J. (2017) **Advanced vibration signal processing on experimental wind turbine gearbox data**. In: *Wind Energy Science Conference*, Copenhagen, Denmark.

II Peeters, C., Daems, P. J., Verstraeten, T., Helsen, J. (2019) **Combining signal processing and AI to optimize condition monitoring on wind turbines**. In: *Wind Energy Science Conference*, Cork, Ireland.

Awards

I Best PhD student paper award, July 2017, International Conference on Damage Assessment of Structures, Fukuoka, Japan.

Curriculum Vitae

Education

Vrije Universiteit Brussel, Belgium
Ph.D. student,
Doctoral school of Natural Sciences & (Bioscience) Engineering.
September 2015 (graduation date: November 21st 2019)

- Title of the PhD research: *Advanced signal processing for the identification and diagnosis of the condition and remaining useful life of rotating machinery*
- Description: The research focuses on developing novel processing techniques that can be used in a robust and scalable way as to accommodate the rising interest from industry in this type of tools.
- Joint PhD degree with University of INSA Lyon.
- Advisors: Jérôme Antoni and Jan Helsen

KU Leuven, Belgium
Master of Science in Mechanical Engineering
September 2013 - July 2015
Specialized in automotive and vibro-acoustic engineering. Master thesis: *Practical realisations of metamaterial concepts for noise and vibration reduction*. Advisor: Wim Desmet.

ATHENS Exchange program, Universidad Politecnica de Madrid, Spain
November 2014
Training course on simulation tools for sound reinforcement

KU Leuven, Belgium

Bachelor of Science in Engineering

September 2010 - July 2013

Bachelor thesis: *Design of a satellite clamping mechanism for VEGA launchers*, in association with QINETIQ SPACE.

Heilig Hart scholen, Heist-op-den-Berg

September 2004 - July 2010

Ancient Greek - Mathematics (8h./week).

Acknowledgments

This PhD dissertation would not have been possible without the necessary funding. Therefore, a few parties need to be thanked for their support which was vital to this research.

First, the author would like to thank the agency for Innovation by Science and Technology in Belgium for supporting the SBO HYMOP project and the agency for Strategic Initiative Materials in Flanders for supporting the MaSiWEC project.

Second, the author also expresses his gratitude to *Fonds Wetenschappelijk Onderzoek - Vlaanderen* or *FWO* for their support with providing funding for a long stay abroad of the author at the university of INSA Lyon.

Third, the author also expresses his gratitude towards the researchers, agencies, and companies that made their experimental data sets available for research. This includes data such as the wind turbine condition monitoring round robin data set of the National Renewable Energy Laboratory in Colorado, USA. Also, the wind farm operators and OEMs are thanked for their contribution in this research.

Bibliography

[1] MARKETS and MARKETS. *Predictive Maintenance Market by Component (Solutions and Services), Deployment Mode, Organization Size, Vertical (Government and Defense, Manufacturing, Energy and Utilities, Transportation and Logistics), and Region - Global Forecast to 2024*. 2019. URL: https://www.marketsandmarkets.com/Market-Reports/operational-predictive-maintenance-market-8656856.html (visited on 07/12/2019).

[2] PWC. *Predictive Maintenance 4.0, Beyond the hype: PdM 4.0 delivers results*. 2019. URL: https://www.pwc.be/en/documents/20180926-pdm40-beyond-the-hype-report.pdf (visited on 07/12/2019).

[3] James M WAKIRU et al. "A review on lubricant condition monitoring information analysis for maintenance decision support". In: *Mechanical systems and signal processing* 118 (2019), pp. 108–132.

[4] David VALIŠ and Libor ŽÁK. "Oil additives used as indicator and input for preventive maintenance optimisation". In: *International Conference on Military Technologies (ICMT) 2015*. IEEE. 2015, pp. 1–6.

[5] Vivek S DAVE et al. "Lubricant-sensitivity assessment of SPRESS B820 by near-infrared spectroscopy: a comparison of multivariate methods". In: *Journal of pharmaceutical sciences* 106.2 (2017), pp. 537–545.

[6] Jim BARKER, Stephen COOK, and Paul RICHARDS. "Sodium contamination of diesel fuel, its interaction with fuel additives and the resultant effects on filter plugging and injector fouling". In: *SAE International Journal of Fuels and Lubricants* 6.3 (2013), pp. 826–838.

[7] Sam GEORGE et al. “Effect of diesel soot on lubricant oil viscosity”. In: *Tribology international* 40.5 (2007), pp. 809–818.

[8] Anand PRABHAKARAN and CR JAGGA. “Condition monitoring of steam turbine-generator through contamination analysis of used lubricating oil”. In: *Tribology International* 32.3 (1999), pp. 145–152.

[9] Manoj KUMAR, Parboti SHANKAR MUKHERJEE, and Nirendra MOHAN MISRA. “Advancement and current status of wear debris analysis for machine condition monitoring: a review”. In: *Industrial Lubrication and Tribology* 65.1 (2013), pp. 3–11.

[10] Wei CAO, Wen Juan WANG, and Rui WANG. “Wear Trend Prediction of Gearbox Based on Oil Monitoring Technology”. In: *Advanced Materials Research*. Vol. 411. Trans Tech Publ. 2012, pp. 576–579.

[11] Bin FAN et al. “Modeling and experimental investigations on the relationship between wear debris concentration and wear rate in lubrication systems”. In: *Tribology International* 109 (2017), pp. 114–123.

[12] Y HU et al. “Development of an interactive friction model for the prediction of lubricant breakdown behaviour during sliding wear”. In: *Tribology International* 110 (2017), pp. 370–377.

[13] Davor LJUBAS, Hrvoje KRPAN, and Ivica MATANOVIĆ. “Influence of engine oils dilution by fuels on their viscosity, flash point and fire point”. In: *Nafta: exploration, production, processing, petrochemistry* 61.2 (2010), pp. 73–79.

[14] Masoud AFRAND et al. “Prediction of dynamic viscosity of a hybrid nano-lubricant by an optimal artificial neural network”. In: *International Communications in Heat and Mass Transfer* 76 (2016), pp. 209–214.

[15] Michael Clifford KOCSIS, Thomas BRIGGS, and Garrett ANDERSON. “The impact of lubricant volatility, viscosity and detergent chemistry on low speed pre-ignition behavior”. In: *SAE International Journal of Engines* 10.3 (2017), pp. 1019–1035.

[16] Subramaniam BAGAVATHIAPPAN et al. “Infrared thermography for condition monitoring–A review”. In: *Infrared Physics & Technology* 60 (2013), pp. 35–55.

[17] Timothy VERSTRAETEN et al. “Fleetwide data-enabled reliability improvement of wind turbines”. In: *arXiv preprint arXiv:1903.11518* (2019).

[18] E GRINZATO et al. “Monitoring of the Scrovegni Chapel by IR thermography: Giotto at infrared”. In: *Infrared Physics & Technology* 43.3-5 (2002), pp. 165–169.

[19] Carosena MEOLA. “Infrared thermography of masonry structures”. In: *Infrared physics & technology* 49.3 (2007), pp. 228–233.

[20] EJ JOHNSON et al. “Evaluation of infrared thermography as a diagnostic tool in CVD applications”. In: *Journal of crystal growth* 187.3-4 (1998), pp. 463–473.

[21] Mohd Shawal JADIN and Soib TAIB. “Recent progress in diagnosing the reliability of electrical equipment by using infrared thermography”. In: *Infrared physics & technology* 55.4 (2012), pp. 236–245.

[22] Yang CAO, Xiaoming GU, and Qi JIN. “Infrared technology in the fault diagnosis of substation equipment”. In: *2008 China International Conference on Electricity Distribution*. IEEE. 2008, pp. 1–6.

[23] František LIZÁK and Michal KOLCUN. “Improving reliability and decreasing losses of electrical system with infrared thermography”. In: *Acta Electrotechnica et Informatica* 8.1 (2008), pp. 60–63.

[24] Harri T KIISKINEN et al. “Infrared thermography examination of paper structure”. In: *Tappi journal* 80.4 (1997), pp. 159–162.

[25] A WYCKHUYSE and Xavier MALDAGUE. “A study of wood inspection by infrared thermography, part I: Wood pole inspection by infrared thermography”. In: *Research in nondestructive evaluation* 13.1 (2001), pp. 1–12.

[26] Carosena MEOLA et al. “The use of infrared thermography for nondestructive evaluation of joints”. In: *Infrared physics & technology* 46.1-2 (2004), pp. 93–99.

[27] R KAFIEH, T LOTFI, and Rassoul AMIRFATTAHI. “Automatic detection of defects on polyethylene pipe welding using thermal infrared imaging”. In: *Infrared Physics & Technology* 54.4 (2011), pp. 317–325.

[28] U SREEDHAR et al. “Automatic defect identification using thermal image analysis for online weld quality monitoring”. In: *Journal of Materials Processing Technology* 212.7 (2012), pp. 1557–1566.

[29] Minh Phong LUONG. “Fatigue limit evaluation of metals using an infrared thermographic technique”. In: *Mechanics of materials* 28.1-4 (1998), pp. 155–163.

[30] ML PASTOR et al. “Applying infrared thermography to study the heating of 2024-T3 aluminium specimens under fatigue loading”. In: *Infrared Physics & Technology* 51.6 (2008), pp. 505–515.

[31] Hsin WANG et al. “Infrared temperature mapping of ULTIMET alloy during high-cycle fatigue tests”. In: *Metallurgical and Materials Transactions A* 31.4 (2000), pp. 1307–1310.

[32] S BAGAVATHIAPPAN et al. “Condition monitoring of exhaust system blowers using infrared thermography”. In: *Insight-Non-Destructive Testing and Condition Monitoring* 50.9 (2008), pp. 512–515.

[33] Dong-Yeon KIM et al. “Fault diagnosis of ball bearings within rotational machines using the infrared thermography method”. In: *Journal of the Korean Society for Nondestructive Testing* 30.6 (2010), pp. 558–563.

[34] Jin Ju SEO et al. “Infrared thermographic diagnosis mechnism for fault detection of ball bearing under dynamic loading conditions”. In: *Advanced materials research*. Vol. 295. Trans Tech Publ. 2011, pp. 1544–1547.

[35] LD FAVRO et al. “Thermal wave imaging of aircraft structures”. In: *Review of Progress in Quantitative NonDestructive Evaluation*. Springer, 1995, pp. 461–466.

[36] B VENKATARAMAN et al. “Correlation of infrared thermographic patterns and acoustic emission signals with tensile deformation and fracture processes”. In: *AIP Conference Proceedings*. Vol. 557. 1. AIP. 2001, pp. 1443–1450.

[37] C BADULESCU et al. “Applying the grid method and infrared thermography to investigate plastic deformation in aluminium multicrystal”. In: *Mechanics of Materials* 43.1 (2011), pp. 36–53.

[38] Jalaj KUMAR, Sony BABY, and Vikas KUMAR. “Thermographic studies on IMI-834 titanium alloy during tensile loading”. In: *Materials Science and Engineering: A* 496.1-2 (2008), pp. 303–307.

[39] Hakim AIT-AMOKHTAR, Claude FRESSENGEAS, and S BOUDRAHEM. “The dynamics of Portevin–Le Chatelier bands in an Al–Mg alloy from infrared thermography”. In: *Materials Science and Engineering* 488.1-2 (2008), pp. 540–546.

[40] AA GOWEN et al. “Applications of thermal imaging in food quality and safety assessment”. In: *Trends in food science & technology* 21.4 (2010), pp. 190–200.

[41] Miquel VELLVEHI et al. “Irradiance-based emissivity correction in infrared thermography for electronic applications”. In: *Review of scientific instruments* 82.11 (2011), p. 114901.

[42] S HUTH et al. “Lock-in IR-thermography-A novel tool for material and device characterization”. In: *Diffusion And Defect Data Part B Solid State Phenomena*. Citeseer. 2002, pp. 741–746.

[43] Ronald D LUCIER and Joseph C BOGNET. “Thermography In Commercial Nuclear Power Stations Applications, Limitations, And Realities”. In: *Thermosense XI: Intl Conf on Thermal Infrared Sensing for Diagnostics and Control*. Vol. 1094. International Society for Optics and Photonics. 1989, pp. 56–62.

[44] Clemente IBARRA-CASTANEDO et al. “Recent progresses in the inspection of aerospace components by infrared thermography”. In: *17th World Conference on Nondestructive Testing*. 2008.

[45] M FIDALI. “An idea of continuous thermographic monitoring of machinery”. In: *image* 7.11 (2008), p. 15.

[46] BB LAHIRI et al. “Quantification of defects in composites and rubber materials using active thermography”. In: *Infrared Physics & Technology* 55.2-3 (2012), pp. 191–199.

[47] ATP SO et al. “Fuzzy logic based automatic diagnosis of power apparatus by infrared imaging”. In: *[1993] Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems*. IEEE. 1992, pp. 187–192.

[48] Baoshu LI et al. “HV power equipment diagnosis based on infrared imaging analyzing”. In: *2006 International Conference on Power System Technology*. IEEE. 2006, pp. 1–4.

[49] Jianbo SHI and Jitendra MALIK. “Normalized cuts and image segmentation”. In: *Departmental Papers (CIS)* (2000), p. 107.

[50] M LYDIA et al. “A comprehensive review on wind turbine power curve modeling techniques”. In: *Renewable and Sustainable Energy Reviews* 30 (2014), pp. 452–460.

[51] Andrew KUSIAK, Haiyang ZHENG, and Zhe SONG. “Models for monitoring wind farm power”. In: *Renewable Energy* 34.3 (2009), pp. 583–590.

[52] Kirsten TRACHT et al. “Failure probability prediction based on condition monitoring data of wind energy systems for spare parts supply”. In: *CIRP Annals* 62.1 (2013), pp. 127–130.

[53] M WÄCHTER et al. “Efficient load and power monitoring by stochastic methods”. In: *EWEA 2015 Annual Event* (2015).

[54] Jannis TAUTZ-WEINERT and Simon J WATSON. “Using SCADA data for wind turbine condition monitoring–a review”. In: *IET Renewable Power Generation* 11.4 (2016), pp. 382–394.

[55] Wenxian YANG, Richard COURT, and Jiesheng JIANG. “Wind turbine condition monitoring by the approach of SCADA data analysis”. In: *Renewable Energy* 53 (2013), pp. 365–376.

[56] Wenxian YANG et al. “Wind turbine condition monitoring: technical and commercial challenges”. In: *Wind Energy* 17.5 (2014), pp. 673–693.

[57] Jamie L GODWIN and Peter MATTHEWS. “Classification and detection of wind turbine pitch faults through SCADA data analysis”. In: *IJPHM Special Issue on Wind Turbine PHM* (2013), p. 90.

[58] Mari Cruz GARCIA, Miguel A SANZ-BOBI, and Javier DEL PICO. "SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox". In: *Computers in Industry* 57.6 (2006), pp. 552–568.

[59] ASAE ZAHER et al. "Online wind turbine fault detection through automated SCADA data analysis". In: *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 12.6 (2009), pp. 574–593.

[60] S CATMULL. "Self-organising map based condition monitoring of wind turbines". In: *EWEA Annual Conf*. Vol. 2011. 2011.

[61] SJ WATSON, I KENNEDY, and CS GRAY. "The use of physics of failure modelling in wind turbine condition monitoring". In: *EWEA Annual Conf*. Vol. 2011. 2011, pp. 309–312.

[62] Meik SCHLECHTINGEN, Ilmar Ferreira SANTOS, and Sofiane ACHICHE. "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description". In: *Applied Soft Computing* 13.1 (2013), pp. 259–270.

[63] Michael WILKINSON et al. "Comparison of methods for wind turbine condition monitoring with SCADA data". In: *IET Renewable Power Generation* 8.4 (2014), pp. 390–397.

[64] Peng SUN et al. "A generalized model for wind turbine anomaly identification based on SCADA data". In: *Applied Energy* 168 (2016), pp. 550–567.

[65] Andrew KUSIAK and Zijun ZHANG. "Analysis of wind turbine vibrations based on SCADA data". In: *Journal of Solar Energy Engineering* 132.3 (2010), p. 031008.

[66] Kyusung KIM et al. *Use of SCADA data for failure detection in wind turbines*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2011.

[67] William G GARLICK, Roger DIXON, and Simon J WATSON. "A model-based approach to wind turbine condition monitoring using SCADA data". In: ICSE. 2009.

[68] Meik SCHLECHTINGEN and Ilmar Ferreira SANTOS. “Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection”. In: *Mechanical systems and signal processing* 25.5 (2011), pp. 1849–1875.

[69] Christopher S GRAY and Simon J WATSON. “Physics of failure approach to wind turbine condition based maintenance”. In: *Wind Energy* 13.5 (2010), pp. 395–405.

[70] Dick BRETELER et al. “Physics based methodology for wind turbine failure detection, diagnostics & prognostics”. In: *EWEA 2015 Annual Event* (2015).

[71] Yingning QIU et al. “An electro-thermal analysis of a variable-speed doubly-fed induction generator in a wind turbine”. In: *Energies* 8.5 (2015), pp. 3386–3402.

[72] Bindi CHEN et al. “Bayesian network for wind turbine fault diagnosis.” In: (2012).

[73] Yingning QIU et al. “Wind turbine SCADA alarm analysis for improving reliability”. In: *Wind Energy* 15.8 (2012), pp. 951–966.

[74] Andrew KUSIAK and Wenyan LI. “The prediction and diagnosis of wind turbine faults”. In: *Renewable energy* 36.1 (2011), pp. 16–23.

[75] Philip CROSS and Xiandong MA. “Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines”. In: *International Journal of Automation and Computing* 12.1 (2015), pp. 25–34.

[76] Hui LI et al. “An improved fuzzy synthetic condition assessment of a wind turbine generator system”. In: *International Journal of Electrical Power & Energy Systems* 45.1 (2013), pp. 468–476.

[77] Jian LI et al. “Normal behavior models for the condition assessment of wind turbine generator systems”. In: *Electric Power Components and Systems* 42.11 (2014), pp. 1201–1212.

[78] Hubert W SCHREIER and Michael A SUTTON. “Systematic errors in digital image correlation due to undermatched subset shape functions”. In: *Experimental Mechanics* 42.3 (2002), pp. 303–310.

[79] XW YE, CZ DONG, and T LIU. “Image-based structural dynamic displacement measurement using different multi-object tracking algorithms”. In: *Smart Structures and Systems* 17.6 (2016), pp. 935–956.

[80] Shwu-Huey YEN, Chun-Hui WANG, and Jui-Chen CHIEN. “Accurate and robust ROI localization in a camshift tracking application”. In: *Multimedia Tools and Applications* 74.23 (2015), pp. 10291–10312.

[81] Zuobin WANG et al. “Phase-shifted image matching algorithm for displacement measurement”. In: *Optical Engineering* 35 (1996).

[82] Jie GUO et al. “Dynamic displacement measurement of large-scale structures based on the Lucas–Kanade template tracking algorithm”. In: *Mechanical Systems and Signal Processing* 66 (2016), pp. 425–436.

[83] Yoshio FUKUDA et al. “Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm”. In: *IEEE Sensors Journal* 13.12 (2013), pp. 4725–4732.

[84] Eann A PATTERSON et al. “Calibration and evaluation of optical systems for full-field strain measurement”. In: *Optics and Lasers in Engineering* 45.5 (2007), pp. 550–564.

[85] RNF CARMO et al. “Assessing steel strains on reinforced concrete members from surface cracking patterns”. In: *Construction and Building Materials* 98 (2015), pp. 265–275.

[86] Chien-Chou CHEN et al. “Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders”. In: *Measurement* 75 (2015), pp. 134–146.

[87] Yoshio FUKUDA, Maria Q FENG, and Masanobu SHINOZUKA. “Cost-effective vision-based system for monitoring dynamic response of civil engineering structures”. In: *Structural Control and Health Monitoring* 17.8 (2010), pp. 918–936.

[88] UP POUDEL, G FU, and J YE. “Structural damage detection using digital video imaging technique and wavelet transformation”. In: *Journal of Sound and Vibration* 286.4-5 (2005), pp. 869–895.

[89] DLBR JURJO et al. “Experimental methodology for the dynamic analysis of slender structures based on digital image processing techniques”. In: *Mechanical Systems and Signal Processing* 24.5 (2010), pp. 1369–1382.

[90] E CAETANO, S SILVA, and J BATEIRA. “A vision system for vibration monitoring of civil engineering structures”. In: *Experimental Techniques* 35.4 (2011), pp. 74–82.

[91] Piotr KOHUT and Piotr KUROWSKI. “Application of modal analysis supported by 3D vision-based measurements”. In: *Journal of Theoretical and Applied Mechanics* 47.4 (2009), pp. 855–870.

[92] Chul Min YEUM and Shirley J DYKE. “Vision-based automated crack detection for bridge inspection”. In: *Computer-Aided Civil and Infrastructure Engineering* 30.10 (2015), pp. 759–770.

[93] Jónatas VALENÇA et al. “Automatic concrete health monitoring: assessment and monitoring of concrete surfaces”. In: *Structure and Infrastructure Engineering* 10.12 (2014), pp. 1547–1554.

[94] Mahmoud R HALFAWY and Jantira HENGMEECHAI. “Integrated vision-based system for automated defect detection in sewer closed circuit television inspection videos”. In: *Journal of Computing in Civil Engineering* 29.1 (2014), p. 04014024.

[95] Takahide SAKAGAMI. “Remote nondestructive evaluation technique using infrared thermography for fatigue cracks in steel bridges”. In: *Fatigue & Fracture of Engineering Materials & Structures* 38.7 (2015), pp. 755–779.

[96] Stephanie GERMAN et al. “Machine vision-enhanced postearthquake inspection”. In: *Journal of Computing in Civil Engineering* 27.6 (2013), pp. 622–634.

[97] Xiao-Wei YE, CZ DONG, and Tan LIU. “A review of machine vision-based structural health monitoring: methodologies and applications”. In: *Journal of Sensors* 2016 (2016).

[98] JC SU, CK HUANG, and YS TARNG. “An automated flank wear measurement of microdrills using machine vision”. In: *Journal of Materials Processing Technology* 180.1-3 (2006), pp. 328–335.

[99] BM KUMAR and MM RATNAM. “Machine vision method for non-contact measurement of surface roughness of a rotating workpiece”. In: *Sensor review* 35.1 (2015), pp. 10–19.

[100] Jianfeng ZHONG et al. “Vision-based system for simultaneous monitoring of shaft rotational speed and axial vibration using non-projection composite fringe pattern”. In: *Mechanical Systems and Signal Processing* 120 (2019), pp. 765–776.

[101] E Dehghan NIRI and S SALAMONE. “A probabilistic framework for acoustic emission source localization in plate-like structures”. In: *Smart Materials and Structures* 21.3 (2012), p. 035009.

[102] Francesco CIAMPA and Michele MEO. “Acoustic emission source localization and velocity determination of the fundamental mode A0 using wavelet analysis and a Newton-based optimization technique”. In: *Smart Materials and Structures* 19.4 (2010), p. 045027.

[103] Tribikram KUNDU, Hayato NAKATANI, and Nobuo TAKEDA. “Acoustic source localization in anisotropic plates”. In: *Ultrasonics* 52.6 (2012), pp. 740–746.

[104] Christian U GROSSE and Masayasu OHTSU. *Acoustic emission testing*. Springer Science & Business Media, 2008.

[105] Martine WEVERS. “Listening to the sound of materials: acoustic emission for the analysis of material behaviour”. In: *Ndt & E International* 30.2 (1997), pp. 99–106.

[106] TMand ROBERTS and M TALEBZADEH. “Acoustic emission monitoring of fatigue crack propagation”. In: *Journal of Constructional Steel Research* 59.6 (2003), pp. 695–712.

[107] A CHOUDHURY and N TANDON. “Application of acoustic emission technique for the detection of defects in rolling element bearings”. In: *Tribology international* 33.1 (2000), pp. 39–45.

[108] Mohamed ELFORJANI and David MBA. “Detecting natural crack initiation and growth in slow speed shafts with the acoustic emission technology”. In: *Engineering failure analysis* 16.7 (2009), pp. 2121–2129.

[109] N TANDON, GS YADAVA, and KM RAMAKRISHNA. “A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings”. In: *Mechanical systems and signal processing* 21.1 (2007), pp. 244–256.

[110] Naresh TANDON and Achintya CHOUDHURY. “A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings”. In: *Tribology international* 32.8 (1999), pp. 469–480.

[111] Saad AL-DOSSARY, RI Raja HAMZAH, and David MBA. “Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing”. In: *Applied acoustics* 70.1 (2009), pp. 58–81.

[112] LM ROGERS. “The application of vibration signature analysis and acoustic emission source location to on-line condition monitoring of anti-friction bearings”. In: *Tribology international* 12.2 (1979), pp. 51–58.

[113] M ELFORJANI and D MBA. “Accelerated natural fault diagnosis in slow speed bearings with acoustic emission”. In: *Engineering Fracture Mechanics* 77.1 (2010), pp. 112–127.

[114] Abdullah M AL-GHAMD and David MBA. “A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size”. In: *Mechanical systems and signal processing* 20.7 (2006), pp. 1537–1571.

[115] Yongyong HE, Xinming ZHANG, and Michael I FRISWELL. “Defect diagnosis for rolling element bearings using acoustic emission”. In: *Journal of vibration and acoustics* 131.6 (2009), p. 061012.

[116] Babak EFTEKHARNEJAD et al. “The application of spectral kurtosis on acoustic emission and vibrations from a defective bearing”. In: *Mechanical Systems and Signal Processing* 25.1 (2011), pp. 266–284.

[117] N JAMALUDIN, David MBA, and RH BANNISTER. “Condition monitoring of slow-speed rolling element bearings using stress waves”. In: *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* 215.4 (2001), pp. 245–271.

[118] Zhanwen WU, Gongtian SHEN, and Junjiao ZHANG. “Characteristics of acoustic emission signals in the rolling bearing on giant wheel”. In: *Proceedings of the 30th European Conference on Acoustic Emission Testing and 7th International Conference on Acoustic Emission*. Vol. 532. 2012.

[119] Takashi SAKO and Osamu YOSHIE. “Diagnostic method of low speed rolling element bearing using AE envelope waveform”. In: *TENCON 2010-2010 IEEE Region 10 Conference*. IEEE. 2010, pp. 724–729.

[120] Juha MIETTINEN and Pentti PATANIITTY. “Acoustic emission in monitoring extremely slowly rotating rolling bearing”. In: *Proceedings of COMADEM*. Vol. 99. 1999, pp. 289–297.

[121] Nordin JAMALUDIN and D MBA. “Monitoring extremely slow rolling element bearings: part I”. In: *NDT & E International* 35.6 (2002), pp. 349–358.

[122] Mari NAGATA et al. “Evaluation of tribological properties of bearing materials for marine diesel engines utilizing acoustic emission technique”. In: *Tribology International* 46.1 (2012), pp. 183–189.

[123] Dubravko MILJKOVIĆ. “Brief review of motor current signature analysis”. In: *HDKBR Info magazin* 5.1 (2015), pp. 14–26.

[124] W. QIAO and D. LU. “A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis x2014;Part I: Components and Subsystems”. In: *IEEE Transactions on Industrial Electronics* 62.10 (2015), pp. 6536–6545. ISSN: 0278-0046. DOI: 10.1109/TIE.2015.2422112.

[125] Javier SERRANO-GONZÁLEZ and Roberto LACAL-ARÁNTEGUI. “Technological evolution of onshore wind turbines — a market-based analysis”. In: *Wind Energy* 19.12 (2016), pp. 2171–2187.

[126] W. QIAO and D. LU. “A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis x2014;Part II: Signals and Signal Processing Methods”. In: *IEEE Transactions on Industrial Electronics* 62.10 (2015), pp. 6546–6557. ISSN: 0278-0046. DOI: 10.1109/TIE.2015.2422394.

[127] Marcelo Nesci SOARES et al. “Vibration-Based Rotor-Side-Converter Open-Switch-Fault Detection in DFIGs for Wind Turbines”. In: *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE. 2018, pp. 1–6.

[128] M El Hachemi BENBOUZID. “A review of induction motors signature analysis as a medium for faults detection”. In: *IEEE transactions on industrial electronics* 47.5 (2000), pp. 984–993.

[129] Neelam MEHALA. “Current signature analysis for condition monitoring of motors”. In: *International Journal of Electronics and Computer Science Engineering* 1.3 (2011), pp. 1629–1633.

[130] TG AMARAL et al. “Image Processing based Classifier for Detection and Diagnosis of Induction Motor Stator Fault”. In: *Image Processing*. IntechOpen, 2009.

[131] SM SHASHIDHARA and PS RAJU. “Stator Winding Fault Diagnosis of Three Phase Induction Motor by Park’s Vector Approach”. In: *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* 2.7 (2013), pp. 2901–2906.

[132] Hayder E ALWAN and Qais S AL-SABBAGH. “Detection of Static Air-Gap Eccentricity in Three Phase Induction Motor by Using Artificial Neural Network (ANN)”. In: *Journal of Engineering* 15.4 (2009), pp. 4176–4192.

[133] VN GHATE and SV DUDUL. “Artificial neural network based fault classifier for three phase induction motor”. In: *International Journal of Computational Intelligence Research* 5.1 (2009), pp. 25–37.

[134] M KHAZAEE et al. “Feature-level fusion based on wavelet transform and artificial neural network for fault diagnosis of planetary gearbox using acoustic and vibration signals”. In: *Insight-Non-Destructive Testing and Condition Monitoring* 55.6 (2013), pp. 323–330.

[135] H CROHAS and P LEPERT. “Damage-detection monitoring method for offshore platforms is field-tested”. In: *Oil Gas J.;(United States)* 80.8 (1982).

[136] GOOGLE. *Google Ngram Viewer*. 2019. URL: https://books.google.com/ngrams/graph?content=vibration+monitoring%2C+structural+monitoring%2C+bearing+monitoring%2C+gear+monitoring&year_start=1900&year_end=2000&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cvibration%20monitoring%3B%2Cc0%3B.t1%3B%2Cstructural%20monitoring%3B%2Cc0%3B.t1%3B%2Cbearing%20monitoring%3B%2Cc0%3B.t1%3B%2Cgear%20monitoring%3B%2Cc0 (visited on 07/05/2019).

[137] E Peter CARDEN and Paul FANNING. "Vibration based condition monitoring: a review". In: *Structural health monitoring* 3.4 (2004), pp. 355–377.

[138] WIKIPEDIA. *Cyber-Physical Systems*. 2019. URL: https://en.wikipedia.org/wiki/Cyber-physical_system (visited on 07/10/2019).

[139] Trillium NETWORK. *The rise of big data and industry 4.0*. 2019. URL: http://trilliummfg.ca/the-rise-of-big-data-and-industry-4-0/ (visited on 07/10/2019).

[140] K. SHVACHKO et al. "The Hadoop Distributed File System". In: *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*. 2010, pp. 1–10. DOI: 10.1109/MSST.2010.5496972.

[141] Wind EUROPE. *Wind energy in Europe in 2018*. 2019. URL: https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Statistics-2018.pdf (visited on 07/12/2019).

[142] J HELSEN et al. "Experimental investigation of bearing slip in a wind turbine gearbox during a transient grid loss event". In: *Wind Energy* 19.12 (2016), pp. 2255–2269.

[143] María Isabel BLANCO. "The economics of wind energy". In: *Renewable and Sustainable Energy Reviews* 13.6 (2009), pp. 1372–1382.

[144] Yanhui FENG et al. "Use of SCADA and CMS signals for failure detection and diagnosis of a wind turbine gearbox". In: *European Wind Energy Conference and Exhibition 2011, EWEC 2011*. Sheffield. 2011, pp. 17–19.

[145] Pramod BANGALORE. “Load and Risk Based Maintenance Management of Wind Turbines”. PhD thesis. Chalmers University of Technology, 2014.

[146] J HELSEN et al. “Experimental dynamic identification of mode-shape driving wind turbine grid loss event on nacelle testrig”. In: *Renewable Energy* 85 (2016), pp. 259–272.

[147] NREL. *Statistics Show Bearing Problems Cause the Majority of Wind Turbine Gearbox Failures*. 2019. URL: https://www.energy.gov/eere/wind/articles/statistics-show-bearing-problems-cause-majority-wind-turbine-gearbox-failures (visited on 07/12/2019).

[148] Brian P GRANEY and Ken STARRY. “Rolling element bearing analysis”. In: *Materials Evaluation* 70.1 (2012), p. 78.

[149] Robert B RANDALL and Jerome ANTONI. “Rolling element bearing diagnostics—a tutorial”. In: *Mechanical Systems and Signal Processing* 25.2 (2011), pp. 485–520.

[150] Biswanath SAMANTA, Khamis R AL-BALUSHI, and Saeed A AL-ARAIMI. “Artificial neural networks and genetic algorithm for bearing fault detection”. In: *Soft Computing* 10.3 (2006), pp. 264–271.

[151] Didier RÉMOND, Jérôme ANTONI, and Robert RANDALL. “Editorial for the special issue on Instantaneous Angular Speed (IAS) processing and angular applications”. In: *Mechanical Systems and Signal Processing* 1.44 (2014), pp. 1–4.

[152] CMMNO. *International Conference on Condition Monitoring of Machinery in Non-Stationary Operation*. 2019. URL: https://cmmno2018.unican.es/ (visited on 07/12/2019).

[153] Quentin Leclere HUGO ANDRE Ilyes Khelf. “Harmonic Product Spectrum revisited and adapted for rotating machine monitoring based on IAS”. In: *Surveillance9*. 2017.

[154] Laurence RENAUDIN et al. “Natural roller bearing fault detection by angular measurement of true instantaneous angular speed”. In: *Mechanical Systems and Signal Processing* 24.7 (2010), pp. 1998–2011.

[155] Ahmed YOUSEF BEN SASI et al. “Instantaneous angular speed monitoring of electric motors”. In: *Journal of Quality in Maintenance Engineering* 10.2 (2004), pp. 123–135.

[156] PD MCFADDEN and JD SMITH. “Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review”. In: *Tribology international* 17.1 (1984), pp. 3–10.

[157] David HOCHMANN and Eric BECHHOEFER. “Envelope bearing analysis: theory and practice”. In: *2005 IEEE Aerospace Conference*. IEEE. 2005, pp. 3658–3666.

[158] P BORGHESANI et al. “A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions”. In: *Mechanical Systems and Signal Processing* 38.1 (2013), pp. 23–35.

[159] D HO and RB RANDALL. “Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals”. In: *Mechanical systems and signal processing* 14.5 (2000), pp. 763–788.

[160] PD MCFADDEN and JD SMITH. “Model for the vibration produced by a single point defect in a rolling element bearing”. In: *Journal of sound and vibration* 96.1 (1984), pp. 69–82.

[161] J ANTONI and RB RANDALL. “A stochastic model for simulation and diagnostics of rolling element bearings with localized faults”. In: *Journal of vibration and acoustics* 125.3 (2003), pp. 282–289.

[162] J ANTONI and RB RANDALL. “Unsupervised noise cancellation for vibration signals: part II—a novel frequency-domain algorithm”. In: *Mechanical Systems and Signal Processing* 18.1 (2004), pp. 103–117.

[163] J ANTONI and RB RANDALL. “Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms”. In: *Mechanical Systems and Signal Processing* 18.1 (2004), pp. 89–101.

[164] Nader SAWALHI and Robert B RANDALL. “The application of spectral kurtosis to bearing diagnostics”. In: *Proceedings of ACOUSTICS*. 2004, pp. 3–5.

[165] D ABBOUD et al. “Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment”. In: *Mechanical Systems and Signal Processing* 84 (2017), pp. 200–226.

[166] Dany ABBOUD et al. “Deterministic-random separation in nonstationary regime”. In: *Journal of Sound and Vibration* 362 (2016), pp. 305–326.

[167] RB RANDALL and N SAWALHI. “Editing time signals using the real cepstrum”. In: *MFPT Conference, Virginia Beach (May 2011)*. 2011.

[168] J LEE et al. “Rexnord technical services: Bearing data set”. In: *Moffett Field, CA: IMS, Univ. Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames* (2007).

[169] Bruce P BOGERT, Michael JR HEALY, and John W TUKEY. “The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking”. In: *Proceedings of the symposium on time series analysis*. Vol. 15. chapter. 1963, pp. 209–243.

[170] Alan V OPPENHEIM, Ronald W SCHAFER, and Thomas G STOCKHAM JR. “Nonlinear filtering of multiplied and convolved signals”. In: *Audio and Electroacoustics, IEEE Transactions on* 16.3 (1968), pp. 437–466.

[171] Robert B RANDALL. “A history of cepstrum analysis and its application to mechanical problems”. In: *International Conference at Institute of Technology of Chartres, France*. 2013, pp. 11–16.

[172] Donald G CHILDERS, David P SKINNER, and Robert C KEMERAIT. “The cepstrum: A guide to processing”. In: *Proceedings of the IEEE* 65.10 (1977), pp. 1428–1443.

[173] RB RANDALL and Jens HEE. “Cepstrum analysis”. In: *Wireless World* 88 (1982), pp. 77–80.

[174] A POLYDOROS and Adly T FAM. “The differential cepstrum: definition and properties”. In: *Proc. IEEE Int. Symp. Circuits Syst.* 1981, pp. 77–80.

[175] J ANTONI, J DANIERE, and F GUILLET. “Blind identification of nonminimum phase systems using the mean differential cepstrum”. In: *Signal Processing Conference, 2000 10th European*. IEEE. 2000, pp. 1–4.

[176] Robert Bond RANDALL. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011.

[177] TC RATHBONE. “Vibration tolerance”. In: *Power Plant Engineering* 43.1939 (1939), pp. 721–724.

[178] HG YATES. “Vibration diagnosis in marine geared turbines”. In: *Trans. North East Coast Institute of Eng. and Shipbuilders* 65 (1949).

[179] Eugene P WIGNER. “On the quantum correction for thermodynamic equilibrium”. In: *Part I: Physical Chemistry. Part II: Solid State Physics*. Springer, 1997, pp. 110–120.

[180] Jean VILLE. “Theorie et application dela notion de signal analytique”. In: *Câbles et transmissions* 2.1 (1948), pp. 61–74.

[181] Ingrid DAUBECHIES, Jianfeng LU, and Hau-Tieng WU. “Synchrosqueezed wavelet transforms: A tool for empirical mode decomposition”. In: *arXiv preprint arXiv:0912.2437* (2009).

[182] Dmytro IATSENKO, Peter VE MCCLINTOCK, and Aneta STEFANOVSKA. “Extraction of instantaneous frequencies from ridges in time–frequency representations of signals”. In: *Signal Processing* 125 (2016), pp. 290–303.

[183] Boualem BOASHASH and Peter BLACK. “An efficient real-time implementation of the Wigner-Ville distribution”. In: *IEEE transactions on acoustics, speech, and signal processing* 35.11 (1987), pp. 1611–1618.

[184] H-I CHOI and William J WILLIAMS. “Improved time-frequency representation of multicomponent signals using exponential kernels”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37.6 (1989), pp. 862–871.

[185] Yunxin ZHAO, Les E ATLAS, and Robert J MARKS. “The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38.7 (1990), pp. 1084–1091.

[186] Boualem BOASHASH. *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press, 2015.

[187] Quentin LECLERE, Hugo ANDRÉ, and Jérôme ANTONI. “A multi-order probabilistic approach for Instantaneous Angular Speed tracking debriefing of the CMMNO 14 diagnosis contest”. In: *Mechanical Systems and Signal Processing* 81 (2016), pp. 375–386.

[188] Ralph A WIGGINS. “Minimum entropy deconvolution”. In: *Geoexploration* 16.1-2 (1978), pp. 21–35.

[189] Jakub OBUCHOWSKI, Radosław ZIMROZ, and Agnieszka WYŁOMAŃSKA. “Blind equalization using combined skewness–kurtosis criterion for gearbox vibration enhancement”. In: *Measurement* 88 (2016), pp. 34–44.

[190] Ron POTTER. “A new order tracking method for rotating machinery”. In: *Sound and Vibration* 24.9 (1990), pp. 30–34.

[191] KR FYFE and EDS MUNCK. “Analysis of computed order tracking”. In: *Mechanical Systems and Signal Processing* 11.2 (1997), pp. 187–205.

[192] KS WANG and Philippus Stephanus HEYNS. “Application of computed order tracking, Vold–Kalman filtering and EMD in rotating machine vibration”. In: *Mechanical Systems and Signal Processing* 25.1 (2011), pp. 416–430.

[193] Hyuno KIM et al. “Visual encoder: robust and precise measurement method of rotation angle via high-speed RGB vision”. In: *Optics Express* 24.12 (2016), pp. 13375–13386.

[194] Alvaro COLLET, Manuel MARTINEZ, and Siddhartha S SRINIVASA. “The MOPED framework: Object recognition and pose estimation for manipulation”. In: *The International Journal of Robotics Research* 30.10 (2011), pp. 1284–1306. DOI: 10.1177/0278364911401765. eprint: https://doi.org/10.1177/0278364911401765. URL: https://doi.org/10.1177/0278364911401765.

[195] David G LOWE. “Distinctive image features from scale-invariant keypoints”. In: *International journal of computer vision* 60.2 (2004), pp. 91–110.

[196] Herbert BAY et al. “Speeded-up robust features (SURF)”. In: *Computer vision and image understanding* 110.3 (2008), pp. 346–359.

[197] Young-shin KWON and Won-jong KIM. “Development of a new high-resolution angle-sensing mechanism using an RGB sensor”. In: *IEEE/ASME Transactions on Mechatronics* 19.5 (2014), pp. 1707–1715.

[198] Takamasa SUZUKI et al. “Two-dimensional small-rotation-angle measurement using an imaging method”. In: *Optical Engineering* 45.4 (2006), p. 043604.

[199] Weimin LI et al. “Method of rotation angle measurement in machine vision based on calibration pattern with spot array”. In: *Applied optics* 49.6 (2010), pp. 1001–1006.

[200] Kok-Meng LEE and Debao ZHOU. “A real-time optical sensor for simultaneous measurement of three-DOF motions”. In: *IEEE/ASME Transactions on Mechatronics* 9.3 (2004), pp. 499–507.

[201] Tomohiko KADOWAKI, Kazuyuki KOBAYASHI, and Kajiro WATANABE. “Rotation angle measurement of high-speed flying object”. In: *SICE-ICASE, 2006. International Joint Conference*. IEEE. 2006, pp. 5256–5259.

[202] Yunfan WANG, Lijuan WANG, and Yong YAN. “Rotational speed measurement through digital imaging and image processing”. In: *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*. IEEE. 2017, pp. 1–6.

[203] Tianyu WANG et al. “Rotational speed measurement using a low-cost imaging device and image processing algorithms”. In: *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE. 2018, pp. 1–6.

[204] Jacek URBANEK et al. “Comparison of amplitude-based and phase-based methods for speed tracking in application to wind turbines”. In: *Metrology and measurement systems* 18.2 (2011), pp. 295–304.

[205] Henry K KWOK and Douglas L JONES. “Improved instantaneous frequency estimation using an adaptive short-time Fourier transform”. In: *IEEE transactions on signal processing* 48.10 (2000), pp. 2964–2972.

[206] Shiufun CHEUNG and Jae S LIM. “Combined multiresolution (wideband/narrow-band) spectrogram”. In: *IEEE Transactions on signal processing* 40.4 (1992), pp. 975–977.

[207] ZK PENG et al. “Polynomial chirplet transform with application to instantaneous frequency estimation”. In: *IEEE Transactions on Instrumentation and Measurement* 60.9 (2011), pp. 3222–3229.

[208] S Chandra SEKHAR and TV SREENIVAS. “Effect of interpolation on PWVD computation and instantaneous frequency estimation”. In: *Signal processing* 84.1 (2004), pp. 107–116.

[209] Konstantinos C GRYLLIAS and Ioannis A ANTONIADIS. “Estimation of the instantaneous rotation speed using complex shifted Morlet wavelets”. In: *Mechanical Systems and Signal Processing* 38.1 (2013), pp. 78–95.

[210] José M ALLER et al. “Sensorless speed measurement of AC machines using analytic wavelet transform”. In: *IEEE Transactions on Industry Applications* 38.5 (2002), pp. 1344–1350.

[211] François AUGER and Patrick FLANDRIN. “Improving the readability of time-frequency and time-scale representations by the reassignment method”. In: *IEEE Transactions on signal processing* 43.5 (1995), pp. 1068–1089.

[212] Kunihiko KODERA, Roger GENDRIN, and C de VILLEDARY. “Analysis of time-varying signals with small BT values”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1 (1978), pp. 64–76.

[213] Juanjuan SHI et al. “Generalized stepwise demodulation transform and synchrosqueezing for time–frequency analysis and bearing fault diagnosis”. In: *Journal of Sound and Vibration* 368 (2016), pp. 202–222.

[214] Zhipeng FENG, Xiaowang CHEN, and Ming LIANG. “Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions”. In: *Mechanical Systems and Signal Processing* 52 (2015), pp. 360–375.

[215] Songtao XI et al. “A frequency-shift synchrosqueezing method for instantaneous speed estimation of rotating machinery”. In: *Journal of Manufacturing Science and Engineering* 137.3 (2015), p. 031012.

[216] Dmytro IATSENKO, Peter VE MCCLINTOCK, and Aneta STEFANOVSKA. “Linear and synchrosqueezed time-frequency representations revisited. Part I: Overview, standards of use, related issues and algorithms”. In: *arXiv preprint arXiv:1310.7215* (2013).

[217] Dmytro IATSENKO, Peter VE MCCLINTOCK, and Aneta STEFANOVSKA. “Linear and synchrosqueezed time-frequency representations revisited. part ii: Resolution, reconstruction and concentration”. In: *arXiv preprint arXiv:1310.7274* (2013).

[218] Dmytro IATSENKO, Peter VE MCCLINTOCK, and Aneta STEFANOVSKA. “Linear and synchrosqueezed time–frequency representations revisited: Overview, standards of use, resolution, reconstruction, concentration, and algorithms”. In: *Digital Signal Processing* 42 (2015), pp. 1–26.

[219] Ross F BARRETT and David A HOLDSWORTH. “Frequency tracking using hidden Markov models with amplitude and phase information”. In: *IEEE Transactions on Signal Processing* 41.10 (1993), pp. 2965–2976.

[220] Stephan SCHMIDT, Philippus Stephanus HEYNS, and Johan Pieter DE VILLIERS. “A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process”. In: *Mechanical Systems and Signal Processing* 100 (2018), pp. 630–646.

[221] Igor DJUROVIĆ and LJubiša STANKOVIĆ. “An algorithm for the Wigner distribution based instantaneous frequency estimation in a high noise environment”. In: *Signal Processing* 84.3 (2004), pp. 631–643.

[222] Shibin WANG et al. “Nonlinear squeezing time–frequency transform for weak signal detection”. In: *signal processing* 113 (2015), pp. 195–210.

[223] Radosław ZIMROZ et al. “Measurement of instantaneous shaft speed by advanced vibration signal processing-application to wind turbine gearbox”. In: *Metrology and Measurement Systems* 18.4 (2011), pp. 701–712.

[224] Cédric PEETERS et al. “Vibration-based angular speed estimation for multi-stage wind turbine gearboxes”. In: *Journal of Physics: Conference Series* 842.1 (2017), p. 012053. URL: http://stacks.iop.org/1742-6596/842/i=1/a=012053.

[225] Frédéric BONNARDOT et al. “Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)”. In: *Mechanical Systems and Signal Processing* 19.4 (2005), pp. 766–785.

[226] Jacek URBANEK, Tomasz BARSZCZ, and Jerome ANTONI. “A two-step procedure for estimation of instantaneous rotational speed with large fluctuations”. In: *Mechanical Systems and Signal Processing* 38.1 (2013), pp. 96–102.

[227] Francois COMBET and Leonid GELMAN. “An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor”. In: *Mechanical systems and signal processing* 21.6 (2007), pp. 2590–2606.

[228] A-O BOUDRAA et al. “IF estimation using empirical mode decomposition and nonlinear Teager energy operator”. In: *Control, Communications and Signal Processing, 2004. First International Symposium on*. IEEE. 2004, pp. 45–48.

[229] RB RANDALL and WA SMITH. “Use of the Teager Kaiser Energy Operator to estimate machine speed”. In: *Paper to be presented at PHM Europe conference, Bilbao, Spain*. 2016, pp. 5–8.

[230] O CARDONA-MORALES, LD AVENDAÑO, and G CASTELLANOS-DOMÍNGUEZ. “Nonlinear model for condition monitoring of non-stationary vibration signals in ship driveline application”. In: *Mechanical Systems and Signal Processing* 44.1-2 (2014), pp. 134–148.

[231] Barbara F LA SCALA and Robert R BITMEAD. “Design of an extended Kalman filter frequency tracker”. In: *IEEE Transactions on Signal Processing* 44.3 (1996), pp. 739–742.

[232] Francois COMBET and Radoslaw ZIMROZ. “A new method for the estimation of the instantaneous speed relative fluctuation in a vibration signal based on the short time scale transform”. In: *Mechanical Systems and Signal Processing* 23.4 (2009), pp. 1382–1397.

[233] Michael FELDMAN. “Hilbert transform in vibration analysis”. In: *Mechanical systems and signal processing* 25.3 (2011), pp. 735–802.

[234] Michael D COATS and RB RANDALL. “Order-Tracking with and without a tacho signal for gear fault diagnostics”. In: *Proceedings of Acoustics*. 2012, pp. 1–6.

[235] Robert B RANDALL and Wade A SMITH. “Accuracy of speed determination of a machine from frequency demodulation of response vibration signals”. In: (2018).

[236] Jérôme ANTONI et al. “Feedback on the Surveillance 8 challenge: vibration-based diagnosis of a Safran aircraft engine”. In: *Mechanical Systems and Signal Processing* 97 (2017), pp. 112–144.

[237] D HANSON et al. “Cyclostationarity and the cepstrum for operational modal analysis of mimo systems—Part I: Modal parameter identification”. In: *Mechanical systems and signal processing* 21.6 (2007), pp. 2441–2458.

[238] Dmytro IATSENKO, Peter VE MCCLINTOCK, and Aneta STEFANOVSKA. “On the extraction of instantaneous frequencies from ridges in time-frequency representations of signals”. In: *arXiv preprint arXiv:1310.7276* (2013).

[239] C.M. BISHOP. *Pattern recognition and machine learning*. Springer, 2006.

[240] M-Ch PAN and Y-F LIN. “Further exploration of Vold–Kalman-filtering order tracking with shaft-speed information—I: Theoretical part, numerical implementation and parameter investigations”. In: *Mechanical systems and signal processing* 20.5 (2006), pp. 1134–1154.

[241] Jason R BLOUGH. “A survey of DSP methods for rotating machinery analysis, what is needed, what is available”. In: *Journal of sound and vibration* 262.3 (2003), pp. 707–720.

[242] Jiri TUMA. “Setting the passband width in the Vold-Kalman order tracking filter”. In: *Twelfth International Conference on Sound and Vibration (ICSV12), Lisbon, Portugal, July*. 2005, pp. 11–14.

[243] Robert B RANDALL and Nader SAWALHI. “A new method for separating discrete components from a signal”. In: *Sound and Vibration* 45.5 (2011), p. 6.

[244] N SAWALHI and R RANDALL. “Signal pre-whitening using cepstrum editing (liftering) to enhance fault detection in rolling element bearings”. In: *Proceedings of the 24 international congress on condition monitoring and diagnostic engineering management (comadem2011), may*. 2011, pp. 330–336.

[245] P BORGHESANI et al. “Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions”. In: *Mechanical Systems and Signal Processing* 36.2 (2013), pp. 370–384.

[246] Wade A SMITH and Robert B RANDALL. “Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study”. In: *Mechanical Systems and Signal Processing* 64 (2015), pp. 100–131.

[247] Robert B RANDALL and Nader SAWALHI. “Cepstral removal of periodic spectral components from time signals”. In: *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*. Springer, 2014, pp. 313–324.

[248] Agusmian Partogi OMPUSUNGGU. “Automated cepstral editing procedure (ACEP) as a signal pre-processing in vibration-based bearing fault diagnostics”. In: ().

[249] Steven F BOLL. “Suppression of acoustic noise in speech using spectral subtraction”. In: *Acoustics, Speech and Signal Processing, IEEE Transactions on* 27.2 (1979), pp. 113–120.

[250] Rainer MARTIN. “Spectral subtraction based on minimum statistics”. In: *power* 6 (1994), p. 8.

[251] Katia LEBART, Jean-Marc BOUCHER, and PN DENBIGH. “A new method based on spectral subtraction for speech dereverberation”. In: *Acta Acustica united with Acustica* 87.3 (2001), pp. 359–366.

[252] Sunil KAMATH and Philipos LOIZOU. “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise”. In: *IEEE international conference on acoustics speech and signal processing*. Vol. 4. Citeseer. 2002, pp. 4164–4164.

[253] David L DONOHO and Jain M JOHNSTONE. “Ideal spatial adaptation by wavelet shrinkage”. In: *Biometrika* 81.3 (1994), pp. 425–455.

[254] L PASTI et al. “Optimization of signal denoising in discrete wavelet transform”. In: *Chemometrics and intelligent laboratory systems* 48.1 (1999), pp. 21–34.

[255] Rami COHEN. “Signal denoising using wavelets”. In: *Project Report, Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa* (2012).

[256] Blackman RALPH BEEBE and Tukey JOHN. “The measurement of power spectra”. In: (1958).

[257] J ANTONI. “Cyclic spectral analysis of rolling-element bearing signals: facts and fictions”. In: *Journal of Sound and vibration* 304.3 (2007), pp. 497–529.

[258] P BORGHESANI et al. “Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals”. In: *Mechanical Systems and Signal Processing* 40.1 (2013), pp. 38–55.

[259] S SHENG. “Wind turbine gearbox condition monitoring round robin study–Vibration analysis”. In: *Contract* 303 (2012), pp. 275–3000.

[260] PD MCFADDEN and JD SMITH. “The vibration produced by multiple point defects in a rolling element bearing”. In: *Journal of sound and vibration* 98.2 (1985), pp. 263–273.

[261] Robert RANDALL and Wade SMITH. “New cepstral methods for the diagnosis of gear and bearing faults under variable speed conditions”. In: *ICSV23 conference, Athens, July*. 2016, pp. 10–14.

[262] Jerome ANTONI. “Fast computation of the kurtogram for the detection of transient faults”. In: *Mechanical Systems and Signal Processing* 21.1 (2007), pp. 108–124.

[263] Radoslaw ZIMROZ and Walter BARTELMUS. “Gearbox condition estimation using cyclo-stationary properties of vibration signal”. In: *Key Engineering Materials*. Vol. 413. Trans Tech Publ. 2009, pp. 471–478.

[264] J ANTONI, J DANIERE, and F GUILLET. “Effective vibration analysis of ic engines using cyclostationarity. part ia methodology for condition monitoring”. In: *Journal of sound and vibration* 257.5 (2002), pp. 815–837.

[265] Jérôme ANTONI et al. “Cyclostationary modelling of rotating machine vibration signals”. In: *Mechanical systems and signal processing* 18.6 (2004), pp. 1285–1314.

[266] AC MCCORMICK and AK NANDI. “Cyclostationarity in rotating machine vibrations”. In: *Mechanical systems and signal processing* 12.2 (1998), pp. 225–242.

[267] I ANTONIADIS and G GLOSSIOTIS. “Cyclostationary analysis of rolling-element bearing vibration signals”. In: *Journal of sound and vibration* 248.5 (2001), pp. 829–845.

[268] C CAPDESSUS, M SIDAHMED, and JL LACOUME. “Cyclostationary processes: application in gear faults early diagnosis”. In: *Mechanical systems and signal processing* 14.3 (2000), pp. 371–385.

[269] Jérôme ANTONI, Ge XIN, and Nacer HAMZAOUI. “Fast computation of the spectral correlation”. In: *Mechanical Systems and Signal Processing* 92 (2017), pp. 248–277.

[270] Robert B RANDALL, Jérôme ANTONI, and S CHOBSAARD. “The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals”. In: *Mechanical systems and signal processing* 15.5 (2001), pp. 945–962.

[271] Emmanuel CANDES, Justin ROMBERG, and Terence TAO. “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. In: *arXiv preprint math/0409186* (2004).

[272] Emmanuel CANDES and Terence TAO. “Near optimal signal recovery from random projections: Universal encoding strategies?” In: *arXiv preprint math/0410542* (2004).

[273] David L DONOHO et al. “Compressed sensing”. In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.

[274] David L DONOHO. “For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution”. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 59.6 (2006), pp. 797–829.

[275] *Web of Science research records for sparsity*. `http://wcs.webofknowledge.com/RA/analyze.do?product=WOS&SID=E6K2nUlrn1a36enVBn3&field=PY_PublicationYear_PublicationYear_en&yearSort=true`. Accessed: 2019-03-12.

[276] Jerome ANTONI. “The infogram: Entropic evidence of the signature of repetitive transients”. In: *Mechanical Systems and Signal Processing* 74 (2016), pp. 73–94.

[277] Carlos A CABRELLI. “Minimum entropy deconvolution and simplicity: A noniterative algorithm”. In: *Geophysics* 50.3 (1985), pp. 394–413.

[278] C-S WANG, Jin TANG, and Bo ZHAO. “An improvement on D norm deconvolution: A fast algorithm and the related procedure”. In: *Geophysics* 56.5 (1991), pp. 675–680.

[279] William Charles GRAY. *Variable norm deconvolution*. 19. Stanford University Ph. D. thesis, 1979.

[280] Geoff L MCDONALD and Qing ZHAO. “Multipoint optimal minimum entropy deconvolution and convolution fix: Application to vibration fault detection”. In: *Mechanical Systems and Signal Processing* 82 (2017), pp. 461–477.

[281] H ENDO and RB RANDALL. “Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter”. In: *Mechanical Systems and Signal Processing* 21.2 (2007), pp. 906–919.

[282] Rafael José Gomes de OLIVEIRA and Mauro Hugo MATHIAS. “Maximization of the Signal Impulsiveness Combining Envelope Technique with Minimum Entropy Deconvolution”. In: *Applied Mechanics and Materials*. Vol. 392. Trans Tech Publ. 2013, pp. 725–729.

[283] H ENDO, RB RANDALL, and C GOSSELIN. “Differential diagnosis of spall vs. cracks in the gear tooth fillet region: experimental validation”. In: *Mechanical Systems and Signal Processing* 23.3 (2009), pp. 636–651.

[284] Tomasz BARSZCZ and Nader SAWALHI. “Fault detection enhancement in rolling element bearings using the minimum entropy deconvolution”. In: *Archives of acoustics* 37.2 (2012), pp. 131–141.

[285] Dae-Ho KWAK et al. “Fault detection of roller-bearings using signal processing and optimization algorithms”. In: *Sensors* 14.1 (2014), pp. 283–298.

[286] Wade A SMITH et al. “Use of cyclostationary properties to diagnose planet bearing faults in variable speed conditions”. In: *10th DST group international conference on health and usage monitoring systems, 17th Australian aerospace congress*. 2017, pp. 26–28.

[287] J ANTONI. “Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions”. In: *Journal of Sound and vibration* 304.3-5 (2007), pp. 497–529.

[288] Marco BUZZONI, Jérôme ANTONI, and Gianluca D’ELIA. “Blind deconvolution based on cyclostationarity maximization and its application to fault identification”. In: *Journal of Sound and Vibration* 432 (2018), pp. 569–601.

[289] Mingfang WANG et al. “Harmonic L2/L1 Norm for bearing fault diagnosis”. In: *IEEE Access* (2019).

[290] Petre STOICA, Randolph L MOSES, et al. “Spectral analysis of signals”. In: (2005).

[291] Roger A HOM and Charles R JOHNSON. “Matrix analysis”. In: *Cambridge University Express* (1985).

[292] RV MISES and Hilda POLLACZEK-GEIRINGER. “Praktische Verfahren der Gleichungsauflösung.” In: *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 9.2 (1929), pp. 152–164.

[293] Audrey REPETTI et al. “Euclid in a Taxicab: Sparse Blind Deconvolution with Smoothed l1/l2 Regularization”. In: *IEEE signal processing letters* 22.5 (2015), pp. 539–543.

[294] Michael ZIBULEVSKY and Barak A PEARLMUTTER. “Blind source separation by sparse decomposition in a signal dictionary”. In: *Neural computation* 13.4 (2001), pp. 863–882.

[295] Boaz BARAK, Jonathan A KELNER, and David STEURER. “Rounding sum-of-squares relaxations”. In: *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. ACM. 2014, pp. 31–40.

[296] Morten MORUP, Kristoffer Hougaard MADSEN, and Lars Kai HANSEN. “Approximate L0 constrained non-negative matrix and tensor factorization”. In: *2008 IEEE International Symposium on Circuits and Systems*. IEEE. 2008, pp. 1328–1331.

[297] Niall HURLEY and Scott RICKARD. “Comparing measures of sparsity”. In: *IEEE Transactions on Information Theory* 55.10 (2009), pp. 4723–4741.

[298] Patrik O HOYER. “Non-negative matrix factorization with sparseness constraints”. In: *Journal of machine learning research* 5.Nov (2004), pp. 1457–1469.

[299] Xiaodong JIA et al. “Sparse filtering with the generalized lp/lq norm and its applications to the condition monitoring of rotating machinery”. In: *Mechanical Systems and Signal Processing* 102 (2018), pp. 198–213.

[300] P BORGHESANI et al. “The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery”. In: *Mechanical Systems and Signal Processing* 44.1-2 (2014), pp. 118–133.

[301] Hai QIU et al. “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics”. In: *Journal of sound and vibration* 289.4-5 (2006), pp. 1066–1090.

[302] Jérôme ANTONI and Pietro BORGHESANI. “A statistical methodology for the design of condition indicators”. In: *Mechanical Systems and Signal Processing* 114 (2019), pp. 290–327.

[303] Henrique Dias Machado de AZEVEDO, Alex Maurício ARAÚJO, and Nadège BOUCHONNEAU. “A review of wind turbine bearing condition monitoring: State of the art and challenges”. In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 368–379.

[304] Bin LU et al. “A review of recent advances in wind turbine condition monitoring and fault diagnosis”. In: *Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009. IEEE*. IEEE. 2009, pp. 1–7.

[305] Pierre TCHAKOUA et al. “Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges”. In: *Energies* 7.4 (2014), pp. 2595–2630.

[306] Jaouher Ben ALI et al. “Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning”. In: *Applied Acoustics* 132 (2018), pp. 167–181.

[307] Junda ZHU et al. “Survey of condition indicators for condition monitoring systems”. In: *Annu. Conf. Progn. Heal. Manag. Soc.* Vol. 5. 2014, pp. 1–13.

[308] Gianluca D’ELIA et al. “Evolution of gear condition indicators for diagnostics of planetary gearboxes”. In: *The International Conference Surveillance 8*. FRA. 2015.

[309] P VEČEŘ, Marcel KREIDL, and R ŠMÍD. “Condition indicators for gearbox condition monitoring systems”. In: *Acta Polytechnica* 45.6 (2005).

[310] Akhand RAI and SH UPADHYAY. “A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings”. In: *Tribology International* 96 (2016), pp. 289–306.

[311] Vikas SHARMA and Anand PAREY. “A review of gear fault diagnosis using various condition indicators”. In: *Procedia Engineering* 144 (2016), pp. 253–263.

[312] Harry J DECKER, Robert F HANDSCHUH, and James J ZAKRAJSEK. “An enhancement to the NA4 gear vibration diagnostic parameter”. In: (1994).

[313] James J ZAKRAJSEK, Dennis P TOWNSEND, and Harry J DECKER. *An analysis of gear fault detection methods as applied to pitting fatigue failure data*. Tech. rep. NATIONAL AERONAUTICS and SPACE ADMINISTRATION CLEVELAND OH LEWIS RESEARCH CENTER, 1993.

[314] I Soltani BOZCHALOOI and Ming LIANG. “A smoothness index-guided approach to wavelet parameter selection in signal denoising and fault detection”. In: *Journal of Sound and Vibration* 308.1-2 (2007), pp. 246–267.

[315] Joshua S RICHMAN and J Randall MOORMAN. “Physiological time-series analysis using approximate entropy and sample entropy”. In: *American Journal of Physiology-Heart and Circulatory Physiology* 278.6 (2000), H2039–H2049.

[316] William A GARDNER. *Cyclostationarity in communications and signal processing*. Tech. rep. STATISTICAL SIGNAL PROCESSING INC YOUNTVILLE CA, 1994.

[317] Amod V DANDAWATE and Georgios B GIANNAKIS. “Statistical tests for presence of cyclostationarity”. In: *IEEE Transactions on signal processing* 42.9 (1994), pp. 2355–2369.

[318] William A GARDNER, Antonio NAPOLITANO, and Luigi PAURA. “Cyclostationarity: Half a century of research”. In: *Signal processing* 86.4 (2006), pp. 639–697.

[319] Jérôme ANTONI. “Cyclostationarity by examples”. In: *Mechanical Systems and Signal Processing* 23.4 (2009), pp. 987–1036.

[320] Goran D ŽIVANOVIĆ and William A GARDNER. “Degrees of cyclostationarity and their application to signal detection and estimation”. In: *Signal processing* 22.3 (1991), pp. 287–297.

[321] Amani RAAD, Jérôme ANTONI, and Ménad SIDAHMED. “Indicators of cyclostationarity: Theory and application to gear fault monitoring”. In: *Mechanical Systems and Signal Processing* 22.3 (2008), pp. 574–587.

[322] Scott ENSERINK and Douglas COCHRAN. “A cyclostationary feature detector”. In: *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*. Vol. 2. IEEE. 1994, pp. 806–810.

[323] MATLAB. *Signal-based condition indicators*. 2019. URL: `https://nl.mathworks.com/help/predmaint/ug/signal-based-condition-indicators.html` (visited on 08/05/2019).

[324] Anil K JAIN. “Data clustering: 50 years beyond K-means”. In: *Pattern recognition letters* 31.8 (2010), pp. 651–666.

[325] David J. C. MACKAY. “Bayesian interpolation”. In: *Neural computation* 4.3 (1992), pp. 415–447.

[326] Jim E GRIFFIN, Philip J BROWN, et al. “Inference with normal-gamma prior distributions in regression problems”. In: *Bayesian Analysis* 5.1 (2010), pp. 171–188.

[327] Bernhard SCHÖLKOPF, Alexander J SMOLA, Francis BACH, et al. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[328] Ralph C SMITH. *Uncertainty quantification: theory, implementation, and applications*. Vol. 12. Siam, 2013.

[329] Peter J ROUSSEEUW. “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.

[330] Huageng LUO et al. “Effective and accurate approaches for wind turbine gearbox condition monitoring”. In: *Wind Energy* 17.5 (2014), pp. 715–728.

[331] Jeremy SHELDON et al. “Robust wind turbine gearbox fault detection”. In: *Wind Energy* 17.5 (2014), pp. 745–755.

[332] David SIEGEL et al. “A comparative study on vibration-based condition monitoring algorithms for wind turbine drive trains”. In: *Wind Energy* 17.5 (2014), pp. 695–714.

[333] Zijun ZHANG, Anoop VERMA, and Andrew KUSIAK. “Fault analysis and condition monitoring of the wind turbine gearbox”. In: *IEEE Transactions on Energy Conversion* 27.2 (2012), pp. 526–535.

[334] Nader SAWALHI, Robert B RANDALL, and David FORRESTER. “Separation and enhancement of gear and bearing signals for the diagnosis of wind turbine transmission systems”. In: *Wind Energy* 17.5 (2014), pp. 729–743.

[335] Prasanna TAMILSELVAN et al. “A Two-Stage Diagnosis Framework for Wind Turbine Gearbox Condition Monitoring”. In: *IJPHM Special Issue on Wind Turbine PHM (Color)* (2013), p. 21.

[336] Hirotugu AKAIKE. “A new look at the statistical model identification”. In: *Selected Papers of Hirotugu Akaike*. Springer, 1974, pp. 215–222.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : Peeters DATE de SOUTENANCE :

Prénoms : Cédric

TITRE : Advanced signal processing for the identification and diagnosis of the condition of rotating machinery

NATURE : Doctorat Numéro d'ordre : 2019LYSEI107

Ecole doctorale :
MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE

Spécialité : MECANIQUE

RESUME :
Cette thèse porte sur des méthodes innovantes de contrôle de l'état de santé des machines tournantes par l'analyse des signaux vibratoires. En effet, la surveillance de l'état de santé des machines contribue à des améliorations substantiels des points de vue économique et de sureté. Afin d'y aboutir, l'une des manières les plus populaires est de recueillir les vibrations de la machine. La plupart de ces vibrations sont directement liées au comportement périodique des sous-systèmes de la machine tels que les arbres de rotation, engrenages, champs électriques rotationnels, etc. Cette connaissance peut être exploitée afin de concevoir une méthodologie adaptée à chaque type de défaut. Cette thèse s'intéresse aux étapes de la mise en œuvre de cette méthodologie. En règle générale, la première condition préalable à l'analyse avancée de l'information récoltée est la disponibilité de la vitesse intantanée de rotation. Cette vitesse doit être connue car la plupart des techniques du traitement du signal sont adaptées aux conditions de fonctionnement stationnaires. Ainsi, la connaissance de la vitesse permetra de compenser les fluctuations de vitesse, par exemple par le ré-échantillonnage angulaire du signal de vibration. Malgré l'existence d'outils de mesure permettant l'estimation de la vitesse tels que les codeurs et les tachymètres, cette thèse étudie le potentiel d'estimer la vitesse instantanée de rotation à partir des signaux vibratoires. Après l'estimation de la vitesse et le ré-échantillonnage angulaire, une étape suivante courante consiste à séparer le signal en composantes déterministes et stochastiques. Dans ce sens, l'efficacité et l'applicabilité de la procédure d'édition du cepstre sont anlysées. Ensuite, différentes méthodes de filtrage sont appliquées au signal résiduel afin d'améliorer le rapport signal sur bruit. Pour cette fin, les méthodes existantes utilisant des critères conventionnels sont étudiées en parallèle avec une nouvelle méthodologie aveugle de filtrage. La dernière étape du processus de traitement consiste à diagnostiquer le défaut potentiel. Ainsi, des indicateurs statistiques ont été calculés sur le signal obtenu après traitement et suivis dans le temps pour vérifier leurs variations. Dans de nombreux cas, la signature du défaut présente un comportement cyclostationnaire. Par conséquent, cette thèse examine également différentes techniques d'analyse de la cyclostationnarité. Enfin, les performances des différentes méthodes de traitement sont validées sur deux ensembles de données expérimentales de vibrations issues de boîtes de vitesses d'éoliennes.

MOTS-CLÉS : condition monitoring, signal processing, fault detection speed estimation, blind filtering

Laboratoire (s) de recherche : Laboratoire Vibrations Acoustique

Directeur de thèse: Jérôme Antoni & Jan Helsen

Président de jury :

Composition du jury :
Randall Robert, Professeur des Universités, UNSW, Président
Martin Nadine, Directeur de Recherche, CNRS Gibsa-lab , Rapporteure
Nejad Amir, Maître de Conférences HDR, NTNU, Rapporteur
Leclere Quentin, Maître de Conférences HDR, INSA-Lyon, Examinateur
Capdessus Cécile, Maître de Conférences, Université d'Orléans, Examinatrice
Antoni Jérôme, Professeur des Universités, INSA-Lyon, Directeur de thèse
Helsen Jan, Maître de Conférences, VUB, Directeur de thèse
Guillaume Patrick, Professeur des Universités, VUB, Co-directeur de thèse