

Investigation of Wind Turbine Rotor Imbalances Using Drivetrain Signals

MSc: Sustainable Energy Technology

Ilaria Leo

Investigation of Wind Turbine Rotor Imbalances Using Drivetrain Signals

by

Ilaria Leo

<u>Student Name</u>	<u>Student Number</u>
Ilaria Leo	5846196

Main supervisor: D. Zappalá
Post-doc Assistant: Sumit K Pal
Project Duration: January, 2024 - October 2024
Faculty: Faculty of Aerospace Engineering, Delft

Acknowledgments

This thesis represents the culmination of my journey as a Master's student at TU Delft. The past two years have been an incredible rollercoaster of emotions, filled with both rewarding and challenging moments that have led to significant personal growth.

I personally believe that every success is a shared journey, shaped by the contributions and support of many. For this reason, I would like to shine a light on the people who have been by my side throughout these months.

First of all, I want to express my sincere thanks to my main and daily supervisors, Dr. Donatella Zappalá and Sumit Kumar Pal, for their invaluable guidance and support throughout this entire process. I appreciate all the feedback received during our meetings, which were always constructive and stimulating.

I would like to express my heartfelt thanks to my family, especially my parents, for their endless support in helping me achieve my goals and realize my dreams. I am forever grateful for all you have done to make this journey possible.

To my friends at TU Delft, thank you for being an essential part of this chapter of my life. It wouldn't have been the same without you, and I will take care of all the memories we've shared. A special thank you to Margherita, Elena, and Federica for all the moments we've spent together and for always being there to encourage me during difficult times. Thank you for being my family abroad.

To my dear friend Benny, you are like a sister to me. To all my friends back home, thank you for always being there for me. We've grown together over the years, and I'm grateful for each step of the journey we continue to share.

Last but not least, I would like to thank Fernando for your constant presence and unconditional support. Having you by my side made me feel like everything is possible.

*Ilaria Leo
Delft, October 2024*

Abstract

Rotor imbalances—such as mass imbalance, pitch misalignment, and yaw misalignment—are critical faults in wind turbine systems. These imbalances cause uneven load distribution on components, leading to excessive wear, failures, increased operational costs due to unplanned downtime, and reduced energy output. Despite advancements in monitoring technologies, current maintenance strategies in wind turbines still rely on time-based manual inspections, as they lack reliable automated detection systems. This thesis addresses the need for a more efficient fault detection framework by integrating already available drivetrain Condition Monitoring System (CMS) vibration signals—commonly used to detect drivetrain component failures, like in gears and bearings—with traditional SCADA data. The aim is to extract signal features that capture the system's dynamic behavior and effectively detect and diagnose rotor imbalances. Notably, this approach overcomes the limitation of current systems not having direct measurements from the blades by leveraging operational data already collected from wind turbines.

Building on prior research, the proposed approach combines frequency and time-domain analyses and focuses on two key data sources: drivetrain vibration measurements and rotor speed data from the SCADA system. A decoupled simulation framework integrates aeroelastic simulations from OpenFAST with a multi-body drivetrain model in SIMPACK, specifically for the 10 MW DTU reference wind turbine. The results show that drivetrain velocity signals, particularly in the side-to-side direction, are highly sensitive to rotor imbalances, enabling accurate trend analysis. Features such as peak amplitudes at 1P and 3P frequencies form the basis of the fault detection and diagnosis criteria proposed in this thesis. By using the median values of their distributions, imbalances can be effectively detected and diagnosed. This approach also supports the implementation of a decision tree framework for real-time fault classification across various operating conditions.

The methodology was tested under both above and below-rated wind speeds, first in steady-state conditions and then in turbulent inflow scenarios. Additionally, health state indicators are proposed to recognize fault severity levels by clustering median value features within predefined ranges for low, medium, and high severity. This comprehensive monitoring approach effectively tracks fault progression across the imbalance scenarios under study. As a result, the proposed method lays the foundation for a future data-driven system that can reduce reliance on manual inspections and provide a scalable solution for predictive maintenance in wind turbine operations.

Contents

Acknowledgments	i
Abstract	ii
Nomenclature	x
1 Introduction	1
1.1 Research Objective and Questions	2
1.2 Report outline	3
2 State of the Art	4
2.1 An Overview of O&M	4
2.2 Wind Turbines Monitoring: Systems and Techniques	5
2.2.1 Vibration-based CMSs	8
2.3 Drivetrain Systems State of the Art	8
2.4 Damage of Wind Turbine Blades	10
2.5 Rotor Imbalances	11
2.5.1 Dynamic Effects of Rotor Imbalances	12
2.6 Signal Processing	13
2.7 Frequency Domain Analysis	13
2.7.1 Order Analysis	14
2.7.2 Windowing	15
2.8 Time-domain Analysis	16
2.9 Data Fusion	18
3 Methodology and Tools	21
3.1 The 10 MW Medium-speed NTNU Drivetrain Model	21
3.2 Decoupled Simulations Layout	24
3.3 Steady State Simulation Parameters	25
3.4 Turbulence	26
3.5 Analysis Method Outline	28
3.5.1 Frequency Domain Analysis and Comparison	29
3.5.2 Time Domain Analysis	31
3.6 Signal Choice	34
3.6.1 A Closer Look at the System Dynamics	35
4 Results: Steady State Analysis	37
4.1 Frequency Domain Analysis	37
4.1.1 FFT Analysis	37
4.1.2 Average Peak Amplitudes 3D Plots	38
4.1.3 Relative Amplitude Change Detection	40
4.1.4 Correlation Analysis	42
4.1.5 Box Plots and Median Values Multi-comparison Analysis	47
4.1.6 Pitch Misalignment vs. Mass Imbalance	50
4.2 Time Domain Analysis	51
4.3 Summary	54
5 Results: Turbulent Conditions Analysis	55
5.1 Number of Seeds Sensitivity Analysis	55
5.2 Frequency Domain Analysis	57
5.2.1 Averaged Peak Amplitudes Analysis	58
5.2.2 Correlation Analysis	61

5.2.3	Box Plots and Median Values Multi-comparison Analysis	63
5.3	Time Domain	66
5.4	Results Findings Summary	68
6	Discussion, Recommendations, and Limitations	70
6.1	Decision Tree for Fault Detection and Diagnosis	70
6.2	Health Monitoring	73
6.3	Recommendations and Limitations	77
7	Conclusions	78
	References	80
A	10 MW DTU Design Specifics	84
B	COT Resampling algorithm	85
B.1	Function Code	85
B.2	Algorithm flow-diagram	86
C	Steady State Analysis	87
C.1	3D averaged peak amplitudes plot	87
C.2	Relative change detection histograms	89
C.3	Box plots and Multiple Comparison analysis for MI vs PM imbalance cases	90
C.4	Time domain analysis for MI and YM	91
D	Steady-state data distribution and statistical median value analysis for all wind speeds	94
D.1	Box plots amplitude distribution analysis	94
D.2	Median values comparison analysis	96
E	Turbulence Analysis	98
E.1	3D averaged peak amplitude comparison plot	98
E.2	Relative change detection histograms comparison	101
E.3	Turbulence correlation analysis	102

List of Figures

1.1	Wind energy installation outlook 2022-2027 (GW) [2]	1
2.1	Condition monitoring categorization overview, inspired by Tchakoua et Al. [9]	5
2.2	Wind turbine components (source: https://windmillstech.com/wind-turbine-components/)	6
2.3	Sensor position in drive train vibration-based condition monitoring system [17]	8
2.4	Schematic layout of the two most common drivetrain configurations (a) with and (b) without gearbox (illustration by Amir Nejad [12])	9
2.5	View of wind turbine blade's parts [21]	11
2.6	Rotor imbalances classification and effects	12
2.7	Windows functions, inspired by Barszcz [10]	16
2.8	a) Raw level b) Feature level and c) Decision level data fusion methods, reproduced from Niu et al. (2019) [33]	19
3.1	NTNU 10 MW drivetrain multibody system model [36]	22
3.2	NTNU 10 MW gearbox schematic layout [36]	22
3.3	SIMPACK model sensor's location	23
3.4	Decoupled simulation procedure	24
3.5	DTU 10 MW turbine power curve	26
3.6	Example of turbulent inflow wind conditions of streamwise (longitudinal) direction component "u", generated through TurbSim	27
3.7	Example of turbulence effect on the rotor speed signal for the PM1 case at U4	28
3.8	Schematic data analysis workflow	33
3.9	Gearbox housing vibration signal for PM2 case and wind speed U4. Comparison of the different velocities measurements along the x,y, and z axes	36
4.1	FFT plot analysis for $v_{GB, \gamma}$ signal for $U_4=12$ m/s. Cases: (A)MI1, (B)PM1, (C)YM1	38
4.2	Averaged peak amplitudes of $v_{GB, \gamma}$ signal under varying health states and wind speed conditions: (A) MI - 1P, (B) PM - 1P, and (C) YM - 3P	39
4.3	Averaged 1P peak amplitudes of the rotor speed signal, ω_{rotor} : (A) MI,(B) PM	40
4.4	Normalised relative change in the averaged amplitudes of $v_{GB, \gamma}$: (A) MI - 1P, (B) PM - 1P, (C) YM - 3P	41
4.5	Correlation matrices between selected signals for 1P peak amplitudes ((A) MI, (B) PM and 3P amplitudes ((C) YM)	42
4.6	Mass imbalance - 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) MI1; (c) MI2; (d) MI3	43
4.7	Mass imbalance - 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6	43
4.8	Mass imbalance - 1P amplitudes scatter plots of $v_{GB, \gamma}$ and ω_{rotor} signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) MI1; (c) MI2; (d) MI3	44
4.9	Mass imbalance - 1P amplitudes scatter plots of $v_{GB, \gamma}$ and ω_{rotor} signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6	44
4.10	Pitch misalignment- 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) PM1; (c) PM2; (d) PM3	45

4.11 Pitch misalignment- 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6	45
4.12 Yaw misalignment- 3P amplitudes scatter plots of $v_{MB (front), \gamma}$ and $v_{MB (rear), \gamma}$ signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) YM1; (c) YM2; (d) YM3.	46
4.13 Yaw misalignment- 3P amplitudes scatter plots of $v_{MB (front), \gamma}$ and $v_{MB (rear), \gamma}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6	46
4.14 MI box plot- 1P amplitude distribution and variance of the selected monitoring signals aggregated for all wind speeds	47
4.15 PM box plot - 1P amplitude distribution and variance of the selected monitoring signals aggregated for all wind speeds	48
4.16 YM box plot - 3P amplitude distribution and variance of the selected monitoring signals aggregated for all wind speeds	48
4.17 MI - Multiple comparison median values analysis for CMS signals	49
4.18 PM - Multiple comparison median values analysis for CMS signals	49
4.19 YM - Multiple comparison median values analysis for CMS signals	49
4.20 $v_{GB, \gamma}$ signal box plots peak amplitude distribution analysis for healthy case compared to MI1, MI2, MI3 and PM1, PM2, PM3 imbalance scenarios	50
4.21 $v_{GB, \gamma}$ amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test)	50
4.22 ω_{rotor} signal box plots peak amplitude distribution analysis for healthy case compared to MI1, MI2, MI3 and PM1, PM2, PM3 imbalance scenarios	51
4.23 ω_{rotor} amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test)	51
4.24 PDFs of $a_{MB (front), \gamma}$ signals for healthy, PM1, PM2, PM3 conditions, at U_4	52
4.25 Healthy vs PM1/PM2/PM3 - $a_{MB, (front), \gamma}$, U_4 : histogram data fitting of the simulated dataset with respect to theoretical normal distribution.	52
4.26 Healthy vs PM1/PM2/PM3 - $a_{MB, (front), \gamma}$, U_4 : NPPs for quantile comparison of simulated data with theoretical normal distribution.	53
4.27 Bar plots for (CI_{CF}) , $(CI_{Kurtosis})$ and (CI_{RMS}) for healthy vs PM1/PM2/PM3 faulty conditions	53
5.1 (A) 1P frequency amplitudes for CMS signals, (B) 3P frequency amplitudes for CMS signals	56
5.2 1P and 3P amplitudes for PM3 at U_4 : relative percentage difference with respect to the one seed input turbulent conditions: (A) Main bearing velocities side- side direction, front and rear (B) Gearbox housing and generator velocities for side-side direction,	56
5.3 FFT plot analysis of $v_{GB, \gamma}$ signal resampled with COT algorithm for different wind speeds: (A) MI1, (B) PM1, (C) YM1	57
5.4 FFT plot Healthy vs YM1/YM2/YM3 fault conditions for $v_{GB, \gamma}$ signal in order domain	57
5.5 Averaged peak amplitudes comparison of $v_{GB, \gamma}$ signal under varying health states and turbulent inflow conditions. (A) MI - 1P, (B) PM - 1P, and (C) YM - 3P	58
5.6 Averaged 1P peak amplitudes comparison of ω_{rotor} signal: (A) MI, (B) PM	59
5.7 MI - 1P: Normalised relative change of the averaged amplitudes, comparing turbulent and steady-state conditions through SS error bars	59
5.8 PM - 1P: Normalised relative change of the averaged amplitudes, comparing turbulent and steady-state conditions through SS error bars	60
5.9 YM - 3P: Normalised relative change of the averaged amplitudes, comparing turbulent and steady-state conditions through SS error bars	60
5.10 Correlation coefficients values for mass imbalance scenario signal pairs comparing steady-state and turbulent input conditions	61
5.11 Correlation coefficients values for pitch misalignment scenario signal pairs comparing steady-state and turbulent input conditions	62
5.12 Correlation coefficients values for yaw misalignment scenario signal pairs comparing steady-state and turbulent input conditions	62

5.13 MI: 1P amplitudes distribution and variance analysis aggregated for all wind speeds . . .	63
5.14 PM: 1P amplitudes distribution and variance analysis aggregated for all wind speeds . . .	63
5.15 YM: 3P amplitudes distribution and variance analysis aggregated for all wind speeds . . .	64
5.16 Healthy condition vs MI1/MI2/MI3 - ω_{rotor} amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test)	64
5.17 Healthy condition vs MI1/MI2/MI3 and PM1/PM2/PM3 - $v_{GB, Y}$ amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test) - wind speeds: U_1, U_3, U_4, U_6	65
5.18 Healthy condition vs YM1/YM2/YM3 - $v_{GB, Y}$ amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test), wind speeds: U_1, U_3, U_4, U_6	65
5.19 PDF for $a_{MB(front), Y}$ signal comparing healthy conditions with multiple pitch misalignment degrees ($U_4 = 12$ m/s)	67
5.20 Time domain torsional mode comparison of PM3 signal in steady-state compared to turbulent inflow conditions	67
5.21 Histogram data fitting of the simulated dataset with respect to theoretical normal distribution. Study case: pitch misalignment imbalance for front main bearing acceleration signal	67
5.22 NPP for quantile comparison of simulated data with theoretical normal distribution. Healthy conditions for wind speed $U_4 = 12$ m/s compared with multiple degrees of pitch misalignment (PM1, PM2, PM3), for $v_{MB(front), Y}$ signal	68
6.1 Detection and diagnosis decision tree for three different imbalance scenarios under study (MI, PM, YM)	71
6.2 ω_{rotor} amplitude distribution median values comparison after significant statistic difference detection	73
6.3 ω_{rotor} amplitude distribution median values magnitude for different health states: Healthy vs MI1/MI2/MI3	73
6.4 Mass Imbalance fault severity assessment tree applied to ω_{rotor} signal	74
6.5 1P - Amplitude distribution median values magnitude for different health states: Healthy vs PM1/PM2/PM3: $v_{MB(front), Y}$, $v_{GB, Y}$ and $v_{Gen, Y}$ signals	74
6.6 Pitch misalignment fault severity assessment tree	75
6.7 3P - Amplitude distribution median values magnitude for different health states: Healthy vs YM1/YM2/YM3: $v_{GB, Y}$ signal: (A) U_3 ; (B) U_6	76
6.8 Yaw misalignment fault severity assessment tree through $Mdn(3P)$ values and for all CMS signals	76
B.1 Flow-diagram for COT function code (B.1)	86
C.1 MI - 1P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	87
C.2 PM - 1P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	88
C.3 YM - 3P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	88
C.4 MI - 1P: relative change averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $v_{Gen, Y}$; (C) $v_{MB (front), Y}$; (D) $v_{MB (rear), Y}$	89
C.5 PM - 1P: relative change averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $v_{Gen, Y}$; (C) $v_{MB (front), Y}$; (D) $v_{MB (rear), Y}$	89
C.6 YM - 3P: relative change averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	90
C.7 Front main bearing boxplot	90
C.8 Front main bearing multiple comparison test	90
C.9 Front main bearing boxplot	91
C.10 Front main bearing multiple comparison test	91

C.11	Generator signal boxplot	91
C.12	Generator signal multiple comparison test	91
C.13	PDFs of $a_{MB (front), Y}$ signals for healthy, MI1, MI2, MI3 conditions, at U_4	91
C.14	Bar plots for (CI_{CF}) , $(CI_{Kurtosis})$ and (CI_{RMS}) for healthy vs MI1/MI2/MI3 faulty conditions	91
C.15	Healthy vs MI1/MI2/MI3 - $a_{MB (front), Y}$, U_4 : histogram data fitting of the simulated dataset with respect to theoretical normal distribution.	92
C.16	Healthy vs MI1/MI2/MI3 - $a_{MB (front), Y}$, U_4 : NPPs for quantile comparison of simulated data with theoretical normal distribution.	92
C.17	PDFs of $a_{MB (front), Y}$ signals for healthy, MI1, MI2, MI3 conditions, at U_4	92
C.18	Bar plots for (CI_{CF}) , $(CI_{Kurtosis})$ and (CI_{RMS}) for healthy vs YM1/YM2/YM3 faulty conditions	92
C.19	Healthy vs YM1/YM2/YM3 - $a_{MB (front), Y}$, U_4 : histogram data fitting of the simulated dataset with respect to theoretical normal distribution.	93
C.20	Healthy vs YM1/YM2/YM3 - $a_{MB (front), Y}$, U_4 : NPPs for quantile comparison of simulated data with theoretical normal distribution.	93
D.1	Mass Imbalance 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds	94
D.2	Pitch misalignment 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds	95
D.3	Yaw misalignment 3P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds	95
D.4	Mass Imbalance 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds	96
D.5	Pitch misalignment 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds	96
D.6	Yaw misalignment 3P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds	97
E.1	MI - 1P: Averaged peak amplitudes comparison comparison under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	98
E.2	PM - 1P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	99
E.3	YM - 3P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	100
E.4	MI - 1P: Relative change in averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $v_{Gen, Y}$; (C) $v_{MB (front), Y}$; (D) $v_{MB (rear), Y}$	101
E.5	PM - 1P: Relative change in averaged frequencies amplitudes for PM1/PM2/PM3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $v_{Gen, Y}$; (C) $v_{MB (front), Y}$; (D) $v_{MB (rear), Y}$	101
E.6	YM - 3P: Relative change in averaged frequencies amplitudes for YM1/YM2/YM3 compared and normalized with respect to healthy conditions. (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$	102
E.7	Mass imbalance - 1P amplitudes scatter plots of $v_{GB, Y}$, Y , and ω_{rotor} signals. Each subplot resembles a single wind speed for different health conditions: (a) U1; (b) U3; (c) U4; (d) U6	102
E.8	Pitch misalignment- 1P amplitudes scatter plots of $v_{GB, Y}$, and $v_{Gen, Y}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U1; (b) U3; (c) U4; (d) U6	103
E.9	Yaw misalignment- 1P amplitudes scatter plots of $v_{GB, Y}$, and $v_{Gen, Y}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U1; (b) U3; (c) U4; (d) U6	103

List of Tables

3.1	Drivetrain system modes	23
3.2	OpenFAST output parameter for SIMPACK decoupled drivetrain analysis	25
3.3	Chosen fault cases and their severity levels, inspired by [23]	25
3.4	Wind speeds simulation setup	26
3.5	Color codes and nomenclature adopted to represent the different imbalance cases (Chapters 4 and 5)	26
3.6	Turbulence intensity level for the chosen wind speeds	28
3.7	Vibration transducer and their related required frequency range, reproduced from Tavner et al. (2020) [53]	34
3.8	Signals selected for the analysis in the frequency domain and their nomenclature	35
3.9	Signals chosen for time domain analysis and their nomenclature	35
4.1	Signals relative change ranges with respect to healthy conditions for different imbalance scenarios and fault severities	41
4.2	Time domain condition indicators for $v_{MB(front),Y}$ comparing Healthy vs PM1/PM2/PM3 cases	54
5.1	Time domain analysis in turbulent conditions. Condition indicator results are compared between healthy conditions and three pitch misalignment degrees PM1/PM2/PM3 for $v_{MB(front),Y}$ signal	68
A.1	10 MW DTU reference turbine specifics [37]	84
A.2	Drivetrain properties proposed by DTU[36]	84

Nomenclature

Abbreviations

Abbreviation	Definition
APDF	Amplitude probability distribution function
CF	Crest factor
CMS	Condition monitoring system
DND	Deviation normal distribution
GB	Gearbox
Gen	Generator
HSS	High-speed shaft
LCOE	Levelized cost of electricity
LSS	Low speed shaft
MB	Main Bearing
MBS	Multi-body simulation
MI	Mass imbalance
O&M	Operation and maintenance
PM	Pitch misalignment
SCADA	Supervisory control analysis data
YM	Yaw misalignment
WECS	Wind energy conversion system
WT	Wind turbine

Symbols

Symbol	Definition	Unit
v	Velocity	$[m/s]$
a	Acceleration	$[m/s^2]$
$\widetilde{1P}$	1P frequency peak amplitude	$[m/s]/[m/s^2]$
$\widetilde{3P}$	3P frequency peak amplitude	$[m/s]/[m/s^2]$

1

Introduction

The World Energy Transitions Outlook 2023 by the International Renewable Energy Agency (IRENA) [1], presents a vision for transforming the global energy landscape in line with the goals of the Paris Agreement. It outlines a pathway to limit global temperature rise to 1.5°C and achieve net-zero CO2 emissions by mid-century. This pathway necessitates a comprehensive transformation of energy consumption and production. In 2022, approximately 300 GW of renewable energy was added worldwide, constituting 83% of the new capacity, while fossil fuels and nuclear accounted for only 17%. By 2050, the share of renewables in the global energy mix is expected to rise, from 16% in 2020, to 77%. This scenario requires a twelve-fold increase in renewable electricity capacity by 2050, driven by high electrification rates in transport and buildings and annual capacity additions need to average 1,066 GW from 2023 to 2050 to stay on track for the 1.5°C target [1]. According to the statistics, the power sector in particular, was responsible for 40% of global CO2 emissions in 2022 and, following the expected projection. It will experience substantial growth in solar PV and wind power share in the coming decades, solidifying their role as pillars of the global energy transition [1].

Among the various renewable energy technologies, wind energy has shown significant growth and potential. The Global Wind Report 2023 by the Global Wind Energy Council [2], revealed that in 2022, 77.6 GW of new wind power capacity was added to power grids worldwide, increasing the total installed wind capacity to 906 GW, marking a year-on-year growth of 9%. Additionally, the report anticipates that 680 GW of wind capacity has to be added globally between 2023 and 2027, of which 130 GW offshore, with an average annual installation rate of nearly 26 GW. The two projected percentage shares between onshore and offshore are shown in Figure 1.1. The forecast predicts that the total installed wind capacity will reach 2 TW by 2030.

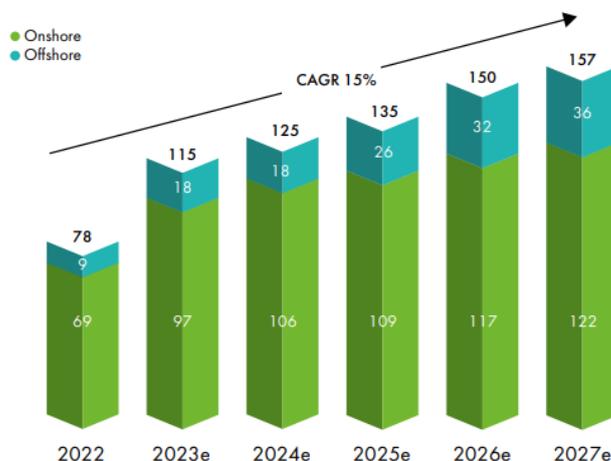


Figure 1.1: Wind energy installation outlook 2022-2027 (GW) [2]

In the European Union, it's been determined that to meet the goals of the "REPowerEU" plan—focused on diversifying energy sources with more clean energy—there is an urgent need to add 30 GW of wind energy capacity each year by 2030. However, progress is falling short, with only 16 GW installed last year. The wind sector faces challenges such as rising costs due to inflation and slow permitting processes, which have delayed around 80 GW of projects [2]. The EU aims to streamline these processes by prioritizing renewable projects and passing new legislation like the "Net-Zero Industry Act" which focuses on simplifying permitting procedures for new factories to strengthen the supply chain. Ultimately, the EU should create favorable market conditions and invest in industrial growth to secure large-scale wind deployment and maintain global competitiveness in renewable energy. Wind turbine system reliability is a critical factor in the success of a wind energy project.

This growth in wind energy capacity came alongside an increase in wind turbine's rated capacity. This upscaling over the last few years has not come without any challenges. From a design perspective, the increase in size and complexity of wind turbines has introduced numerous technical difficulties. Among the most critical components affected by these changes, we can find the blades. These larger structures are prone to a variety of reliability issues, which can significantly impact the overall performance and lifespan of the turbines. Ensuring the structural integrity and durability of the blades has become a paramount concern in the development of modern wind turbines. It has been noticed and analyzed by many companies like DNV [3], that due to both their increase in size and operation in more harsh environmental conditions, like the one experienced by offshore wind farms, wind turbine's blade faults is arising as one of the main contributors in the turbine's operation and maintenance (O&M) costs. Furthermore, those faults are responsible for causing imbalance issues in the turbines' rotor affecting the reliability of the whole system. Poor reliability directly affects both the wind project's revenue stream through increased operation and maintenance (O&M) costs and reduced availability to generate power due to turbine downtime.

Therefore, it is of major importance when it comes to O&M, to be equipped with a robust condition monitoring system (CMS), to detect, isolate, estimate, and perform prognoses on component degradation. It is a challenging procedure that is becoming essential to reduce unplanned maintenance and downtime through predictive condition-based maintenance which can avoid severe failures.

In addition to optimizing costs and enhancing technical performance, addressing a crucial social aspect of wind turbine operation and maintenance (O&M) is imperative: safeguarding worker safety by refining maintenance strategies and reducing unplanned interventions caused by critical failures. Wind turbine repairs require highly skilled technicians who are exposed to various hazards, including working at significant heights, performing suspended tasks, handling electrical systems, and operating in confined spaces. These risks are intensified by challenging environmental and weather conditions, particularly in offshore operations where personnel transfers pose additional constraints.

A key example of these challenges is found in nacelle maintenance, which involves specific dangers such as injury from moving parts if the nacelle rotates unexpectedly, burns from hot surfaces, and proximity to high-voltage cables. Accessing the nacelle often necessitates climbing long vertical ladders, especially when lifts are unavailable, placing considerable physical demands on technicians. These occupational hazards are highlighted by the European Agency for Safety and Health at Work (EU-OSHA) [4], and a related exploratory study by Cunha et al. [5] further explores both the physical and psychological risks faced by workers in clean energy sectors. This study emphasizes the need to address the unique safety challenges inherent to "green jobs," ensuring comprehensive protection for technicians in the wind energy industry.

1.1. Research Objective and Questions

Fault detection and diagnosis in industrial processes is a highly active research area, driven by the potential to significantly reduce maintenance costs and improve system reliability through the implementation of advanced monitoring techniques and algorithms. In the context of mechanical faults, specific fault signatures can be observed in vibration signals, either in the time or frequency domain. This research targets the application of these techniques to wind turbine systems by analyzing simulated drivetrain condition monitoring system (CMS) vibration signals, combined with SCADA signal measurements, to detect rotor imbalance scenarios. Rotor imbalances are well-known fault conditions that can significantly compromise the operational efficiency, safety, and reliability of power generation. Imbal-

ances induce excessive vibrations and loading, leading to accelerated wear, increased maintenance costs, and potential system failures. In wind turbines, they can damage drivetrain components and undermine structural integrity, resulting in costly repairs, operational downtime, and increased risk for technicians working in potentially hazardous environments. Early detection of rotor imbalances is thus critical, as it helps mitigate these risks and contributes to a reduction in the levelized cost of electricity (LCOE) and the operation and maintenance (O&M) costs of wind turbines, ultimately supporting the competitiveness of wind power as a renewable energy source.

Rotor imbalances in the turbine's blades can be categorized into two types: mass imbalance and aerodynamic imbalance [6]. Mass imbalance occurs when the rotor's weight distribution is unevenly spread around its center of gravity, often due to causes such as ice accumulation or blade damage. This imbalance induces irregular centrifugal and inertial forces during rotation, which lead to increased vibrations and oscillations. Aerodynamic imbalance, on the other hand, arises from deviations in the aerodynamic properties of the rotor blades, such as those caused by pitch angle misalignment or yaw misalignment of the turbine with respect to the wind direction.

As a result, to enable early and reliable detection of imbalance conditions, it is crucial to extract interpretable features from CMS signal analysis. These features are key to developing an accurate decision tree algorithm, which is essential for effective fault detection and diagnosis. This approach helps prevent issues such as reduced component lifetime, increased safety risks, and reduction in energy yield.

Therefore, this thesis poses the following primary research question:

"Can rotor imbalance be efficiently detected and diagnosed using existing condition monitoring sensors on the wind turbine drivetrain to prevent severe component failures and ensure reliable power production?"

The presented research question can be further divided into more subquestions that will be addressed through this thesis report:

1. Which signal analysis techniques are most suitable for detecting the presence of rotor imbalances? Which drivetrain signals are the most effective and reliable for their detection and diagnosis?
2. Which signal features ensure robust fault detection?
3. Can a decision tree algorithm be developed for fault detection and classification?
4. Can health indicators be constructed for the specific imbalance scenarios under investigation? To what extent can they identify fault severity levels?

1.2. Report outline

The report is structured as follows: Chapter 2 provides a theoretical background essential for this research, covering O&M strategies, wind turbine monitoring systems, drivetrain components, rotor imbalances, blade faults, and key signal processing techniques. Chapter 3 outlines the simulation framework and presents a detailed explanation of the methodology and analysis procedures employed. Chapter 4 presents the results for steady-state input conditions, while Chapter 5 addresses results under turbulent conditions. Chapter 6 includes an in-depth discussion of the results introducing a decision tree framework for fault detection and diagnosis and health assessment. The study's limitations and recommendations are also explored. Finally, conclusions are provided in Chapter 7 that summarizes the key findings and underlines areas for future research and improvements.

2

State of the Art

This chapter provides an overview of the current state-of-the-art relevant to this research and its associated framework. Section 2.1 focuses on existing wind turbine O&M strategies, while Section 2.2 outlines the various monitoring systems employed in wind turbines and the corresponding detection techniques. Section 2.3 presents a summary of the key components constituting the wind turbine drivetrain, offering a foundation for understanding the drivetrain model used in the simulations discussed in Chapter 3. Additionally, Section 2.4 briefly introduces the most common types of damage observed in wind turbine blades. Section 2.5 delves into rotor imbalance cases, providing an overview of ongoing research in the field, followed by a detailed discussion of their structural dynamics effects in Section 2.5.1. Section 2.6 introduces the fundamentals of signal processing, with particular focus on the frequency domain (Section 2.7 and its subsections) and time-domain analysis (Section 2.8). Finally, Section 2.9 summarizes various system data fusion approaches and their classifications.

2.1. An Overview of O&M

As wind power advances toward becoming a major utility source, reducing the LCOE is crucial for making wind power competitive with conventional energy sources. A significant challenge with wind power systems is the high cost of O&M, especially for offshore wind farms, which leads to power loss due to the extended downtime of turbines. The O&M cost for offshore wind turbines, in particular, is estimated to be around 20-25 % of the total income, up to three to five times greater than onshore due to the difficult-to-access locations, the high cost of the specialist personnel and access equipment needed [7, 8], therefore robust condition monitoring systems (CMS) and fault diagnosis strategies of wind turbines represent a high-level priority to reduce those high costs. The monitoring processes can be carried out in two ways: online, which offers immediate feedback on the equipment's status, or offline, where data is gathered at regular intervals using separate measurement systems not integrated with the equipment, it comprehends the whole system and equipment of sensors for data acquisition and processing. The system provides continuous indications of components (and hence wind turbine) conditions based on techniques that can include vibration analysis, acoustics, oil analysis, strain measurement, and thermography. Wind turbines (WTs) in particular, utilize monitoring systems to evaluate the condition of essential components such as blades, gearboxes, generators, main bearings, and towers. It has been observed that large fault types such as gearboxes, generators, and blade systems are the ones responsible for the greatest downtime together with complex and costly repair procedures [8].

Figure 2.1 shows that after extracting and analyzing the monitoring system's signal data outputs, two main categories of maintenance strategies can be recognized in the literature [9]: *Corrective Maintenance and Preventive Maintenance*. The former is mainly adopted onshore and is being performed when a fault already occurred and has been located. When the fault is detected two main approaches can be applied depending on the severity and the importance of the component, namely, *Palliative Maintenance* for provisional rehabilitation or *Curative Maintenance* for permanent rehabilitation. On the other hand, preventive maintenance instead is planned to prevent the actual failure from occurring in advance. Different strategy cases can be identified like time-based, which means that regular main-

tenance visits are planned or parameter prevision-based interventions that are scheduled depending on the relative component lifetime information rather than present operations. They can be recognized under the names of *Systematic Maintenance* and *Forecast Maintenance*. Secondly, *Conditional Maintenance* (CBM), depends on a CMS that involves the acquisition, processing, analysis, and interpretation of data and the selection of optimal maintenance actions after detecting the present health conditions of the components. Lastly, *Proactive Maintenance* can be applied which provides condition-based intervention.

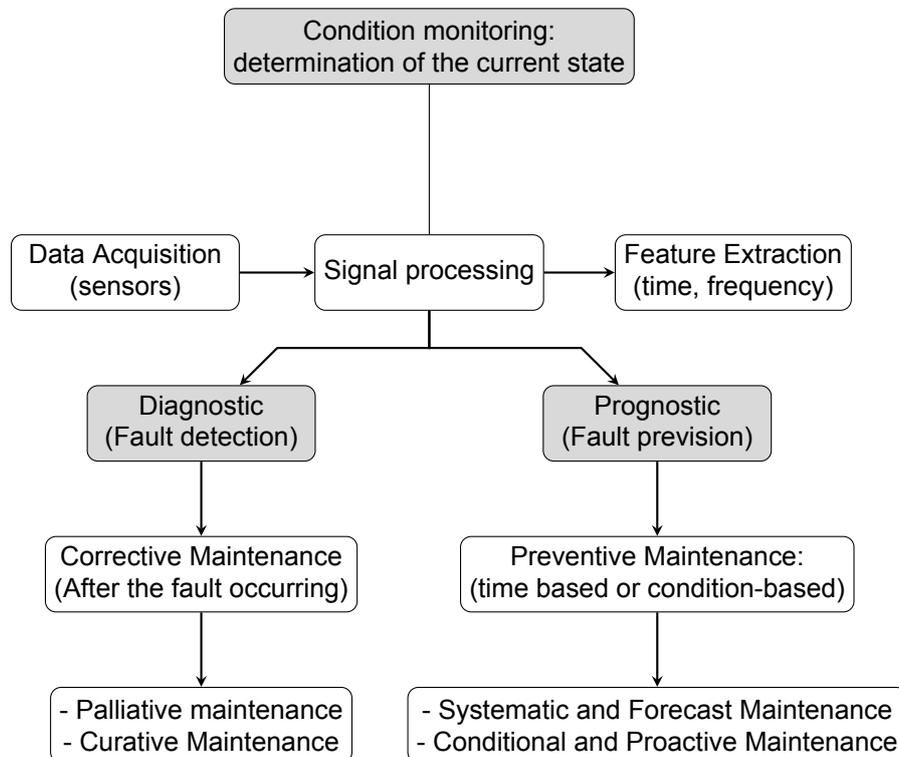


Figure 2.1: Condition monitoring categorization overview, inspired by Tchakoua et Al. [9]

2.2. Wind Turbines Monitoring: Systems and Techniques

As introduced in the previous section, many studies report that 25% up to 35% (offshore) of the LCOE is represented by O&M of wind turbines during their lifetime [8]. Therefore modern WTs are outfitted with systems for active remote monitoring and control of their components in real-time. These systems provide valuable data that can be used to predict, detect, and diagnose faults as they arise.

Wind turbine components can be classified according to multiple criteria and different main subsystems can be identified. First of all, we can find the rotor system, which includes blades and the hub, and then there is the drivetrain which includes the power transmission systems with low-speed shaft (LSS), bearings, couplings, gearbox, high-speed shaft (HSS), and brakes. The electrical setup includes the generator and the power electronics and the control system includes the pitch motor and gears, they allow blade positioning to maximize the energy capture for lower wind speeds or limit it for higher ranges. The yaw system tracks the incoming wind direction and includes gears, brakes, and sensors. Finally, the support structures are represented by the nacelle and the tower. Figure 2.2 gives an overview of the mentioned subsystem components.

CMSs are positioned in the drivetrain of the wind turbines and provide high-resolution monitoring for diagnosis and prognosis of faults before catastrophic damage occurs. They help to increase the reliability of the WT system giving the possibility to schedule maintenance sessions in advance and reducing downtime and lost revenue. It can be both online providing instantaneous feedback information or offline in which data are collected at regular time intervals using measurement systems. Multiple CMS

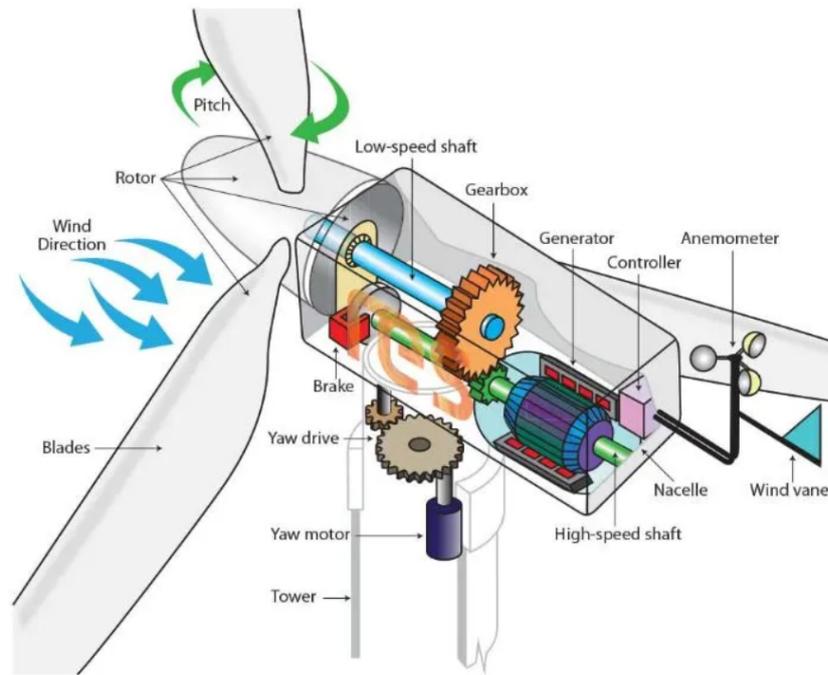


Figure 2.2: Wind turbine components (source: <https://windmillstech.com/wind-turbine-components/>)

concepts are currently available and reported below [10, 11].

- The first most common low-resolution system is represented by *Supervisory Control and Data Acquisition (SCADA)* whose output data are typically sampled at 1 Hz frequency and averaged every ten minutes. These signals are collected using sensors fitted on the WT and can include data from environmental parameters, electrical outputs, control variables like shaft speeds and pitch/yaw system angles data, or thermal data linked to machinery health state. SCADA provides a valuable low-cost monitoring system usually connected to centralized monitoring centers operated by turbine manufacturers or service providers. The large volume of generated SCADA data needs a deep analysis to interpret them due to their low acquisition rate, machine learning techniques are often employed for this purpose. One classification of machine learning techniques for wind turbine condition monitoring is to divide them into supervised and unsupervised learning. A possible categorization of the different SCADA-based monitoring methods is with the following five classes [12]:
 - i. trending, which uses statistical methods for setting alarming thresholds
 - ii. clustering, which is suitable for a wind farm implementation in order to automatically detect and distinguish between 'healthy' and 'faulty' conditions,
 - iii. normal behavior modeling (NBM), which involves the use of regression models that fit measured data and predict future value behavior,
 - iv. Damage modeling uses physical failure modes of interest to estimate failure probability, and finally
 - v. Assessment of alarms and expert systems which involves machine learning and AI knowledge-based systems.
- **Vibration Analysis:** it is the most important method used for WTs with a 58% share of the total CM market [10], in particular for high frequency rotating machinery components, i.e. the drive-train. Multiple sensor types are available for this purpose and the type of sensors used essentially depends on the frequency range of interest, the stiffer the sensor mounting position the higher its accuracy [13]. The most widely spread sensor types are accelerometer, velocity, and displacement sensors. Accelerometers can be further divided into piezoelectric and MEMS sensors, the

former relies on the piezoelectric effect of crystals to generate electrical output proportional to the applied acceleration, and the latter relies on the inertia of a movable proof mass supported to the frame by a mechanical suspension system. Velocity transducers are based on the electromagnetic induction effect which results in a voltage production proportional to the velocity of the vibration that causes the magnet movement. Displacement sensors use non-contact methods like eddy currents or capacitive sensing. Alternative concepts are also available like LDV (non-contact optical sensors) which use the laser Doppler phenomena.

Vibration techniques were initially adopted for WT CM. The underlying principle relies on two fundamental observations [14, 15]:

- i. Each drivetrain component possesses a characteristic vibration frequency, which maintains a steady amplitude under normal operating conditions, albeit varying with drivetrain speed.
- ii. The vibration profile alters when a component begins to deteriorate, with the nature of these changes being dependent on the specific failure mode.

Output signal analysis needs specific knowledge which can be divided into two main groups of processing categories: time domain and frequency domain. Time domain analysis relies on statistical parameter detection of the signal like peak value, root-mean-square, crest factor, and kurtosis. On the other hand, frequency domain analysis relies on techniques like the Fast Fourier Transform (FFT) algorithm and vibration amplitude with power spectrum (PSD) to isolate the frequencies of interest. These analyses will be the main focus of this research work, both the mentioned techniques will be explored and further explained in more detail in sections 2.7 and 2.8.

- **Oil Analysis:** Mechanical components like gears in the gearbox are subject to wear due to several factors like temperature, contact pressure between the teeth, and high operating speeds. In order to reduce this effect lubricants are used during operations and their chemical composition quality monitoring can help in detecting imminent failures by performing a particle contamination analysis and providing early warnings [11, 16].
- **Acoustic Emissions & Ultrasonic Analysis** Strong non-stationary signals like sudden crack propagation cause a rapid release of strain energy and due to the altered metal structure, elastic waves are generated. This method has been proven to be successful in blade crack detection and monitoring of bearing and gearbox together with vibration-based analysis. Although their application results to be quite expensive it is gaining more attention in the field of condition monitoring for its ability to detect faults also in slowly rotating machines. Additionally, the ultrasonic emission method is also known as an extension of the acoustic emissions analysis for structural evaluations, with the advantage that ultrasonic waves experience less noise than acoustic ones. They both proved to be useful for inner structure damage investigation but still need complicated and time-consuming signal processing techniques [11, 16].
- **Electrical Parameters Analysis:** For proper condition monitoring planning also the electrical system can be monitored through online and offline techniques but this is rarely implemented. Methods like the Motor Current Signal Analysis rely on high-frequency acquisition of voltage and current signals. These techniques are at the moment confined to research-related activities but there is significant potential for applying them successfully in the field [11].
- **Strain Measurements:** the implementation of this technique is highly spread in the rotor blade monitoring field, it can be performed with the use of strain gauges (more common) or optical fiber sensors which are still too expensive. They detect if the blade structure deforms under the applied loads, with accuracy dependent on the sensor mounting position. The resulting fault indicators will be peak strain values for damage location and strain rate analysis for failure and crack prognosis [11, 16].
- **Thermography:** this technique relies on the measurement of temperature gradient for damage detection but is not useful for early detection conditions due to its slow development and relation to failure severity level. The infrared radiations emitted by the objects are captured by a thermographic camera, and this process is suitable only for offline detections. It is often used for monitoring electronic components. The drawbacks of this method are the requirement of thermal image processing which can be costly and can also be affected by external conditions [11, 16].

2.2.1. Vibration-based CMSs

Given its widely spread use in wind turbines, it is of interest to mention that in such systems the most important variations concern the sensor's quantity and positioning and the analytical algorithms employed. Usually, all CMS systems rely on accelerometers as core measurement instruments. The typical sensor layout along the wind turbine drivetrain is depicted in figure 2.3 showcasing the architecture of the Gram & Juhl (an industry leader in wind turbine condition monitoring solutions) turbine CMS [17]. Sensors are mounted on the drivetrain side of the wind turbine energy conversion system to detect characteristic vibration signatures for each component. The installation of vibration sensors for CMS is described in ISO 10816-2 [18]. The standard recommends the placement of piezo-resistive or capacitive accelerometers on the housing of the main bearings, the gearbox, and the generator, in both radial and axial directions. The vibration signals shall be analyzed in the frequency band of 10 Hz to 5 kHz. The signature for each gear mesh or bearing is unique and depends on the geometry, load, and speed of the involved components. Insights about the drivetrain state of the art and related components are given in Section 2.3.

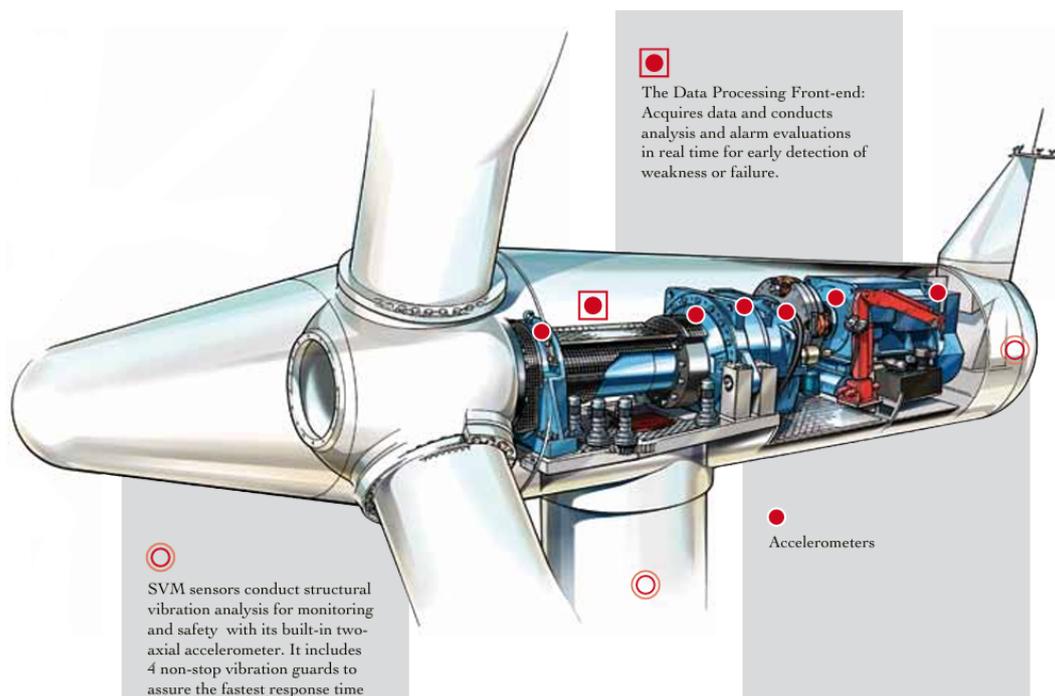


Figure 2.3: Sensor position in drive train vibration-based condition monitoring system [17]

2.3. Drivetrain Systems State of the Art

This section gives an overview of drivetrains and their general characteristics. However, it should be noted, that is not the purpose of this research to dive into the details of their respective technical aspects and more information can be found in the literature.

Drivetrains are the systems that convert the wind's kinetic energy to electrical energy, including in this context the whole power conversion components: main bearing, shafts, gearbox, generator, and power converter. Various wind turbine drivetrain technologies exist, each with distinct advantages and disadvantages concerning factors such as cost, weight, dimensions, manufacturing processes, material usage, efficiency, reliability, and operational and maintenance requirements [12].

The two most common drivetrain systems layouts are represented in Figure 2.4 which include two different concepts: geared drive and direct drive systems. Drivetrain design trends do not rely, like other components, on the wind and site conditions but mainly on costs, logistics and service, and operational monitoring [12]. Geared turbines account for the use of the gearbox, which increases the low speed of the wind-driven rotor several hundred-fold considerably reducing the size of the required generator. This category usually involves induction generators or synchronous generators. However, the presence

of a gearbox can impact the reliability of the wind turbine increasing the percentage of outage time. The second most common option, especially offshore, is the direct drive system which requires the presence of a bigger generator to produce a higher torque. It operates at the same speed as the turbine's rotor with the advantage of eliminating gearbox failure and transmission losses. This technology comes with disadvantages like heavy mass due to the large size and diameter and higher cost compared to geared drive systems generators. The direct drive layout usually involves two systems layout: the permanent magnet direct drive and the electrically excited direct drive.

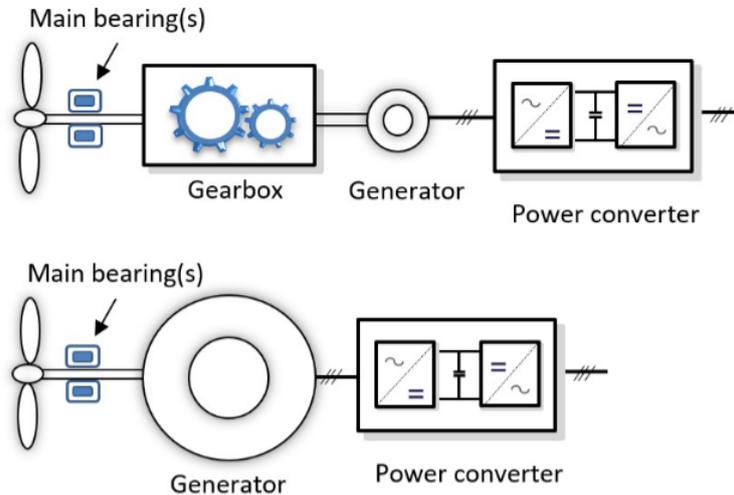


Figure 2.4: Schematic layout of the two most common drivetrain configurations (a) with and (b) without gearbox (illustration by Amir Nejad [12])

The loads and operational conditions the drivetrains must endure are determined by the design load cases specified in the International Electrotechnical Commission (IEC) standards: IEC 61400-1 for land-based wind turbines, and IEC 61400-3-1 and IEC 61400-3-2 for offshore fixed and floating wind application, respectively.

Gaining a more detailed understanding of the specific components that constitute the drivetrain [12], it is known that *main bearings* are one of the components with higher cost implications per fault's severity, existing commercially available designs currently in use are rolling element bearings. These components undergo repeating large-scale fluctuations in loads, and depending on the degree of machinery integration (i.e. the bearing supporting both turbine rotor and generator like in direct drive structures) the severity of a possible failure increases. The main adopted design especially for higher-power wind turbines is a tapered roller main bearing, which has truncated cones roller structures that support both radial and axial loads. Two different drivetrain designs are possible depending on whether they involve a single or double main bearing corresponding to a three-point or four-point support structure. The former configuration will result in a lighter structure due to the lack of a second bearing which also results in a shorter low-speed shaft (LSS), the latter on the other hand, has been demonstrated to be less sensitive to loads as it has been analyzed by Guo et al. in [19].

Due to the increasing trend in wind turbine size and power rating, multiple *gearbox* designs with higher torque densities and speed ratios are currently present. Gearboxes are designed for a minimum of a 20-year life, as specified in the IEC 61400-4 design standards but, it has been observed that reliability during operations results to be lower than the one calculated in reference through the standards. Important is the use of lubricants for gears to reduce the level of wear arising during the operations. To achieve the large step-ups required by high-rated machines like WT (rotational speed must be increased by a factor close to 100), usually, three separate stages in the gearbox are required. The first one attached to the LSS is usually planetary to achieve a sufficiently high ratio and spread the high torque load on multiple contacts due to its higher density [12]. This stage comprises three planet gears in a planet carrier coaxial with a sun gear and a ring gear with the planet gear rotating at the constant centers of the planet carrier. On the other hand, two parallel stages are present which use helical gears that can handle lower torque densities and are implicated both in the intermediate speed shaft (ISS)

and high-speed shaft (HSS) which is then coupled to a generator drive end.

Regarding *generator systems*, direct-drive systems, especially those utilizing rare-earth permanent magnet synchronous generators (PMSGs), are favored for offshore applications due to their high efficiency and reliability. In contrast, geared systems are additionally divided into doubly-fed induction generators (DFIGs) with partial-power converters and brushless generators with full-power converters (GFPCs). DFIG systems are commonly used in medium-sized turbines ranging from 3–6 MW. Both squirrel cage induction generators and PMSGs are employed by GFPC systems. Induction generators generate rotor magnetic fields via rotor currents, resulting in joule rotor losses and decreased efficiency. PMSGs, using rare-earth permanent magnets, avoid these additional losses by directly producing the rotor magnetic field. DFIG systems use carbon brushes and slip rings to conduct currents between the rotor and stator. These components are prone to wear and require frequent maintenance. PMSGs, however, eliminate the need for these elements, leading to lower maintenance demands. Studies have shown that PMSGs have a 40% lower failure rate than comparable DFIGs during the early operational stages. Concerns over the availability of rare-earth elements, which are crucial for PMSGs, have driven the wind industry to explore innovative technologies [12].

Positioned between the generator and the power grid, the *power converter* must meet the operational requirements of both interfaces. The initial common layout for smaller power ratings was constituted by a diode rectifier with a DC/DC boost converter and a two-level voltage source converter. Then it was improved in a back-to-back converter structure which instead of the first two stages listed components can be implemented with multiple parallel two-level voltage source inverters for higher power ratings. Attention also needs to be paid to thermal loading with adequate cooling systems and grid integration

2.4. Damage of Wind Turbine Blades

The increasing trend of global wind energy installation led to a respective increasing trend in blade size, resulting in more challenging design requirements for the loads that the structure needs to withstand and aerodynamic performance. Blades have significant effects on the overall performance of the wind turbine and are costly to manufacture, reaching 25-35 % of the total costs and maintenance [20]. The main manufacturing materials of wind turbine blades are glass-fiber-reinforced polymer (GFRP) and carbon-fiber-reinforced polymer (CFRP). Figure 2.5 shows an overview of the terminology linked to all the blade's parts. The complete blade is formed by the bonding of two half shells along the trailing edge. The occurrence of damages or faults is often caused by extreme weather conditions in harsh environments like storms, lightning strikes, and sudden strong wind phenomena like gusts. DNV observed [3] that the blade damages caused by lightning phenomena originate from the increasing blade tip height and carbon fiber spar caps adoption in manufacturing processes, while leading-edge erosion (gradual wear and damage of the blade's leading edge caused by continuous exposure to operational stresses) phenomena arise in particular due to the increased tip speed of higher power-rated wind turbines. Location regions for most usual damage or failure along the blades are the edges, both near the root and near the tip side, the trailing edge on the high-pressure side, and the leading edge.

Lightning strikes will cause damages, especially at the outmost part of the turbine blade trailing edge, such as delamination, debonding, shell, and tip detachment. Delamination phenomena need particular attention during the manufacturing process since composite materials such as the one used for wind turbine blades are prone to this type of defects that can be responsible for triggering also local buckling modes, these combined effects can lead to the failure of the entire structure. Strong winds will cause blade breakage, cracks (longitudinal or transverse), and failure in the most severe cases, additionally, harsh atmospheric conditions are responsible for the erosion of the leading edge. Ice accumulation is also an issue of great concern because it will affect the rotor's aerodynamic performance, causing imbalanced rotation, aerodynamics loss, unwanted stop, and increased blade fatigue and safety risks.

In order to make sure that the blade is still in good condition and fulfills its functions, damage detection techniques can be adopted to monitor the damage before it causes catastrophic incidents. The main possible strategy is through fault indicators relying on the following methods: strain measurement through strain gauges, acoustic emission, ultrasound, vibration analysis, and thermography which have been introduced previously in section 2.2. Optical inspection on-site through machine vision for image acquisition is also possible. Nevertheless, it is not yet common nowadays to have sensors implemented

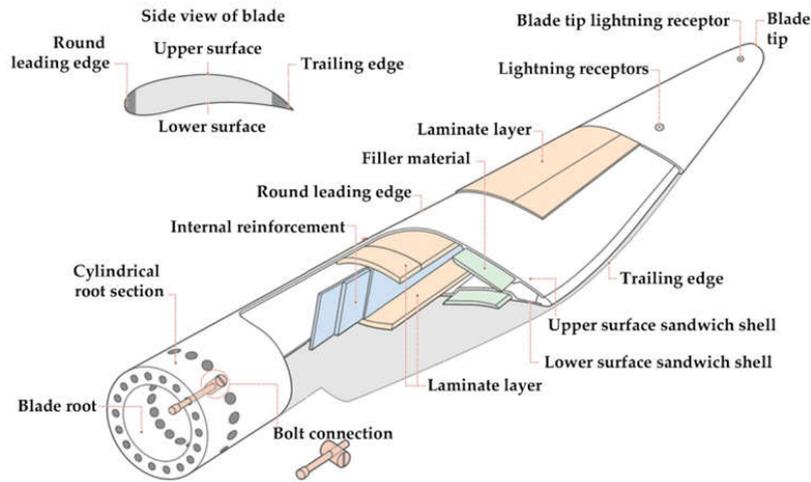


Figure 2.5: View of wind turbine blade's parts [21]

on the blade structures themselves, therefore the possible detection of faults is mainly linked to on-site inspections. Due to the importance of preventing severe damage and faults, it is of particular interest in the present research scenario to investigate new methods for remote detection.

2.5. Rotor Imbalances

Wind turbine blade faults are prone to induce rotor imbalance. In the literature, two main groups of rotor imbalance types can be found: mass and aerodynamic imbalances. Mass imbalance occurs when the mass of a rotating body is not evenly distributed around its center of gravity like in the presence of ice accretion or different blade damage types like the one mentioned in Section 2.4. The related uneven imbalance forces result in gravitational and centrifugal forces. The former will result in torque oscillations that excite torsional modes of the drivetrain, and the latter will result in periodic shear forces in the fixed frame that excite tower transverse bending modes. In the case of aerodynamic imbalance, like pitch misalignment, discrepancies arise in trust and tangential forces due to not optimized lift and drag forces along the blade. The fixed frame combination of shear and bending moments impacts structural and drivetrain response. On the other hand, a second aerodynamic imbalance is represented by yaw misalignment, which occurs when the wind turbine rotor is not correctly aligned with the wind direction for maximum power production, and its resulting load effect leads to shear force excitation affecting the structural response of the turbine. Traditionally, only SCADA signals are utilized. Meanwhile, the drivetrain condition monitoring signals are mainly used for fault detection in gears and bearings. Rotor imbalance detection mainly employs frequency-domain methods. It has been investigated and modeled in many studies that a successful method to do it implies the monitoring of the rotor 1P rotational frequency peak since imbalances in aerodynamic, gravitational, or inertial forces generate periodic loads at this frequency [22, 23].

This thesis aims to explore further cutting-edge fields of research, like the one proposed by Mehlan et al. [23], and Jin Xu et al. [24], both of which introduce a novel approach for investigating rotor imbalances through the use of drivetrain condition monitoring signals. Mehlan et al. analyze the resulting dynamic effects of three different imbalance scenarios by monitoring drivetrain vibration response and they propose a knowledge-based approach for the detection and diagnosis of all the mentioned imbalances for different fault cases. By looking at the 1P amplitudes, they point out that the most promising CMS signals are gearbox housing side-side velocity for pitch misalignment detection and nacelle acceleration for mass imbalance and yaw misalignment detection. Xu et al. instead, presented a successful vibration model for blade crack detection by applying the Morlet wavelet transform for rotational frequency detection of the rotor hub implementing also a health indicator for imbalance severity recognition. Niebsch et al. developed a method to estimate mass and aerodynamic imbalances by physical modeling of rotor dynamics and solving the inverse problem [6]. Although detection is easier to implement, more

research is needed for diagnostics, like classifying imbalance types and estimating fault severity. This research gap will be the target of the presented thesis work. Current literature methods can be broadly classified as physic-based, knowledge-based, or data-driven, and each of them presents different limitations. Physics-based methods require detailed aeroelastic models, but uncertainties arise from limited turbine specifications. Data-driven techniques show high accuracy but struggle with insufficient fault data, requiring expensive equipment like LIDAR to collect more data. Knowledge-based approaches depend on expert knowledge of fault causes and effects, but challenges exist in obtaining and managing this expertise. Each method has inherent limitations based on data availability, modeling accuracy, or domain knowledge [23].

2.5.1. Dynamic Effects of Rotor Imbalances

A short insight into the dynamic of a wind turbine is provided in this section to motivate the adopted analysis choices reported in Chapter 3 for the simulated signals. A schematic overview is also provided in figure 2.6. This Section is not meant to explain the full theory behind systems dynamics which, together with the mathematical background, can be found in the literature [[25], [26]].

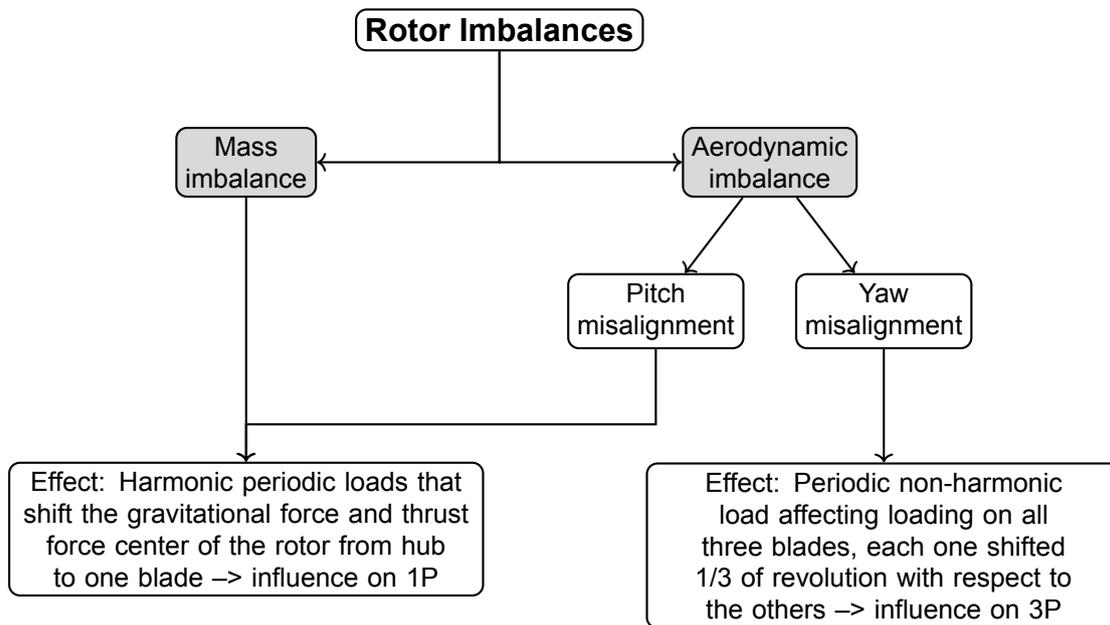


Figure 2.6: Rotor imbalances classification and effects

The loading conditions affecting wind turbines can be characterized according to their periodicity: harmonic, non-harmonic loading but periodic, and random loading. Harmonic loading in wind turbines occurs when variations follow a cosine or sine function, with three primary sources contributing to this effect. The first source is aerodynamic imbalance, which occurs when one blade's pitch angle differs from the others, shifting the center of thrust from the hub to the affected blade. The second source is mass imbalance, arising from uneven mass distribution among the blades, which shifts the rotor's center of gravity away from the hub. Both of these imbalances cause load variations and moments that align with the rotor's rotation period, known as 1P frequency. For offshore turbines, the third source of harmonic loading is small, regular waves, which can be approximated as sinusoidal functions, creating corresponding sinusoidal loads.

On the other hand, rotational sampling phenomena like wind shear, tower shadow, turbulence, and yaw misalignment produce non-harmonic periodic loadings that can be decomposed into different harmonic frequency contributions (1P, 2P, 3P...). Rotational sampling is the phenomenon of variation of the local wind conditions at the position of a blade element that is caused by its rotational motion. For three-bladed rotor turbines, rotational sampling induces load variations on the nacelle structure at a frequency of 3P and its multiples. This occurs because loading phenomena affecting all three blades are offset by one-third of a revolution relative to each other. Consequently, the total force on the hub, which is

the sum of all harmonics on the three blades, results in the cancellation of 1P and 2P signals due to their opposing phases. However, 3P signals are in phase and thus combine constructively, resulting in a non-zero force at the hub.

2.6. Signal Processing

As previously mentioned in Section 2.2, different approaches can be adopted to analyze the signals collected by the monitoring system depending on the type of fault being analyzed and the respective machinery component. Conventional vibration-based condition monitoring techniques rely on tracking key features within the measured signal. The two main and most common analysis techniques used to perform feature extraction of the CMS signals are represented by frequency domain and time domain analysis, a respective detailed overview is given in the following sections. It is important to notice that the majority of all vibration signal-processing techniques require steady-state loading conditions which are necessary to ensure measurement consistency. In most situations, harmonic frequencies necessitate stationarity in time to prevent frequency smearing, which can invalidate spectrum-based methods. However, wind turbines are inherently non-stationary machines due to the variable nature of wind, which dictates the rotor's rotation speed. This variation results in time-varying harmonic frequencies from components like gears, shafts, and bearings. Thus, accurate speed information is essential for numerous signal-processing techniques to handle these fluctuations effectively, necessitating the use of resampling algorithms. Installing an angle encoder or tachometer on one of the rotating shafts in the gearbox is a common, accurate, and reliable method to obtain this speed data. Alternatively, the instantaneous angular speed can be estimated directly from the vibration signal itself. Related to this concept, Computed order tracking (COT) algorithms are explored. COT is primarily used for frequency domain analysis, it tracks specific orders or harmonics over time in the frequency domain, providing insight into the behavior of rotating components.

2.7. Frequency Domain Analysis

When it comes to signal analysis, especially vibration signals, the frequency spectrum can represent a powerful tool to detect specific component contributions and the effect of a fault compared to healthy conditions. The most widely adopted technique to perform it is through the Fast Fourier Transform (FFT) which is an efficient algorithm for calculating the Discrete Fourier Transform (DFT) of a digital discrete-time signal. In this section, just a general overview of the main and relevant concepts for this thesis work is provided according to Barszcz [10]. The full theory and mathematical background can be found in [27].

Fourier Series is applied in order to analyze periodic signals composed of a finite sum of harmonic. This analysis allows the breaking down of complex periodic waveforms into simpler sinusoidal components, which are easier to analyze and interpret when the overall period of the signal $x(t)$ is finite.

$$x(t) = x(t + nT_0) \quad (2.1)$$

The reciprocal of the reported period is equal to the fundamental frequency according to the relation: $f = 1/T_0$. Therefore the Fourier series of the signal $x(t)$ is given by equation 2.2:

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} M_n \cos(2\pi n f_1 t + \phi_n) \quad (2.2)$$

This representation simplifies the analysis of periodic signals by decomposing them into their constituent frequencies. The complex form of the Fourier series, along with its integral form, provides a more compact representation:

$$x(t) = \sum_{-\infty}^{\infty} c_n e^{j2\pi \frac{nt}{T_0}} \quad \text{with} \quad c_n = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-j2\pi \frac{nt}{T_0}} dt \quad (2.3)$$

where: M_n are the amplitudes and Φ_n the phases at the respective frequencies f_n . The Fourier transform of a continuous-time signal instead is not assumed to be periodic and it gives its frequency domain representation according to the following relation for continuous *Fourier transform*, $X(f)$:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (2.4)$$

Conversion back to the time domain can be obtained by Inverse Fourier Transform, defined as:

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \quad (2.5)$$

However, real signals in data acquisition systems are sampled in time, so discrete. To obtain spectra of sampled signals we use *Discrete Fourier Transform* (DFT), which are expressed in its direct and inverse transform by the formulas:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk} \quad (2.6)$$

$$x(n) = \sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}nk} \quad (2.7)$$

where N represents the length of the signal and the number of samples according to the signal sampling frequency f_{samp} . This frequency represents the highest value that can be correctly represented in a sampled signal. It is important to consider that in order to have a correct reconstruction of the signal the limiting cut-off frequency is represented by the Nyquist frequency $f_{samp}/2$. According to Nyquist Shannon's sampling theorem [27] components of the spectrum can only be extracted for frequencies between

$$-\frac{f_{samp}}{2} \leq f \leq \frac{f_{samp}}{2} \quad (2.8)$$

This restriction is necessary to avoid aliasing, a distortion phenomenon where overlapping of frequencies may occur (high-frequency components folding back into lower frequencies) if the sampling rate is below the expected Nyquist rate. Aliasing can cause significant errors in the interpretation of frequency components and must be mitigated by ensuring proper sampling rates and applying anti-aliasing filters when necessary. FFT is a simply very efficient algorithm for calculating the DFT equations which is optimised with respect to computing time and memory consumption. It rapidly computes such transformations by factorizing the DFT matrix into a product of sparse factors. As a result, it manages to reduce the complexity of computing the DFT from $O(N^2)$, which arises if one simply applies the definition of DFT, to $O(N \log N)$, where N is the data size. The full details of the FFT algorithm are beyond the scope of this Thesis and can be found in Oppenheim et al. [27]. Nevertheless, according to the mentioned advantages, FFT is the chosen method to perform frequency domain analysis and such algorithms are commonly present in multiple programming environments such as MATLAB.

2.7.1. Order Analysis

In analyzing rotating machine vibrations it is often desired to have a frequency x-axis based on harmonics or 'orders' of shaft speed. This is done to avoid smearing discrete frequency components due to speed fluctuations. These techniques are applied to asynchronously sampled signals (i.e. with a constant sample rate in Hertz) to obtain the same signal sampled at constant angular increments of a reference shaft. Spectra, where the frequency axis is normalized with shaft rotation frequency, are denoted normalized order spectra.

Using shaft speed measurements, COT is a resampling and interpolation technique that transforms a discrete signal from the time domain to the angular domain. An order (X) refers to a frequency multiple

of the reference machine rotational speed ω , in rev/min, and frequency f in Hz, through the relation [15]:

$$f = X \frac{\omega}{60} \quad (2.9)$$

In normalized order spectra, the positions of vibration signature peaks are invariant to changes in speed, simplifying the spectra' interpretation. By normalizing the frequency axis to the shaft rotation frequency, any cyclic event synchronized with the shaft rotation will produce a spectral component at a consistent position, even when speed varies. This property ensures that the spectral components remain at fixed positions, facilitating easier analysis under variable speed conditions, its accuracy is then linked to the adopted interpolation method.

COT Algorithm

According to K. R. Fyfe et al. (1996) [28], this procedure can be carried out following a few main steps, the synthesized algorithm procedure is reported in this section.

Initially, the time instants T_k ($k = 1 : K$) corresponding to integer rotations of the shaft (i.e., angle equal to $2\pi k$) have to be estimated. Then, we define an angular rotation vector as:

$$\alpha_i = 2 \cdot \pi \frac{iK}{N} \quad (2.10)$$

This vector accounts for the desired angular resolution, given by:

$$\Delta\alpha = \frac{K}{N} \quad (2.11)$$

Subsequently, we derive a vector of time instants through an initial interpolation step:

$$t(i\Delta\alpha) = \text{interpolation}(\{2\pi k, T_k\}, \alpha_i) \quad (2.12)$$

Following that, a secondary interpolation step is employed to obtain the angular resampled signal $x(i\Delta\alpha)$ from the original time domain signal $x(j\Delta t)$:

$$x(i\Delta\alpha) = \text{interpolation}(\{x(j\Delta t), j\Delta t\}, t(i\Delta\alpha)) \quad (2.13)$$

2.7.2. Windowing

When a signal is sampled, especially for non-periodic signals, edge effects can cause spectral leakage, therefore windowing functions are essential to be applied to signals before performing Fourier Transforms. To mitigate this, window functions are applied to the signal to taper its edges smoothly to zero, reducing the boundary discontinuities. Data windowing minimizes the magnitude of the spectrum, so a trade-off is needed when choosing the window to be applied between main lobe width and side-lobe height. A detailed explanation of the window function can be found in [27], in Figure 2.7 a comparison between three different window functions on a 100 sample length is represented. They are all compared against the rectangular window function which corresponds to not applying any windowing, it gives ones for all the signal length components and zeroes otherwise, while the others are used to be multiplied by a data block to smooth the start and end part of the signal window length to zero forcing periodicity between the segments.

The most commonly used window functions are Hann (Hanning) and Hamming [15], whose functions are expressed by the following relationship:

$$\text{Hann} \quad w[n] = 0.5 \left[1 - \cos \left(2\pi \frac{n}{N} \right) \right] \quad (2.14)$$

$$\text{Hamming} \quad w[n] = 0.54 - 0.46 \cos \left(2\pi \frac{n}{N} \right) \quad (2.15)$$

They both belong to the family of raised cosine functions and are characterized by:

$$w[n] = \begin{cases} \alpha + (1 - \alpha) \cos\left(\frac{2\pi}{N}n\right) & -N/2 \leq n \leq N/2 \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

where the parameter α defines the degree of taper for $0.5 \leq \alpha \leq 1$. When $\alpha = 1$ we have the rectangular window with zero taper, $\alpha = 0.54$ corresponds to Hamming window, and when $\alpha = 0.5$ we have the maximum taper with Hann window.

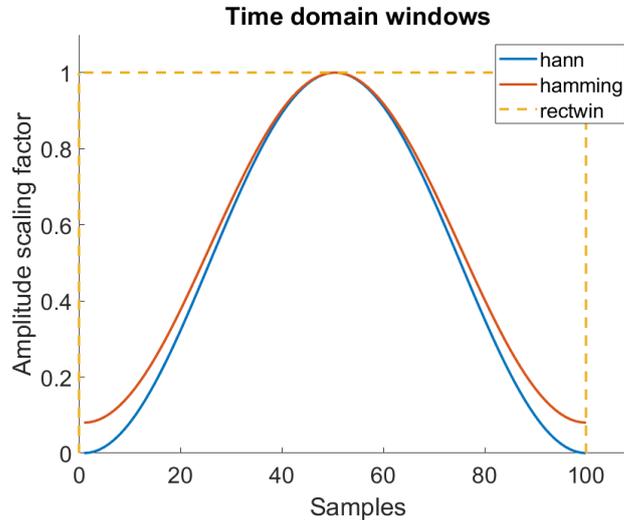


Figure 2.7: Windows functions, inspired by Barszcz [10]

When selecting an appropriate truncation window, it's essential to consider both the characteristics of the signal under analysis and the properties of the system that generates it. These factors play a significant role in determining the most suitable window function for the task. In particular, These two similarly named Hamming and Hanning (more properly referred to as Hann) window functions have a sinusoidal shape. The difference is that the Hanning window touches zero at both ends, removing any discontinuity. The Hamming window is suggested for noisy systems signals providing stronger sidelobes attenuation.

2.8. Time-domain Analysis

The second option for analyzing condition monitoring signals is time-domain vibration analysis. Vibration signals obtained are a series of values representing proximity, velocity, and acceleration, and in time domain analysis, the signal's amplitude is plotted against time. The signal processing method required to obtain information and feature extraction consists of the analysis of statistical parameters that go under the name of *Condition Indicators* [29]. Depending on the component that needs to be analyzed different parameters can be explored. The ones considered in this thesis, able to identify changes in the signals when it comes to fault condition detection, are: Root-mean-square value (RMS), Crest Factor (CF), and Kurtosis. New condition indicators have also been explored in the context of gearbox vibration analysis by P.J. Rzeszucinski [30] and they showed to have a good potential for fault detection. They include monitoring the amplitude of the normal probability density function (APDF) and monitoring the degree of deviations of the data set from a theoretical normal distribution (DND). In particular, they have been applied and tested in the context of bearing and gear fault detection by using residual signals. A residual signal is defined as a synchronous averaged signal in which the components related to shaft speed and gear mesh frequencies with their harmonics are removed. Unlike frequency domain analysis, time-domain signals do not inherently decompose the contributions of different system components, making it necessary to apply filtering or preprocessing techniques to isolate relevant information. In the present research work, filtering of a specific system's mode is applied (see Chapter 3). The following sections will give a detailed overview also of the mathematical implementation of the proposed condition indicators (CIs) for the signal analysis.

Root Mean Square Value (RMS)

RMS is computed from the vibration signal and describes its energy content. Nowadays, however, we deal with digitally sampled signals in CMSs, therefore the expression for RMS definition is given by:

$$s_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i^2)} \quad (2.17)$$

s_{rms} is the root mean square value of dataset s , s_i is the i -th member of s , and N is the number of points in the dataset. In the following sections of the presented work this condition indicator is going to be addressed as CI_{RMS} . Generally, the RMS value of the vibration signal is a good descriptor of the overall condition of the tested component, it is sensitive to loads and speed changes and the main usage is to monitor the overall vibration level that increases as the damage level progresses, additionally, it is useful in detecting an imbalance in rotating machinery [13]. RMS is generally preferred over simple peak value detection due to its sensitivity to noise.

Crest Factor (CF)

A crest factor is represented by the ratio of the input signal peak value to the RMS value:

$$CF = \frac{s_{peak}}{s_{RMS}} \quad (2.18)$$

where $s_{peak} = \max|s|$ is the maximum absolute value value of the signal. As damage progresses the root mean square value of the vibration signal increases its value and the crest factor decreases. The crest factor is often used in gearbox quality monitoring devices and when measurements are conducted at different rotational speeds because it is independent of speed. The second selected condition indicator presented here will be referred to as CI_{CF} throughout this study.

Kurtosis

Kurtosis is a nondimensional statistical measurement of the number of outliers in distribution and vibration analysis, it corresponds to the number of transient peaks, so it describes how peaked or flat the distribution is. A high number of peaks may be indicative of wear, its effectiveness depends on the presence is dependent of the presence of significant impulsiveness in the signal. Its mathematical definition is given by equation 2.19.

$$Kurt = \frac{N \cdot \sum_{i=1}^N (s_i - \bar{s})^4}{\left(\sum_{i=1}^N (s_i - \bar{s})^2\right)^2} \quad (2.19)$$

where N is the number of points in the time history of signal s , and s_i is the i -th point in the signal, while \bar{s} is the mean value of the signal. It is defined as the fourth centralized moment of the signal normalized by the square of the variance and will be addressed with the nomenclature CI_{Kurt} .

CI_{APDF}

This proposed new C.I. [30] is based on the idea of generating a normal Probability Density Function (PDF) for each vibration signal and recording its maximum value. The idea behind this implementation relies on the fact that when a fault develops and progresses with time new contributions start to appear in the signal causing a spreading of the amplitude distribution tails whose widening will cause the maximum value of the normal PDF curve to decrease with respect to healthy conditions, since the area under the curve of a normally distributed signal is equal to 1. Therefore this CI_{APDF} monitors changes in the amplitude of the normal probability density function of the vibration signal.

$$CI_{APDF} = \max[f(s_{APDF})] \quad \text{with} \quad f(s_{APDF}) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(s-\mu)^2}{2\sigma^2}} \quad (2.20)$$

where, $f(s)$ is the normal probability density function, σ is the standard deviation of the vibration signal $s(t)$, and μ is the mean value.

CI_{DND}

The second CI proposed is the Deviation from Normal Distribution (DND), which assesses to what extent a measured signal deviates from a simulated normal distribution [30]. The Normal Probability Plot (NPP), which is a specific type of Quantile-Quantile (Q-Q) plot, is recommended for this purpose. This technique compares the measured vibration signal to an equal number of samples from a theoretical normal distribution. If the signal is normally distributed, the points on the NPP form a straight line. Deviations from this line indicate non-normality. As faults worsen, impulsive events can occur, disrupting the normal distribution and increasing the signal's variance, which leads to noticeable deviations in the NPP. This relationship can be mathematically described as:

$$CI_{DND} = \frac{\sum_{i=1}^N |P_{a,i} - P_{t,i}|}{N} \quad (2.21)$$

where N is the number of data points in the signal, $P_{a,i}$ is the value of the actual curve at the i -th data point, $P_{t,i}$ is the value of the theoretical curve at the i -th data point. The theoretical distribution is derived from the actual signal's mean and the standard deviation values. Once the NPP is generated the area between the two curves can be calculated through equation 2.21.

2.9. Data Fusion

Hall and Llinas in 1997 [31] defined this discipline as: "*data fusion techniques combine data from multiple sensors and related information from associated databases to achieve improved accuracy and more specific inferences than could be achieved by the use of a single sensor alone*". The primary goal is to enhance the quality of information and improve decision-making processes by combining data from different sources into reliable results.

An important assessment to do when building such a system is to choose where to fuse (combine) the different data. Knowledge combination can take place at various levels and different categorization and system architecture are possible. For example in the Data Fusion Techniques review by F. Castanedo (2013) [32], five different categorization criteria are summarized and listed below.

1. System input data sources relations: they can be identified as complementary, redundant, or cooperative data;
2. Input-output data types: data in-data out, data in-feature out, feature in-feature out, feature in-decision out, decision in-decision out;
3. Abstraction level of the employed data: raw measurements, signals, and characteristics (extracted features) or symbols;
4. Classification of the data fusion process into multiple processing levels including an associated database and an information bus that connects them all;
5. Architecture type: centralized, decentralized or distributed, depending on the information processing and data fusion take place in a single processor or locally in each node and then exchanged, or alternatively, in the distributed architecture the data are first processed independently and then sent to the fusion node.

However, the most common and widely spread categorization for knowledge combination is raw data fusion level, feature extraction fusion level, and decision fusion level represented in figure 2.8 [33].

- **Raw data fusion level:** Data fusion at a raw stage can occur when data are of the same type, i.e., representing the same physical parameter. Once the signals are combined in a shared central node, a vector is extracted and classical detection methods and pattern recognition techniques can be applied. This comes with some disadvantages since the fault information detected by one sensor could be lost in the fusion process with other sensor signals that didn't measure it.
- **Feature extraction fusion level:** At this level, features are extracted from each sensor according to the type of raw data and successively all the feature vectors are fused and used for pattern recognition.

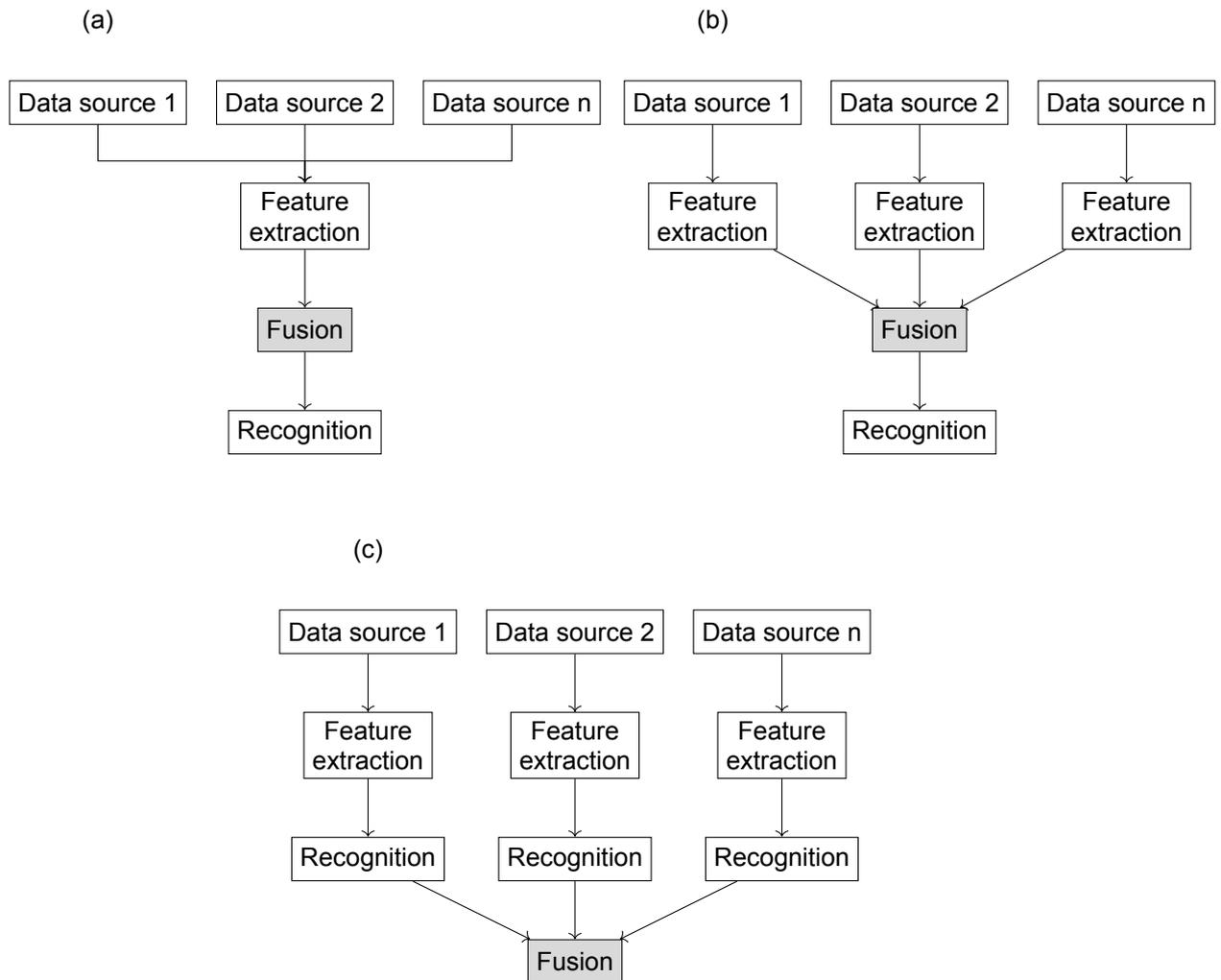


Figure 2.8: a) Raw level b) Feature level and c) Decision level data fusion methods, reproduced from Niu et al. (2019) [33]

Different techniques can be adopted at this point, for example, clustering algorithms group data points based on their similarities without needing prior labels, forming clusters where internal variations are minimized, and differences between clusters are maximized. Artificial Neural Networks (ANNs), inspired by the human brain, consist of interconnected neurons that learn patterns through weighted connections. Multi-layer perceptrons (MLPs) are a common ANN structure capable of classifying complex patterns. Physical models use known characteristics of systems to create predictive models, comparing sensor features to identify matches. While accurate, these models require significant computational power. Parametric templates compare observed data features to predefined classes based on historical data, categorizing features into distinct, non-overlapping boundaries. Knowledge-based models use AI to mimic human reasoning, relying on extensive expert knowledge databases. These systems retrieve specific information in response to queries but require much prior data knowledge, which can be hard to obtain.

- **Decision fusion level:** The decision fusion level involves making separate assessments from each sensor and combining these to reach a final decision. This approach reduces the number of sensors needed while still integrating diverse information to form a comprehensive assessment.

Classical inference uses empirical probabilities from long-term observations to validate assumptions. It is straightforward but limited to repeatable processes and binary decision alternatives. Bayesian inference has some improvements by updating probabilities based on new evidence, using empirical data, historical records, and subjective insights. This method can handle multiple hypotheses and improve accuracy over time, even though it can be complex and relies on

subjective probabilities. Fuzzy logic allows for intermediate values between 0 and 1, giving the possibility of dealing with different degrees of truth. This approach models human-like decision processes more accurately, giving the possibility of handling uncertainties. Finally, heuristic methods mimic human decision-making in groups, using techniques like voting, scoring, and Q-sort models to aggregate opinions and simulate consensus-building processes. It results in an intuitive approach, but on the other hand, these methods may lack mathematical rigor and clear quantifiability. By using these techniques, decision fusion combines sensor data into a unified and actionable assessment, ensuring robust condition monitoring.

3

Methodology and Tools

This chapter provides a comprehensive overview of the simulation environment, the chosen simulation parameters, and the analysis procedures used in this thesis. It begins by introducing the drivetrain model, discussing its characteristics and design in detail (Section 3.1). Section 3.2 presents the layout of the decoupled analysis utilized in this study, highlighting the adopted simulation framework and approach. Next, Section 3.3 outlines the steady-state simulation parameters, which form the basis for analyzing various faulty cases and inflow conditions under different operational scenarios. Section 3.4 describes the turbulence conditions, a critical factor in understanding the drivetrain performance and behavior in realistic environments, which are included in the second set of simulations. Section 3.5 describes the methods employed for the analysis of the drivetrain signals, looking at both the frequency domain (Section 3.5.1) and the time domain (Section 3.5.2), with a focus on the techniques used to compare simulation outputs features. Finally, Section 3.6 addresses the choice of the signals, with an in-depth examination of the system dynamics (Section 3.6.1) and how these considerations influence signal selection.

3.1. The 10 MW Medium-speed NTNU Drivetrain Model

SIMULIA SIMPACK [34] is a complete software package dedicated to the development and simulation of high-fidelity multibody system (MBS) virtual prototypes. The components of mechanical systems are modeled as separate parts and their interconnections are defined by force elements and joints. The level of fidelity obtained depends on this modeling choice. The basic concept of SIMPACK is to create the equation of motion of a mechanical system model based on the given degrees of freedom (DOF) and apply mathematical tools like time integration to solve them. Joints define the DOFs, which can vary from 0 to 6. Depending on the DOFs, MBSs can be divided into three types [35]:

1. Purely torsional multibody models, which focus solely on rotational motion around a fixed axis, ignoring any translational motion or deformation of the bodies. Therefore only one DOF is allowed and the other five are fixed.
2. Rigid multibody models, assume that the bodies are perfectly rigid, meaning they do not deform under load. In these models, all connections and constraints between bodies are treated as rigid joints, and both translational and rotational movements are considered. It is also possible to combine rigid bodies with discrete flexible joint elements, usually to investigate the influence of bearing flexibilities. These types of models include up to 6 DOF.
3. Flexible multibody models, which account for the deformation of bodies under loading, incorporating both translational and rotational motions along with the effects of structural flexibility. Typically, these models can have a large number of DOFs.

This Section presents the 10 MW medium-speed drivetrain model developed by the Norwegian University of Science and Technology (NTNU) [36] which is the chosen tool for the simulations performed in this research work. The NTNU model was designed in SIMPACK via the multibody system approach. To compromise between the accuracy of the dynamic behaviors and computation time, a rigid-flexible

coupled drivetrain model has been implemented. This includes the main shaft, planet carriers, and transmission shafts in the gearbox represented by flexible bodies, and the hub, bedplate, housing, and gears modeled as rigid bodies. The described setup has been implemented for the 10-MW reference wind turbine proposed by the Technical University of Denmark (DTU) in 2013 as a scale-up of the reference 5 MW NREL wind turbine. The key characteristics of the 10 MW DTU turbine are provided in Table A.1 in Appendix A [37]. Additionally given drivetrain parameters for the 10 MW turbine have been reported for completeness in Table A.2. Most drivetrain models available in the literature use high-speed gearboxes ([38],[39]). In contrast, the current model proposes a medium-speed configuration, which balances reliability issues, particularly prevalent in high-speed systems under offshore conditions, and weight challenges associated with scaling up power ratings in direct-drive (DD) systems. Furthermore, from the Global Wind Energy report (2023) [2] it has been reported, that medium-speed drivetrains have an additional benefit since they use one-tenth of the rare earth permanent magnets needed in a DD drivetrain system. The SIMPACK drive train MBS used in this work is shown in Figure 3.1.

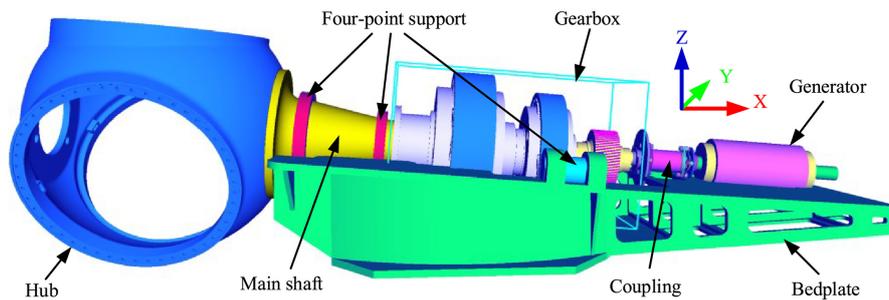
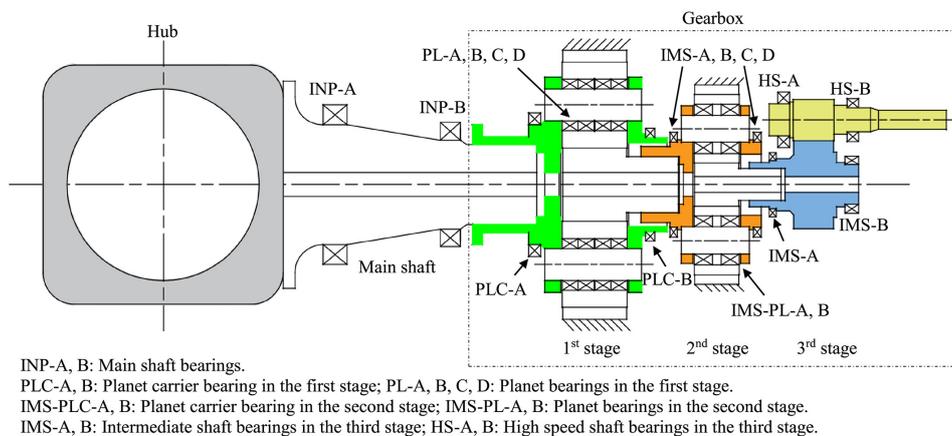


Figure 3.1: NTNU 10 MW drivetrain multibody system model [36]

The main components of the drivetrain are the hub, the main shaft, the main bearing, the gearbox, the coupling, the generator, and the bedplate. Their model is based on the design loads and criteria that are recommended in relevant international standards. More details about the general features of the different components can be found in Section 2.3 and the schematic layout of the NTNU model is shown in Figure 3.2. The MBS features a four-point support layout with two main bearings with a tapered roller design. This can prevent huge non-torque loads from entering the gearbox. The gearbox is designed with two planetary stages and one parallel stage with a 50:1 speed ratio and, lastly, a medium-speed electric generator is included. It is also important to mention that as reported by Mehlan et al. [23] the MBS drivetrain model can simulate accurately gearbox housing velocities, but it has limitations in addressing vibrations in the main bearing and generator housing. This is because these housings are not individually implemented but are treated as part of the bedplate.



INP-A, B: Main shaft bearings.
 PLC-A, B: Planet carrier bearing in the first stage; PL-A, B, C, D: Planet bearings in the first stage.
 IMS-PLC-A, B: Planet carrier bearing in the second stage; IMS-PL-A, B: Planet bearings in the second stage.
 IMS-A, B: Intermediate shaft bearings in the third stage; HS-A, B: High speed shaft bearings in the third stage.

Figure 3.2: NTNU 10 MW gearbox schematic layout [36]

According to ISO 10816-21 [18], drivetrain condition monitoring signals should be acquired by placing piezo-resistive or capacitive accelerometers on key components such as the main bearings, gearbox, and generator coupling. This ensures effective monitoring of system health. It is important to note that, in real-world applications, measurement accuracy and frequency response can be affected by the installation of the transducers, making it essential to secure the most rigid coupling possible. Figure 3.3 illustrates the selected sensor locations utilized in the simulations. In compliance with the standard, sensor measurement directions include horizontal (transverse to the rotor shaft), axial (along the rotor shaft), and vertical. For fault detection involving low characteristic frequencies (ranging from 0.1 to 10 Hz), the standard further recommends the use of velocity signals, which are obtained by integrating the measured accelerations.

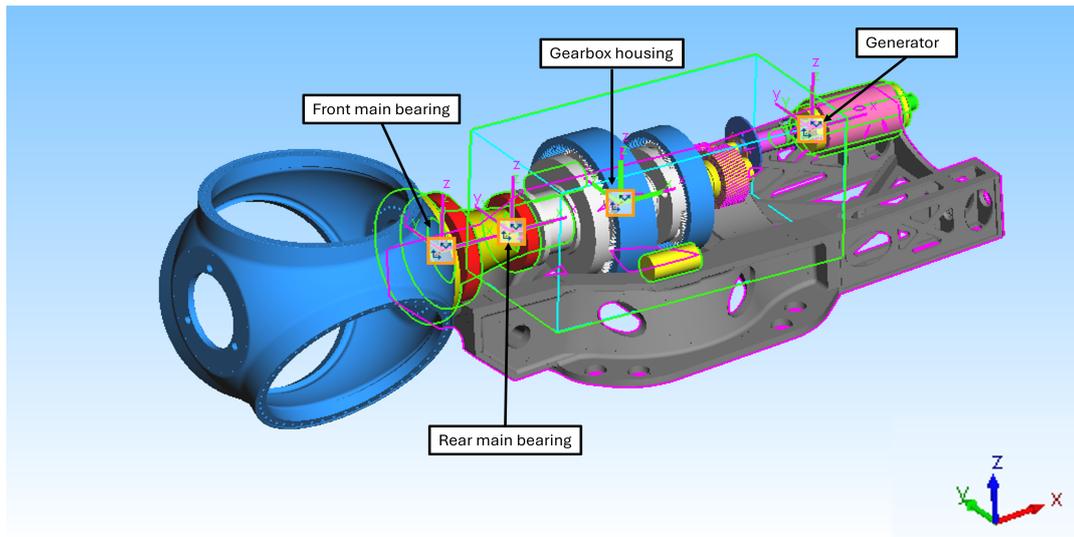


Figure 3.3: SIMPACK model sensor's location

The drivetrain modes can be obtained in SIMPACK through the "Eigenvalues analysis" tool and they are important to identify different contributions when performing analysis of the system. The full system modes of the NTNU model are reported in Table 3.1. Specifically, according to the analysis, the drivetrain torsional mode's natural frequency has been found to be equal to 4.211 Hz. The model resonance analysis and validation related to the torsional mode has been performed by Wang et al. [36].

Table 3.1: Drivetrain system modes

Mode	SIMPACK Value [Hz]
Tower fore-aft	0.24
Tower side-side	0.25
1st drivetrain free-fix	0.51
1st asymmetric flap with yaw	0.55
1st asymmetric flap with tilt	0.59
Collective flap	0.63
1st asymmetric edge	0.93
2nd asymmetric edge with yaw	0.93
2nd asymmetric flap with yaw	1.38
2nd asymmetric flap with tilt	1.59
Rotor symmetric flap	1.76
Rotor symmetric edge	1.81 11
Blade asym. flap & tower fore-aft	2.254
Blade asym. flap with tower fore-aft	2.3

3.2. Decoupled Simulations Layout

Multiple input factors are responsible for mechanical loads that act on a wind turbine, between these features we can find: rotor interactions with different inflow conditions, aero-servo-elastic interaction, and hydrodynamic loads also in the case of an offshore wind turbine. In this study, the focus is on the impact that the aero-servo-elastic interactions of the rotor of a land-based wind turbine have on the drivetrain. Based on the interaction level, drivetrain multibody models can be categorized as either fully coupled or decoupled models. Fully coupled models solve the aero-servo-elastic interaction and/or electromechanical grid interaction alongside drivetrain dynamics simultaneously. Conversely, a two-step process is adopted in decoupled models, and it is employed in this research [35]. The global analysis is conducted using first a standard aero-servo-elastic software to determine the wind turbine loads and dynamics. In this work, OpenFAST is used for this purpose [40]. Subsequently, the resultant loads and dynamics- which include forces, moments, motions, and deflection experienced by the system- are used as input for the drivetrain multibody model developed in SIMPACK. Specifically, inputs to these models include main low-speed shaft (LSS) loads (forces and moments) and nacelle motions [35]. The described procedure is shown in the schematic layout in Figure 3.4.

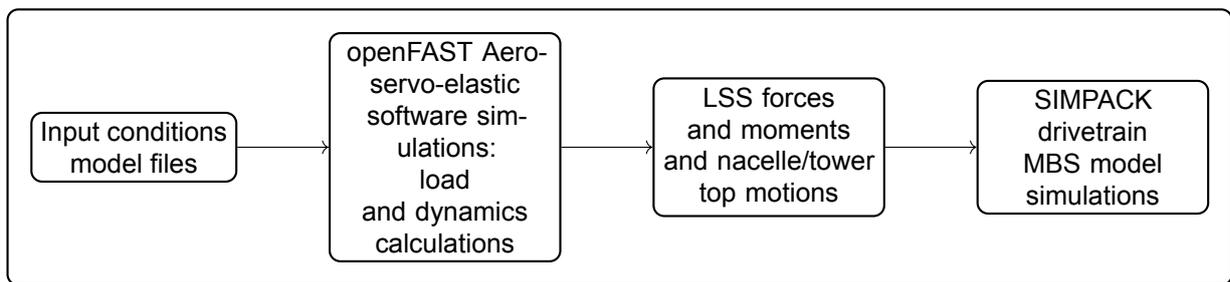


Figure 3.4: Decoupled simulation procedure

The central aim of this thesis is to address the second stage of the decoupled analysis, focusing on multibody simulations performed using SIMPACK. The aero-servo-elastic input files were prepared for various conditions of interest—these include distinct inflow profiles, environmental influences, and both healthy and faulty structural scenarios (details regarding these simulation conditions are elaborated in Sections 3.3 and 3.4). Output data generated by OpenFAST simulations, originally sourced from a previous MSc thesis project, were utilized to develop the input parameters required for the drivetrain analysis. The input variables, presented in Table 3.2, offer the critical information necessary for accurately simulating drivetrain dynamics under the specified conditions.

It is worth noting that, according to IEC 61400-13 standards [41], multiple coordinate systems must be implemented for the measurement of mechanical loads within wind turbine systems, as they are essential for interpreting simulation output parameters [42]. OpenFAST inherently supports these various coordinate systems. However, this work focuses on three primary coordinate systems relevant to the parameters under consideration:

- **Shaft coordinate system** (x_s, y_s, z_s): This system is fixed to the shaft, translating and rotating with the tower, while yawing with the nacelle and furling with the rotor. Importantly, it does not rotate with the rotor itself. All motion-related outputs from the nacelle inertial measurement unit are expressed in this coordinate frame.
- **Azimuth coordinate system** (x_a, y_a, z_a): This system shares its origin with the shaft coordinate system but rotates synchronously with the rotor. It is crucial for capturing rotor dynamics and azimuth-dependent phenomena.
- **Tower base coordinate system** (x_t, y_t, z_t): Fixed to the support platform, this coordinate system translates and rotates with the platform. Its origin is located at the intersection of the tower's centerline and its connection to the base platform. In scenarios where tower deflection is negligible, this system coincides with the tower-top coordinate frame (base plate), which does not rotate with the nacelle's yaw movement.

Table 3.2: OpenFAST output parameter for SIMPACK decoupled drivetrain analysis

Variable name	Description
GenSpeed	Angular speed of the HSS and generator (rpm)
RotThrust	LSS thrust force (this is constant along the shaft and is equivalent to the rotor thrust force) directed along the xa- and xs-axes (kN)
LSShftFys	Nonrotating LSS shear force (this is constant along the shaft) directed along the ys-axis (kN)
LSShftFzs	Nonrotating LSS shear force (this is constant along the shaft) directed along the zs-axis (kN)
LSShftTq	LSS torque (this is constant along the shaft and is equivalent to the rotor torque) about the xa- and xs-axes (kN·m)
LSSTipMys	Nonrotating LSS bending moment at the shaft tip (teeter pin for two-bladed turbines, apex of rotation for three-bladed turbines) about the ys-axis (kN·m)
LSSTipMzs	Nonrotating LSS bending moment at the shaft tip (teeter pin for two-bladed turbines, apex of rotation for three-bladed turbines) about the zs-axis (kN·m)
YawBrTDxt	Tower-top / yaw bearing fore-aft (translational) deflection (relative to the undeflected position) directed along the xt-axis (m)
YawBrTDyt	Tower-top / yaw bearing side-to-side (translational) deflection (relative to the undeflected position) directed along the yt-axis (m)
YawBrTDzt	Tower-top / yaw bearing axial (translational) deflection (relative to the undeflected position) directed along the zt-axis (m)

3.3. Steady State Simulation Parameters

The analysis performed by Mehlan et al. [23], discussed in Section 2.5, is the starting point of this research work used as a reference and validation, and from which further studies are performed. Therefore, the same values of the simulation parameters are used for the different input conditions. They are reported and explained in the present Section.

The three different types of rotor imbalances explored, introduced in Section 2.5, are mass imbalance, pitch misalignment, and yaw misalignment. For each scenario, three fault levels are selected and analyzed. Each of them will be compared to the reference healthy case conditions, for all the simulated wind speeds. The faulty cases are shown in Table 3.3. As introduced previously, the National Renewable Energy Laboratory's OpenFAST software was used to model both mass and aerodynamic imbalances in the 10 MW DTU wind turbine. The imbalances were imposed individually on one blade. Multiple simultaneous faults in more than one blade are not part of this work.

Table 3.3: Chosen fault cases and their severity levels, inspired by [23]

	Mass Imbalance	Pitch Misalignment	Yaw misalignment
Level 1	0.58 %	1°	5°
Level 2	0.88%	2°	10°
Level 3	1.17%	3°	15°

In accordance with ISO 21940-11:2016, the classification standard for addressing mass imbalance cases is denoted by the "G" scale. As highlighted by H. Schneider (2020) [43], wind turbines undergo a balancing procedure prior to commissioning, with G16 serving as the reference threshold before turbines become fully operational. Consequently, three fault severity levels are examined: G32, G48, and G64, representing mass imbalance increases of 0.58%, 0.88%, and 1.17%, respectively, affecting one blade in the most severe scenario [23].

Building on the findings of Saathoff et al. [44], this study accounts for pitch misalignment by modifying the structural twist of one blade within the aeroelastic model. Their research indicates that approxi-

mately 35.3% of operational wind turbines exhibited pitch misalignments ranging from 0.6° to 2° , with more significant misalignments—exceeding 2° —being reported in only 2.6% of cases [44].

For yaw misalignment, the selected values are based on parameters outlined in key studies within the field [45, 46].

Each case will be analyzed under six different environmental conditions, specifically, six selected wind speeds: three below- and three above-rated wind speeds, as detailed in Table 3.4. This selection is designed to cover the entire range of the turbine's power curve shown in Figure 3.5.

Table 3.4: Wind speeds simulation setup

Case name	Wind speed [m/s]
U1	4
U2	7
U3	10
U4	12
U5	14
U6	20

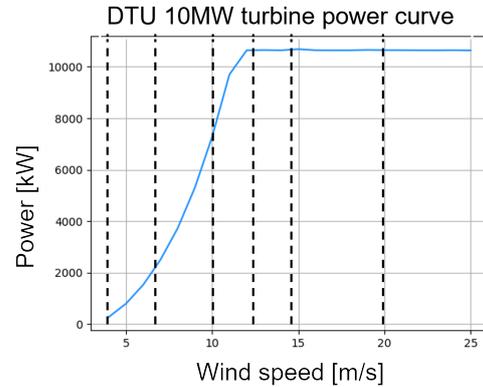


Figure 3.5: DTU 10 MW turbine power curve

Therefore, a total of 60 simulations in steady-state conditions were analyzed, 54 faulty cases combined with 6 healthy reference cases, one for each chosen wind speed. In the following chapters, the fault levels will be referred to using the color scheme and nomenclature reported in Table 3.5.

Table 3.5: Color codes and nomenclature adopted to represent the different imbalance cases (Chapters 4 and 5)

	Representation	Mass Imbalance	Pitch Misalignment	Yaw Misalignment
Level 0	Healthy	Healthy	Healthy	Healthy
Level 1	Low	MI1	PM1	YM1
Level 2	Medium	MI2	PM2	YM2
Level 3	High	MI3	PM3	YM3

3.4. Turbulence

For the second set of simulations, more realistic inflow conditions, including turbulence intensity, were modeled. Turbulent inflow conditions were generated using TurbSim, a stochastic, full-field turbulent wind simulator designed to efficiently produce randomized coherent turbulent structures around a specified mean wind speed [47]. TurbSim employs a statistical model to simulate a time series of three-component wind speed vectors at various points on a two-dimensional vertical rectangular grid, fixed in space.

When simulating flow containing coherent turbulent elements, it is advisable to conduct an ensemble of simulations using identical boundary conditions but varying the random seed. The random seed in the TurbSim file must be an integer within the range of -2147483648 to 2147483647 . Due to the inherent variability in turbulent flows, the manual advises the use of more than 30 different random seeds for a specific set of boundary conditions to ensure robust results [47]. An example of one wind speed vector for turbulent inflow conditions is reported in Figure 3.6.

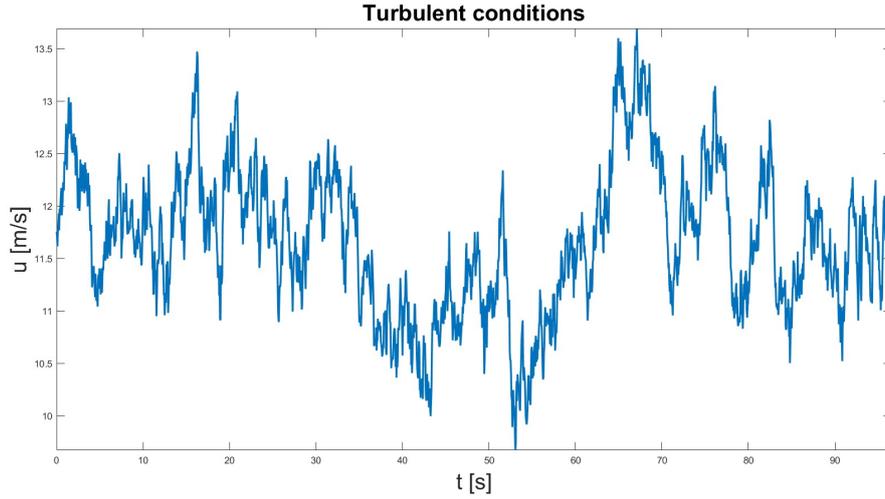


Figure 3.6: Example of turbulent inflow wind conditions of streamwise (longitudinal) direction component "u", generated through TurbSim

Due to limitations in software licensing for SIMPACK and the substantial computational resources required, it was not feasible to execute the full set of simulation seeds for each combination of input parameters. In the study by Mehlan et al. [23], six simulation seeds were used, but reproducing this across all combinations of wind speeds and fault cases was not possible in the present work.

To address this constraint and still validate the steady-state analysis under more realistic turbulence conditions, a strategic reduction was made in the number of wind speeds tested, decreasing from six to four. Specifically, U1, U3, U4, and U6 were chosen (Table 3.4). This reduction ensures that a comparable baseline is maintained between steady-state and turbulent condition simulations by eliminating two wind speeds that were redundant, while still representing all critical regions of the power curve. The fault conditions related to imbalance severity, as outlined in Table 3.3, were consistently applied across these simulations.

Given the technical limitations mentioned above, the number of seeds was restricted to two. As a result, a total of 80 simulations were performed, covering both faulty and healthy cases across four wind speeds and two turbulence seeds, thereby providing a robust dataset for analysis.

The external inflow design parameters are commonly dictated by the specific site characteristics or site classification for the intended wind turbine installation. In the present case, turbulence intensity was quantified in compliance with IEC 61400-1 standards, employing the Normal Turbulence Model (NTM) for Class A wind turbines. This model characterizes the higher turbulence levels typically encountered in onshore environments. In the NTM the representative value for turbulence standard deviation, σ_1 , is calculated taking into account the wind speed velocity at the hub height of the wind turbine, V_{hub} , its formulation is given by:

$$\sigma_1 = I_{ref}(0.75V_{hub} + b) \quad \text{with} \quad I_{ref} = 0.16 \quad b = 5.6m/s \quad (3.1)$$

Therefore the turbulence intensity, I , value is equal to:

$$I = \frac{\sigma_1}{V_{hub}} \quad (3.2)$$

which allows to calculate the turbulence intensity values for the six wind speeds, addressed as: I_{U1} , I_{U2} , I_{U3} , I_{U4} , I_{U5} and I_{U6} . The intensity values are reported in Table 3.6.

Case name	Turbulence intensity [-]
I_{U1}	0.344
I_{U2}	0.248
I_{U3}	0.2096
I_{U4}	0.1947
I_{U5}	0.184
I_{U6}	0.1648

Table 3.6: Turbulence intensity level for the chosen wind speeds

The effects of introducing turbulence into the measured signals, compared to steady-state conditions, are illustrated in Figure 3.7. This figure represents the measurement of the rotor speed for a low-severity pitch misalignment imbalance case under rated power conditions ($U_4 = 12\text{m/s}$). The plot compares the same fault case under two different inflow conditions: with and without turbulence. Notably, the increased randomness in amplitude oscillations of the measured signal due to turbulence can be observed.

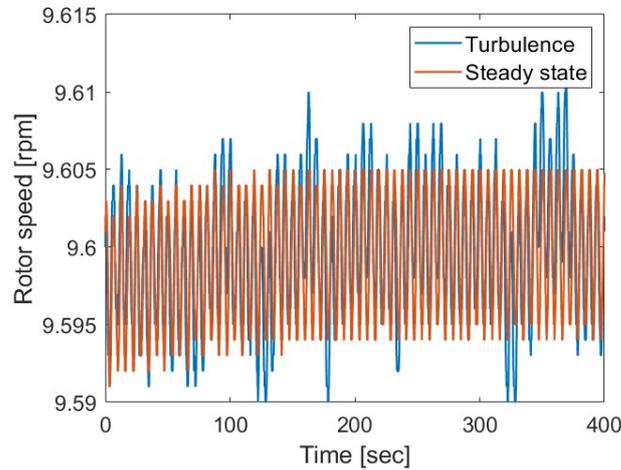


Figure 3.7: Example of turbulence effect on the rotor speed signal for the PM1 case at U_4

3.5. Analysis Method Outline

The signal analysis framework presented in this thesis is schematically outlined in Figure 3.8. It starts with an initial processing phase, applied to a selected subset of signals based on specific criteria detailed in Section 3.6. Before starting the analysis, unstable system transient effects are removed from the signals. These transients typically include initial irregularities or dynamics that do not represent the system's normal or steady-state operation. Such early transients may result from initial conditions, like non-zero initial velocities or forces, which dissipate over time and are irrelevant to the long-term behavior of the system. Including these transient effects in the analysis could lead to misleading results and inaccurate interpretations of the system's true performance.

In this study, the length of the transient-removed signal is 3600 seconds, i.e. 400 seconds are removed. The analysis then diverges into two approaches: frequency domain and time domain analysis reported in Sections 3.5.1 and 3.5.2.

These sections integrate the knowledge from the literature review presented in Chapter 2, along with basic key statistical insights required to implement the selected comparison analysis methodologies. This is done after all signal processing steps have been completed and the relevant features extracted. Chapters 4 and 5 will present the results obtained from the applied methodologies, first under steady-state conditions, followed by turbulent inflow conditions. Subsequently, Chapter 6 will synthesize these results into a comprehensive detection and diagnosis framework. This framework is structured using a

decision fusion approach, as outlined in Section 2.9, where individual assessments from each sensor signal are first performed independently, and then combined to form a final, unified decision tree.

3.5.1. Frequency Domain Analysis and Comparison

The analysis begins by focusing on the left-hand branch of the schematic workflow depicted in Figure 3.8. As discussed in Section 2.7.1, frequency domain analysis involves the implementation of the COT order resampling process. The implemented algorithm code and related flow chart are reported and explained in detail in Appendix B (Section B.2). This step is essential for aligning a varying speed system with the shaft integer rotation period, providing a stable foundation for further analysis.

The frequency domain approach has been adopted in this study to emphasize frequency patterns related to rotor operations of interest, through the extraction of the once-per-revolution rotor frequency (1P) and its third harmonics (3P) from the OpenFAST rotor speed signal. As noted in Section 2.5, 1P and 3P harmonics are crucial for detecting rotor imbalance. To perform accurate spectrum analysis using the Fast Fourier Transform (FFT), a windowing process is first necessary to prevent spectrum leakage or signal aliasing, as described in Section 2.7.2, which could distort the results. The Hann window has been selected in this case to achieve the outlined purpose. Specifically, for 3600-second signals, 60 windows of 60 seconds are extracted. The next analysis stage involves the calculation of the FFT spectrum for each of the 60 windows into which the signals have been partitioned. After the once-per-revolution rotor frequency (1P) and its third harmonics (3P) are extracted from the OpenFAST rotor speed signal, the relative peak amplitudes to the listed frequencies are identified for each signal and extracted for each window. In the following results chapters the nomenclature that is going to be used to indicate the related frequencies' peak amplitudes, is: $1\tilde{P}$ and $3\tilde{P}$. After all the mentioned features are collected for the different healthy and faulty scenarios a comparison analysis which involves multiple steps is performed and reported accordingly to the order presented in Figure 3.8.

1. **FFT analysis plots:** As an initial step, FFT plots are generated within the 0-1 Hz bandwidth for each analyzed signal to visualize the extracted spectral peaks at the identified frequencies of interest. These plots allow for a comparison between the different imbalance scenarios and their respective healthy cases under various inflow conditions. This serves as a preliminary assessment to determine whether the detection of imbalances is feasible.
2. **Averaged peak amplitudes 3D plots:** The peak amplitudes from the FFT results for each analysis window are averaged to obtain a representative value for each selected signal under the simulated operating conditions. This step aims to capture a characteristic amplitude for each condition. Subsequently, the average peak amplitudes corresponding to different fault severities are compared to those from the healthy condition across all selected wind speeds. The results are compiled and presented in a 3D plot to assess whether any discernible trends emerge.
3. **Peak amplitude relative change detection histograms:** In succession, histograms are plotted to quantify how the relative change in average peak amplitude values varies at different wind speeds with increasing severity of the fault compared to healthy conditions.
4. **Signals correlation analysis:** After the initial analysis of averaged peak amplitudes, the focus shifts to examining the complete dataset of peak values corresponding to each window—totaling 60 points for each signal. Analyzing these amplitude distributions, rather than single averaged values, allows for a broader range of analytical techniques.

One of the primary methods explored in this stage is the use of correlograms, which provide a visual representation of the correlation matrix, showing the correlation coefficients between pairs of variables. There are three common methods to calculate these correlation coefficients: Pearson, Spearman, and Kendall.

- *Pearson's* correlation coefficient is most appropriate for linear relationships and assumes that the data is normally distributed.
- *Spearman and Kendall* correlations, on the other hand, are non-parametric methods, making them suitable for more general applications where these assumptions do not hold.

The Spearman correlation coefficient (denoted by ρ) is particularly useful when assessing the monotonic relationship between two variables, regardless of whether the relationship is linear. It

operates on the ranked values of the data, making it less sensitive to outliers and not dependent on any underlying distribution. In contrast, Kendall's coefficient is often preferred for smaller datasets but is more computationally intensive, as it provides enhanced accuracy by comparing pairs of data points directly. Given the large dataset in this study, the Spearman correlation coefficient is ideal for identifying whether a monotonic relationship exists between the amplitude distributions of different signals, even when the assumptions of normality, homoscedasticity (constant variance), and linearity are not met. Spearman's correlation does not rely on these assumptions, making it robust for this analysis. The formula for the Spearman correlation coefficient is shown in Equation 3.3, with its values ranging between -1 and 1, so from a perfect negative monotonic relationship to a positive relationship.

$$\rho(a, b) = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \quad (3.3)$$

Where d is the difference between the ranks of the two variable distribution groups (a and b) for the extracted peak amplitudes and n is their length. Through the calculation of the correlation coefficients, between the different analyzed pairs, it is possible to identify the signals that exhibit the most significant response in terms of peak amplitude change due to imbalanced rotor conditions since correlations describe the mutual relationship between two variables.

After the correlation coefficients are calculated and plotted in correlograms, the amplitude distribution points are further investigated through scatter plots to visualize the relationship between signal pairs (one on the x-axis and one on the y-axis) across different fault cases and wind speeds. Each point on the scatter plot represents a pair of amplitude values—one from each variable—collected from the same measurement window, ensuring that the data points reflect simultaneous measurements. These plots are used to observe clustering patterns between data points from different fault severities and assess how they relate to each other.

5. **Peak amplitude distributions analysis:** In the following step, a more detailed analysis of the peak amplitude data distribution is performed. This can be achieved using box plots, followed by statistical multiple comparison analysis.

A box plot graphically displays the central tendency, spread, and skewness of numerical data based on quartiles, providing a standardized summary of the dataset. In particular, the representation box spans from the lower quartile (25th percentile) to the upper quartile (75th percentile), representing the interquartile range (IQR). Inside the box also the median is shown (50th percentile), and depending on its position, the skewness of the data can also be determined. Outliers, which differ significantly from the rest of the data, are plotted as individual points beyond the whiskers. In this work, the box plots are used to visualize the distribution of the peak amplitudes across different fault cases for the analyzed signals and how they compare with each other.

6. **Median values multi-comparison plots:** In the present stage, a statistical analysis of the amplitude distribution's median values is conducted. This section outlines the main statistical tests considered to guide the decision-making process for selecting the most appropriate method for this study. Detailed explanations can be found in specialized literature [48]. The statistical tests considered include ANOVA, the Kruskal–Wallis test, and the Mann-Whitney U test.

Initially, a classical one-way analysis of variance (ANOVA) was considered, but it turned out to be unsuitable for this dataset due to its strict assumptions. ANOVA requires that data across all groups have equal variances and follow a normal distribution, conditions not met by the simulated dataset. Transforming the data to meet these assumptions could lead to unreliable results, especially when the data significantly deviates from normality. Therefore, using ANOVA would provide inaccurate estimates.

Instead, the non-parametric Kruskal–Wallis test was chosen. This test is an alternative to one-way ANOVA, designed for situations where the assumptions of normality and equal variances do not hold. The Kruskal–Wallis test compares the distributions across groups by ranking the data rather than using the raw numeric values. It evaluates whether the median values of the groups

are statistically different, testing the null hypothesis that all group medians are equal. If the test rejects the null hypothesis, it indicates that at least one group's median differs from the others.

In the Kruskal–Wallis test, data are ranked from smallest to largest across all groups, and a chi-square statistic is used to determine whether the differences in ranks are significant. In cases of tied values, the average rank is assigned to all tied observations. While the test does not evaluate differences in group means, it is highly effective at identifying whether one group stochastically dominates another.

Choosing the right statistical test is essential for minimizing errors, such as false positives (Type I) or false negatives (Type II). If the Kruskal–Wallis test finds significant differences, subsequent pairwise comparisons can be performed to identify which specific groups differ. However, the focus of this analysis is on comparing all the fault groups only with the healthy condition.

For pairwise comparisons, the Mann-Whitney U test (also known as the Wilcoxon rank-sum test) is used and chosen as the final tool for this analysis stage. This non-parametric test compares two independent groups without assuming a normal distribution, making it more robust for skewed data or datasets that deviate from normality. The Mann-Whitney U test works by ranking all the values from both groups together and calculating a U statistic, which measures the difference between the rank sums of the two groups according to the formula reported in Equation 3.4. This approach is more appropriate for analyzing the fault conditions against the healthy state.

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad (3.4)$$

Where U_1 is the resulting U statistic value for Group 1, n_1 is the sample size of Group 1 and n_2 is the sample size of Group 2, R_1 instead is the sum of the ranks for the observation in Group 1. The same applies to Group 2. The smaller of the two U-values is the test statistic. The reason for this is that the Mann-Whitney U-test is based on the assumption that lower U values indicate a difference between groups, with U-values near 0 indicating a more extreme difference.

Once the U statistic is calculated, it is compared to a critical value from the U distribution, or more commonly, it is converted into a p-value. The p-value is a key concept in hypothesis testing, representing the probability of obtaining a result under the assumption that the null hypothesis (H_0) is true. If the p-value is below a significant threshold of 0.005, we reject the null hypothesis and conclude that there is a statistically significant difference between the two groups' distributions, implying a difference in medians. If this condition is met, the median values of the analyzed groups in the dataset are presented in a multiple comparison plot, clearly distinguishing between groups with a significant positive difference and those without.

This method provides insight into how value ranges vary across different wind conditions and fault severity levels introducing a new potential criterion for the detection and diagnosis of different health state scenarios.

3.5.2. Time Domain Analysis

In this section, we focus on the right-hand branch of the schematic layout shown in Figure 3.8. The time domain approach adopted in this study involves applying a filtering process to isolate the impact of a specific system mode on the mechanical components. This filtering helps prevent the overlap of various system effects and components, which can complicate interpretation. A bandpass filter (BPF) is used to allow only the frequencies within a specific band to pass, effectively suppressing those outside the selected bandwidth and avoiding the introduction of unwanted frequencies into the system. In vibration analysis, a 10% range is commonly employed ([49], [50]) to balance the isolation of the desired frequency for a particular system mode while capturing sufficient signal energy to accurately represent the mode without distortion. This approach is validated by performing the eigenvalue analysis, confirming that this range corresponds to the bandwidth encompassing all the different system modes present in the model. After processing the signals with the explained procedures, key features, also known as condition indicators, are extracted.

A condition indicator is any feature that helps distinguish between normal and faulty operation or predicts remaining useful life. Effective condition indicators cluster similar system statuses while differentiating between distinct statuses [51]. Focusing on indicators closely related to the system's physics and dynamics ensures that the information obtained is clear and relevant for fault detection.

Time domain analysis examines the system's physical and statistical parameters to identify detectable trends under different fault conditions and varying severity levels. The selected statistical indicators were discussed in Section 2.8 and their formulation is given in Equations: 2.18 (CI_{CF}), 2.19 (CI_{Kurt}), 2.17 (CI_{RMS}), 2.20 (CI_{APDF}) and 2.21 (CI_{DND}).

Multiple analysis types are involved in their determination and comparison, and they are going to be addressed also in this case in the same order given in Figure 3.8.

1. **Signals torsional mode PDFs and histograms comparison:** To ensure the validity of the condition indicators, CI_{APDF} and CI_{DND} , the input filtered signals must adhere to the initial assumption that they follow a normal distribution under healthy conditions (as discussed in Section 2.8). This analysis aims to quantify deviations from this assumption in the presence of faults. For each rotor imbalance scenario, the analysis compares four signal conditions: one healthy state and three levels of fault severity for the same vibration measurement. The process begins by evaluating the amplitude probability density function (PDF) condition indicator, its form is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3.5)$$

Therefore, the mean (μ) and standard deviation (σ) of the signals under comparison are calculated to perform the respective PDF allowing a comparison of peak amplitude values. The x-axis spans the range between the minimum and maximum values of the signals. Next, the theoretical PDFs are compared with histogram plots generated from the actual simulated signal data. This comparison assesses how well the theoretical normal distribution aligns with the observed data, providing validation for the assumption that the healthy condition signals follow a normal distribution.

2. **NPP plots for distribution fitting:** Secondly, a comparison analysis between values of the (CI_{DND}) indicator is conducted. This metric represents the area gap present between the deviation of the signal distribution from the theoretical normal distribution. The comparison is performed using Normal Probability Plots (NPP), which involve a point-by-point quantile comparison to determine whether the simulated signal data points align with the expected straight-line distribution.

To robustly assess, beyond visual inspection, whether the dataset signals fit the normal distribution, the Kolmogorov-Smirnov test can be used as a goodness-of-fit verification method. This test is non-parametric, meaning it does not assume a specific distribution for the data, making it applicable in a wide range of scenarios. It is particularly useful because of its sensitivity to differences in both location and shape between the empirical and reference distributions. The test quantifies the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution (in this case, the standard normal distribution) [52]. This can be carried out using MATLAB's built-in 'kstest' function. Since the selected test is non-parametric, it does not make strict assumptions about the structure of the input data, which is advantageous for the dataset under study. The focus of this work is not on the mathematical details of the test, which can be further explored in statistical literature [52], but rather on the insights it offers. The test aims to return a decision regarding the null hypothesis (H_0), which posits whether or not the data distribution vector originates from a standard normal distribution. The test output is $h = 1$ if the null hypothesis is rejected at the 5% significance level; otherwise, the reported output will be $h = 0$.

3. **Bar plot indicators comparison:** The three remaining condition indicator values are directly calculated through the listed formulas and their numerical results are compared across different cases using simple bar plots.

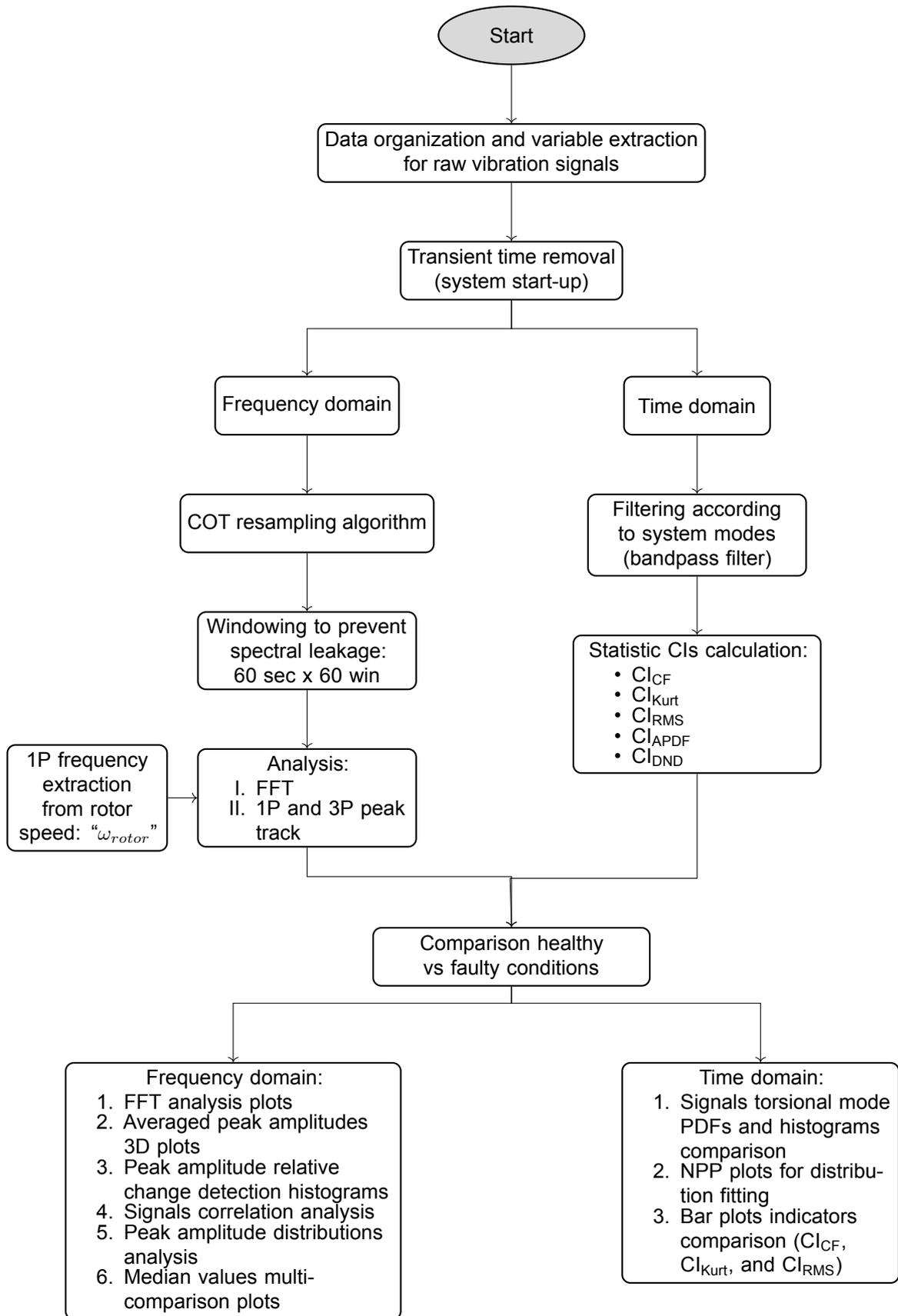


Figure 3.8: Schematic data analysis workflow

3.6. Signal Choice

As discussed in Section 2.2, various sensor types are available for wind turbine CMS. The initial focus of the proposed work is on extracting key signals from the extensive dataset typically collected by standard turbine condition monitoring systems. This dataset includes measurements of accelerations and related quantities such as velocities and displacement. In addition, in this study, loads and mechanical moments are recorded in both the stationary reference frame of the wind turbine system and the rotating reference frame of the blades. The focus of this work will be on vibration monitoring analysis, which is well-established and particularly effective in assessing the condition of mechanical drive components. This approach is widely adopted and focuses on measuring three interrelated quantities: displacement, velocity, and acceleration, which are connected through numerical integration or differentiation. As the vibration frequency increases, displacement tends to decrease, whereas acceleration increases[53]. Therefore, as the frequency rises, it is generally advisable to transition from a displacement sensor to a velocity transducer, and ultimately, to an accelerometer. The optimal frequency ranges for different transducers are detailed in Table 3.7, as reported by Tavner et al. [53].

Table 3.7: Vibration transducer and their related required frequency range, reproduced from Tavner et al. (2020) [53]

Transducer	Typical transducer output response					
	0–2 Hz	2–5 Hz	5–200 Hz	0.2–1 kHz	1–2 kHz	2–20 kHz
Displacement	Good	Good	Good	Fair	Poor	Nil
Velocity	Fair	Good	Good	Good	Fair	Poor
Accelerometer	Poor	Fair	Good	Good	Good	Good

In the drivetrain model used for this study, displacement measurements were not included as outputs, since international standards do not recommend their analysis. However, for future studies, it would be beneficial to also consider displacement measurements. This is because loading induced by rotor imbalances in wind turbines affects low-frequency vibration ranges. Section 2.5.1 explained that unbalanced rotors transmit the most intense loads at once per rotation, known as "1P". The terms 1P, 2P, 3P, and so on, refer to the frequency or periodicity of recurring variations, particularly in loads or system responses. These terms represent the number of occurrences per revolution, with 1P meaning once-per-revolution, 2P meaning twice-per-revolution, and 3P meaning three-times-per-revolution, etc. Therefore, the 1P frequency is equal to the inverse of the time taken for one full rotation, while the 3P frequency is three times that value [25]. This once-per-revolution frequency, along with its harmonics nP , is determined by the rotor's speed ω_{rotor} and can be expressed as:

$$nP[Hz] = n \cdot \left(\text{mean} \left(\frac{\omega_{rotor}}{60} \right) \right). \quad (3.6)$$

Combining the findings from the literature study reported in Table 3.7 and the low-frequency bandwidth range of interest analyzed in this study between 0 and 1 Hz (Section: 3.5.1), simulated drivetrain CMS velocity signals, were selected for the *frequency domain* analysis, in conjunction with the rotor rotational speed vector derived from the OpenFAST simulations. The latter signal is particularly relevant as it effectively captures the impact of types of imbalance that specifically influence individual blades, such as the modeled cases of mass imbalance and pitch misalignment, as discussed in Section 2.5.1. This is due to gravitational imbalance effects and allows the distinction of mass imbalance and pitch misalignment effects from those due to yaw misalignment. Table 3.8 shows the signals used in this work and their nomenclature, which will be consistently used in the analysis of the results in Chapters 4 and 5. It can be noticed that the selected subset of vibration signals of the CMS are in the side-side direction (y-axes Figure 3.3). An extensive explanation for this choice is given in the following Section 3.6.1.

Table 3.8: Signals selected for the analysis in the frequency domain and their nomenclature

Signal	Nomenclature
Front main bearing side-side velocity	$V_{MB (front), Y}$
Rear main bearing side-side velocity	$V_{MB (rear), Y}$
Gearbox housing side-side velocity	$V_{GB, Y}$
Generator side-side velocity	$V_{Gen, Y}$
Rotor rotational speed	ω_{rotor}

Several considerations were made when designing the time-domain analysis. It is well established that mechanical systems are frequently subjected to varying forces and excitations, often due to changing environmental conditions or the development of faults. In rotating machinery, these dynamic phenomena can exhibit significant non-linear behavior, a phenomenon extensively documented over the past several decades, particularly in studies of torsional vibrations such as Kahraman et al. (1997) [54].

As detailed in Section 3.1, the model incorporates two main bearings and four support systems, which are intended to mitigate the transmission of significant non-torque loads to the drivetrain. Consequently, the study by Wang et al. (2019) [36] focuses on resonance evaluation primarily in the torsional direction, as this is the direction most likely to coincide with system excitation frequencies. To avoid the influence of multiple modes that could lead to misleading results, the torsional mode has been selected as the filtering frequency for time-domain analysis, specifically to assess its sensitivity to the rotor imbalance cases under study. The torsional mode's primary frequency of interest, as mentioned in Section 3.1, is equal to 4.21 Hz. Nonetheless, the study of the axial loading drivetrain mode of the system needs also to be addressed since it reflects the impact of non-torque loads such as axial thrust and shear forces which induce blades out of plane motions [55]. Even if the used drivetrain model design mitigates these effects they still need to be investigated but this will be left for future study and improvements.

Furthermore, external excitations, such as rotor imbalances, are known to cause torque fluctuations that directly impact the torsional mode of the drivetrain ([56], [57]). This connection is further explored in Section 3.6.1. During the analysis of the simulated dataset, it was observed that the velocity vibration signals exhibited a multimodal distribution at higher frequency ranges, while the chosen condition indicators for this domain study apply only to normally distributed signals (Section 3.5.2). To address this and ensure more accurate analysis, acceleration signals, always in the side-side direction of measurements (y-axis), were chosen for the time-domain study, they are listed in Table 3.9.

Table 3.9: Signals chosen for time domain analysis and their nomenclature

Signal	Nomenclature
Front main bearing abs acceleration	$a_{MB (front), y}$
Rear main bearing abs acceleration	$a_{MB (rear), y}$
Nacelle abs acceleration	$a_{Nac, y}$
Generator abs acceleration	$a_{Gen, y}$

3.6.1. A Closer Look at the System Dynamics

It has to be mentioned that the model used in this work allows for the detection of forces, moments, and vibration quantities in all three frame directions—axial (x-axis), and radial, which is further divided into vertical (z-axis) and lateral or side-side (y-axis) directions. In the literature study by Mehlan et al. [23], the induced drivetrain responses resulting from the three rotor imbalances have been comprehensively analyzed. This investigation focuses on understanding how these imbalances affect the drivetrain and associated structures. An introduction to the topic was provided in Section 2.5, and the present paragraph addresses these aspects in summary with deeper technical details [23].

- Firstly, it is known that when it comes to *mass imbalance*, the primary excitation occurs through centrifugal forces in the lateral (Y) and vertical (Z) directions within the wind turbine's fixed frame. These periodic shear forces, driven by the rotation of the imbalance mass, excite transverse

bending modes in the tower structure, leading to lateral deflections and vibrations. Additionally, gravitational forces in the rotating frame generate periodic torque oscillations about the X-axis, which in turn excite torsional modes within the drivetrain, significantly impacting critical components such as the gearbox and main shaft.

- *Pitch misalignment* leads to uneven aerodynamic forces across the rotor, and it was found to generate substantial out-of-plane bending moments in the Y (side-side) and Z (vertical) directions within the fixed frame. These bending moments induce vibrations in the main shaft, which propagate through the drivetrain and affect downstream components like the gearbox. Furthermore, the imbalance in tangential forces due to pitch misalignment creates shear forces along the X (fore-aft) and Y (side-side) axes. These forces contribute to lateral bending of the tower, with a lesser impact on fore-aft motion.
- *Yaw misalignment*, on the other hand, implies misalignment of the rotor axis with the wind direction, it results in periodic variations in the effective angle of attack on each blade. This misalignment induces oscillating thrust and tangential forces that generate periodic shear forces along the Y (side-side) and Z (vertical) axes and bending moments in the Y-axis. The primary consequence of these forces is increased swaying of the tower. The dynamic response to yaw misalignment is complex and can vary significantly depending on operational conditions and the specific yaw angle, making it a challenging aspect to manage.

The analysis of these dynamic outcomes, grounded in an understanding of the physical relationships involved, provides insights into the most meaningful signals and directional frames to focus on. According to the study's findings [23], significant responses for all cases occur in the side-side (y-axis) direction. Furthermore, different directions of the current simulated dataset, have been analyzed and compared for each vibration signal. In Figure 3.9 a representative example is reported which clearly shows the higher impact of the modeled faults, on the y-axis vibration measurement on the gearbox housing. As a result, the chosen subset of signals for effective analysis across all imbalances emphasizes the y-axis for the chosen analysis methods. This approach leaves the axial (X) and vertical (Z) directions for further study and potential refinement in future work.

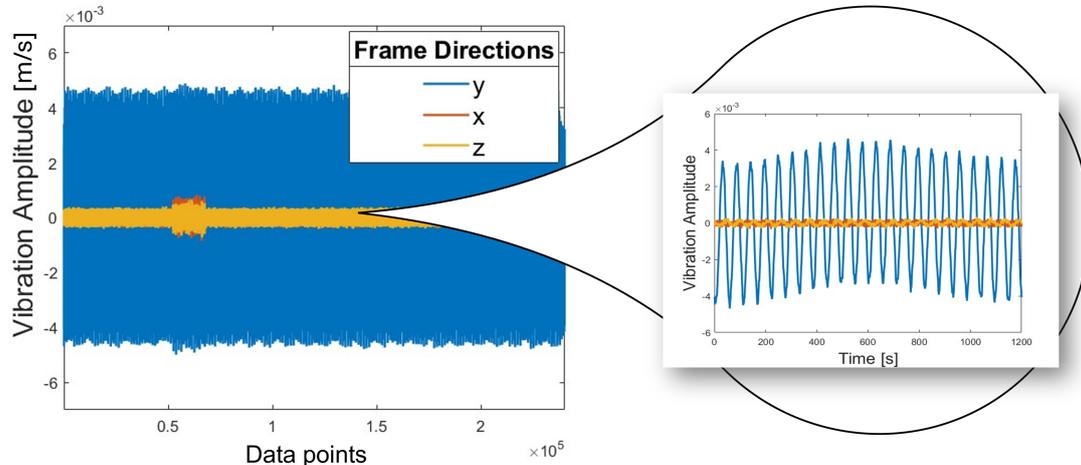


Figure 3.9: Gearbox housing vibration signal for PM2 case and wind speed U4. Comparison of the different velocities measurements along the x,y, and z axes

4

Results: Steady State Analysis

In the following chapter, the results from the initial set of 60 steady-state simulations, performed according to the input conditions detailed in Section 3.3, are presented. Extracting interpretable features is fundamental to ensuring clarity and precision when developing a classification tree algorithm for fault detection. The key objective of the results' analysis, is to identify distinctive features that enable differentiation between various imbalance scenarios and their diagnosis, facilitating accurate fault diagnosis and trend analysis under the variable wind turbine operating conditions.

As outlined in Section 3.5, various analytical approaches were chosen to compare post-processed signals under both healthy and faulty conditions. Section 4.1 begins with a frequency domain analysis, using FFT plots to assess the frequency spectrum of the selected signals (4.1.1). The peak amplitude magnitudes at key frequencies are averaged across windows and displayed in a 3D plot. This allows for a comprehensive evaluation of each signal, considering all combinations of fault scenarios and inflow conditions (as specified in Section 3.3), with further details provided in Subsection 4.1.2. Next, Subsection 4.1.3 uses histograms to detect relative changes in the averaged peak amplitudes, normalized against the healthy condition, helping to identify trends across different fault severity levels. In Subsection 4.1.4, the relationships between signals are explored through correlation matrices and scatter plots of the peak amplitude distributions. These distributions are further analyzed in Subsection 4.1.5 using box plots and statistical tests to compare the median values between groups. Finally, Subsection 4.2 presents the results of a time-domain analysis, focusing on selected condition indicators to compare the signals under steady-state conditions. The chapter concludes with a summary of the key findings in Section 4.3.

4.1. Frequency Domain Analysis

In the case of steady-state conditions, the computed order tracking (COT) resampling algorithm was not applied. This decision stemmed from the observation that, in the absence of turbulence within the simulated dataset, the analysis of the frequency content of the windowed signals did not show any significant variations. Therefore, post-processing analysis for feature extraction was applied directly to the selected signals listed in Section 3.6.

The analysis focused on two key frequencies: 1P and 3P. According to what is discussed in Section 2.5.1, the 1P frequency reflects the effects of harmonic periodic loads caused by the shift in gravitational force and thrust force centers from the hub to one of the blades due to mass imbalance and pitch misalignment. On the other hand, the 3P frequency reflects the effects caused by the misalignment between the direction of the incoming wind and rotor alignment, impacting the loading conditions on all three blades.

4.1.1. FFT Analysis

The analysis presented in this section aims to confirm the feasibility of detecting the dynamic effects previously discussed through frequency domain spectrum inspection. The FFT analysis provides a

representation of the signal spectrum, enabling the identification of contributions from various system components. Figure 4.1 shows the case of the side-to-side velocity of the gearbox housing as a representative signal to demonstrate the impact of the three faults on system behavior. These signals were analyzed under rated power conditions, specifically at wind speed U_4 .

The three plots in Figure 4.1 show different fault scenarios for low-severity cases. The plot on the left depicts the mass imbalance case, followed by the pitch misalignment case in the center, and yaw misalignment on the right. In the first two graphs, the comparison between faulty and healthy signals reveals significant differences in the peak amplitudes at the 1P rotor rotational frequency, as expected, primarily due to shifts in the center of gravitational and thrust forces on the turbine's rotor side. In the third plot, which represents yaw misalignment, no significant change is observed in the 1P peak amplitude. However, in accordance with predictions, a noticeable change is detected at the third harmonic amplitude (3P). This is attributed to the yaw misalignment that affects all three rotor blades.

In the mass imbalance and yaw misalignment cases, additional system modes can be also identified, corresponding to the 3rd tower fore-aft mode and the 1st blade edge-wise mode. The tower fore-aft mode refers to the tower bending motions that can either act against or align with the wind direction. This phenomenon is important because creates a relative velocity on the blades causing variation in the lift force therefore the dynamic motion of the system can be partially damped aerodynamically. The second visible mode instead refers to the second bending mode of a wind turbine rotor and is known as edge-wise deformation, which can also be called lead-lag motion because the rotor deforms in the plane of rotation. As expected, these modes exhibit consistent amplitude levels in both healthy and faulty conditions, since they do not depend on the presence of the imbalance. The comprehensive system's mode list is detailed in Section 3.1. Furthermore, it is evident from the middle graph that the pitch misalignment case exhibits a significantly higher order of magnitude in the 1P amplitude, which hides the peaks associated with other system modes.

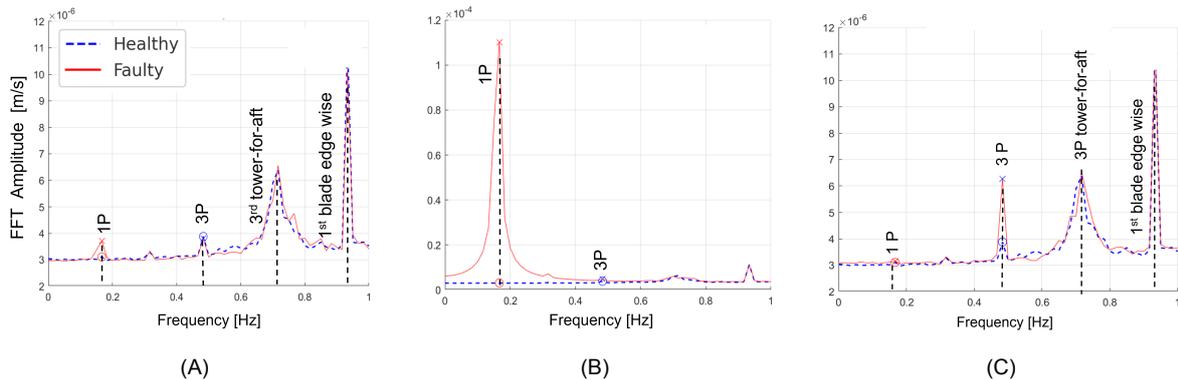


Figure 4.1: FFT plot analysis for $v_{GB, \gamma}$ signal for $U_4 = 12$ m/s. Cases: (A)MI1, (B)PM1, (C)YM1

4.1.2. Average Peak Amplitudes 3D Plots

After extracting and collecting the detected peak amplitudes for each of the 60 simulation cases in the dataset, the averaged values across all 60 windows were calculated to provide a single representative peak amplitude value for each signal, fault condition, and wind speed. These values were then plotted in a 3D layout representation to visualize all combinations together. Figure 4.2 illustrates the behavior of the gearbox housing velocity signal in the side-to-side direction for each imbalance scenario, serving as a representative example. The plots display data across six selected wind speeds and four health states (ranging from healthy conditions to level 3 fault severity) along the x- and y-axes. The corresponding averaged peak amplitude magnitudes for each combination of wind speed and health state are plotted on the z-axis. Only one signal is presented in this Chapter. The other CMS measurements showed similar and consistent behavior and they are presented in Appendix C, Section C.1.

The severity of the imbalance is reflected in the increasing amplitudes at the 1P and 3P frequencies, demonstrating a proportional dependency and upward monotonic trend across all analyzed wind speeds. External wind conditions also play a significant role, with a positive trend observed of increas-

ing amplitude values as wind speed increases looking at the same fault level, likely due to the higher aerodynamic loads. This behavior is noticeable in most of the cases and particularly clear for pitch misalignment scenarios, however, this doesn't always apply to each investigated imbalance scenario. A closer examination of each imbalance case reveals that the amplitude values for the three conditions differ in magnitude, showing varying sensitivities on the drivetrain side for each imbalance scenario.

For the case of mass imbalance, presented in Figure 4.2 (A), the 1P amplitudes under healthy conditions show a slight offset from zero, which reduces the visual slope steepness of the trend compared to the pitch misalignment case (Figure 4.2 (B)). This difference is due to the fact that the sensitivity to mass imbalance is of the order of 10^{-6} , while pitch misalignment results in amplitude magnitudes of the order of 10^{-4} . As a result, the plot resolution is adjusted to reflect higher orders of magnitude for pitch misalignment, but the same trend as in the mass imbalance case can be observed if the numerical values are examined more closely. On the other hand, in the case of yaw misalignment, shown in Figure 4.2 (C), a less pronounced linearly increasing trend is observed. In particular, in the first transition step between healthy and YM1, there is almost no change in the 3P amplitude for all the investigated wind speeds, in particular for the lower wind speeds in the cut-in region of the power curve.

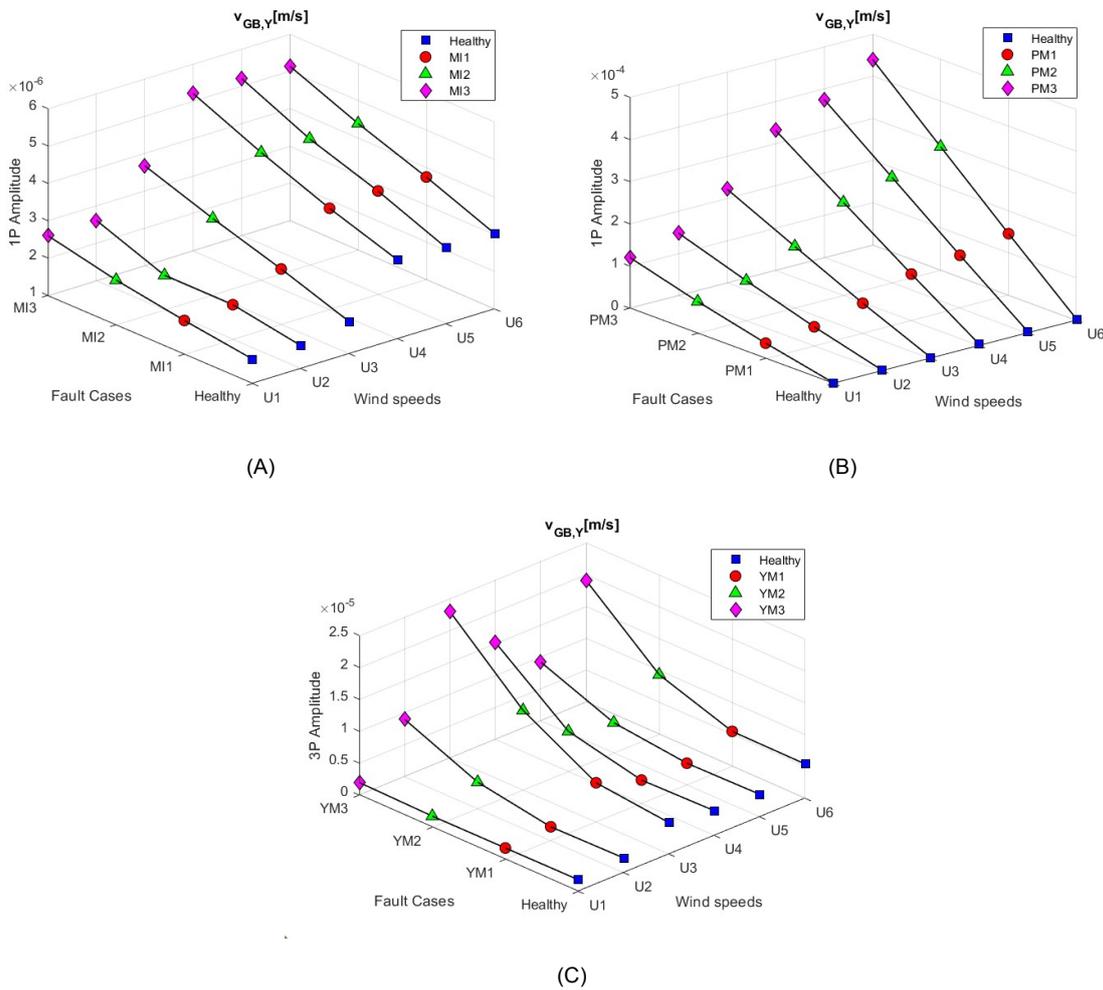


Figure 4.2: Averaged peak amplitudes of $v_{GB, \gamma}$ signal under varying health states and wind speed conditions: (A) MI - 1P, (B) PM - 1P, and (C) YM - 3P

Furthermore, considering the dynamic effects of the first two cases of imbalance (MI and PM) on the rotor side, the rotational speed signal, ω_{rotor} , measured in rpm, was analyzed in the frequency domain. This analysis was not performed for yaw misalignment, as no significant dynamic effects were observed in that case. It should be noted that this signal is not derived from the simulated CMS sensor measurements in SIMPACK but is instead obtained from the openFAST simulations.

Figure 4.3 displays the rotor speed results for all the simulated external input combinations in the cases of mass imbalance and pitch misalignment. For the mass imbalance case, an increasing trend is observed with increasing fault severity levels across the same wind speed, whereas the magnitude tends to decrease slightly with increasing wind speed for the same health condition. This behavior may be attributed to the introduction of aerodynamic damping effects on the blades, which could reduce vibrations in the rotor, leading to lower 1P amplitudes, as well as to the effects of rotor speed regulation by the control system, which can mitigate centrifugal force impacts.

For the pitch misalignment case, an increasing trend is observed, similar to the trend seen in the gearbox housing signal results. This is likely due to the more pronounced mismatch in aerodynamic forces acting on the blades resulting from thrust force imbalances.

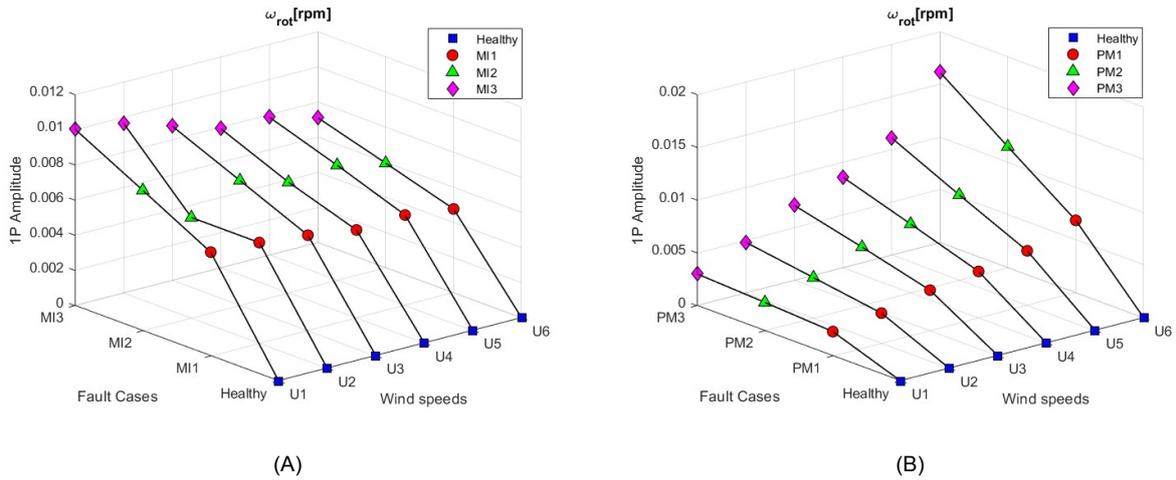


Figure 4.3: Averaged 1P peak amplitudes of the rotor speed signal, ω_{rotor} : (A) MI, (B) PM

4.1.3. Relative Amplitude Change Detection

Due to the low magnitude ranges of the detected peaks, it is beneficial to quantify the trends observed in the 3D plot analysis. For the three faulty cases and six wind speeds, the relative change in amplitude of the collected signals with respect to the healthy case was then calculated and normalized against the corresponding healthy amplitude. The histograms in Figure 4.4, show the results of the gearbox housing vibration signal, $v_{Gen, \gamma}$. The analysis for other signals, which show similar trends, is provided in Appendix C, Section C.2.

These histograms reveal a clear upward trend as the system moves from healthy to faulty conditions across all three imbalance cases. Figures 4.4 (A) and (B) show similar trends for both mass imbalance and pitch misalignment. Pitch misalignment stands out as the most sensitive case, exhibiting amplitudes up to 150 times higher than those corresponding to healthy conditions for the case of PM3 at U6. In contrast, mass imbalance, as we saw in the previous section analysis (4.1.2), is confirmed to be the least sensitive, with amplitude variations reaching no more than a factor of one relative to the healthy case. The 3P amplitudes in the case of yaw misalignment, are not sensitive enough for the detection of low fault severity or cut-in wind speeds. Nevertheless, they are effective at identifying faults of moderate to high severity levels at higher wind speeds above cut-in.

Additionally, it is important to note that the trends for the three cases do not increase linearly with wind speed. Instead, higher values are observed at a wind speed of $U_3 = 10m/s$, which corresponds to the region just before the rated wind speed of $U = 11.4m/s$. Considering the control regions of a wind turbine's power curve, the area corresponding to this wind speed is known as the transition region between the partial load control and full load control. In this region, the control strategy shifts from maximizing the power coefficient through torque control to pitch control, where torque becomes inversely proportional to the rotational speed in order to maintain a constant power output. This causes the tip-speed ratio to deviate from its optimal value. The dynamic effects of the system in this transition region are likely to influence the simulated results, contributing to the detection of higher peak amplitudes at

wind speeds equal to U_3 .

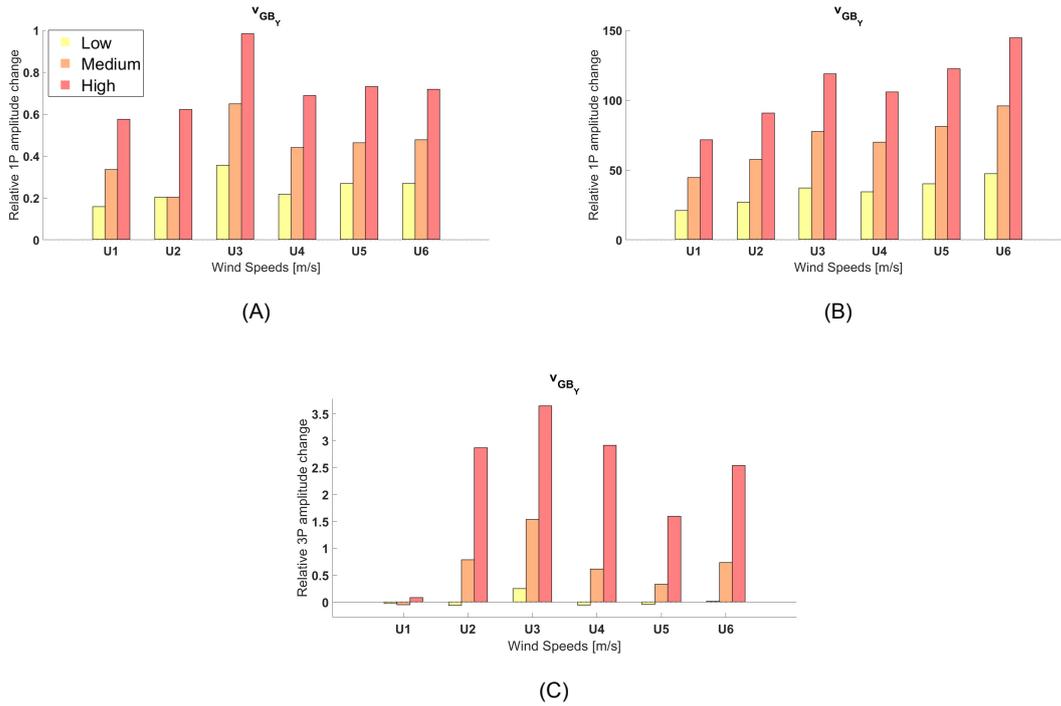


Figure 4.4: Normalised relative change in the averaged amplitudes of $v_{GB, \gamma}$: (A) MI - 1P, (B) PM - 1P, (C) YM - 3P

Table 4.1 presents the amplitude variation ranges for all analyzed signals. It summarizes the relative changes across the six wind speeds for each imbalance case, categorized into low, medium, and high levels, as depicted in the histogram plots.

Focusing on $v_{GB, \gamma}$ signal column range as an example, it is evident that while mass imbalance and yaw misalignment exhibit overlapping variation ranges between the different fault levels, pitch misalignment displays a clear distinction in fault magnitude, and this is valid across most of the signal measurements. This allows for effective fault detection and health assessment of the system contributing to avoiding the possibility of false positive or negative diagnosis. Nonetheless, it's worth it to mention that the results under steady-state wind conditions are expected to be more sensitive and easily detectable compared to turbulent inflow scenarios.

Table 4.1: Signals relative change ranges with respect to healthy conditions for different imbalance scenarios and fault severities

Imbalance Case	Fault Severity Level	$v_{MB(f), Y}$ Range	$v_{MB(r), Y}$ Range	$v_{GB, Y}$ Range	$v_{Gen, Y}$ Range
Mass Imbalance	Low	1-3	0.2-1	0.15-0.3	0.01-0.2
	Medium	2-5	0.4-2	0.2-0.6	0.1-0.4
	High	3-7	1-2.5	0.5-1	0.2-0.5
Pitch Misalignment	Low	90-130	50-100	20-50	15-30
	Medium	190-250	100-190	50-100	35-70
	High	300-400	190-300	70-150	50-100
Yaw Misalignment	Low	-0.05 - 0.3	-0.1 - 0.3	-0.01 - 0.3	-0.004 - 0.3
	Medium	-0.21 - 1.6	-0.1 - 2	-0.05 - 1.5	-0.03 - 2
	High	0.4 - 4	0.3 - 4	0.1- 4	0.05 - 4

4.1.4. Correlation Analysis

This stage of the analysis aims to assess the sensitivity of vibration signals from different sensor locations across all simulated imbalance cases, with the objective of determining their suitability for reliable fault detection. To do this, the correlation between peak amplitude values from various sensor signals has been analyzed. A comprehensive overview of this analysis is shown in the correlograms in Figure 4.5).

Unlike earlier steps, where averaged values were used, this analysis utilizes the full set of peak amplitudes extracted from the 60 windows for each signal, without averaging. The amplitude data for each signal are compared in pairs, meaning the peak amplitudes of two different signals are analyzed simultaneously to identify their correlation, as explained in Section 3.5.1. Rotor speed is also included in the analysis for the first two imbalance cases, for the reasons discussed earlier in Section 4.1.2.

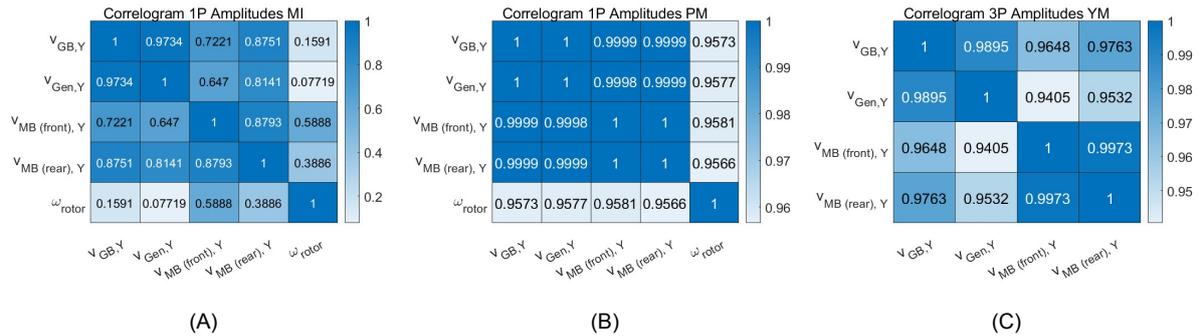


Figure 4.5: Correlation matrices between selected signals for 1P peak amplitudes ((A) MI, (B) PM and 3P amplitudes ((C) YM)

It can be observed that in general, CMS signal pairs, between main bearings, gearbox housing and generator coupling, exhibit high correlation coefficients, indicating a strong relationship in how these components respond to rotor imbalances. The correlation coefficients are equal to or approaching 1 in cases (B) and (C). This suggests that multiple sensor locations can effectively capture the 1P and 3P harmonics, which are indicative of the respective faults. In contrast, lower correlation coefficients are observed in case (A), particularly when the signals correlate to the rotor speed. This signal shows weaker correlations with the other vibration signals, with coefficients ranging from 0.15 to 0.39, indicating different responses and sensitivity to the presence of imbalances.

Scatter plots have been used to visually assess how the behaviors of these signals cluster. For each imbalance case, two types of visualizations are provided. The first set of plots compares the amplitudes corresponding to each health state (healthy, and fault severities 1, 2, and 3) of the machine, across all analyzed inflow conditions, displayed in four separate subplots. The second set includes six subplots, one for each simulated wind speed (from U_1 to U_6), showing how the fault-related amplitudes change as fault severity increases. For the figures presented in this section, in both visualizations, the x- and y-axes represent the frequency peak amplitudes for the analyzed signal pair. As known, 1P frequency peaks are extracted for the mass imbalance and pitch misalignment scenarios, and the 3P peak for the yaw misalignment case. To simplify the notation and improve clarity in the figure subplots, the signal distributions for points' amplitude on the axes will be denoted as "signal[u.m.]".

In this Section, for each imbalance case, the most effective signal pair is presented to highlight the most significant and meaningful results of this analysis. However, similar trends have been observed in other signal pairings due to the consistent correlation coefficients observed in figure 4.5. The first presented imbalance scenario is mass imbalance, followed by pitch misalignment, and finally yaw misalignment.

For the mass imbalance case, the correlation coefficients between the CMS signals range from 0.64 to 0.97. These results have been analyzed across all possible signal combinations. However, no clear clustering of the extracted peak amplitudes was observed, as they mainly tend to mix and overlap. This indicates that it is not possible to differentiate healthy/faulty behavior among the six wind speeds, nor is it possible to distinguish between different fault levels when comparing within the same wind speed.

An illustrative example examining the correlation between the 1P peak amplitudes of the generator coupling and gearbox housing is presented in Figures 4.6 and 4.7. This analysis reveals that, despite the increasing trend observed in the 3D layout of averaged peak amplitudes in Figure 4.2 (A), the differences in magnitude between various combinations of inflow conditions or fault levels are not sufficiently distinct to allow for clear clustering.

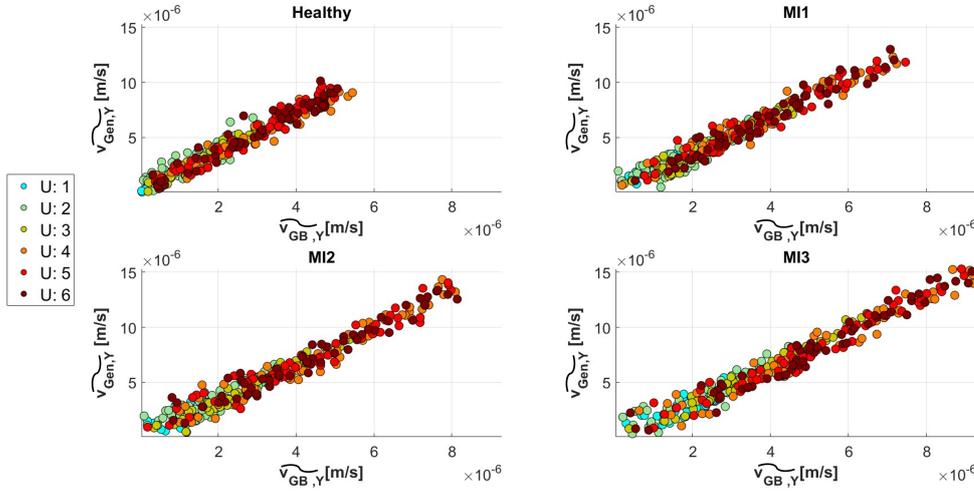


Figure 4.6: Mass imbalance - 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) MI1; (c) MI2; (d) MI3

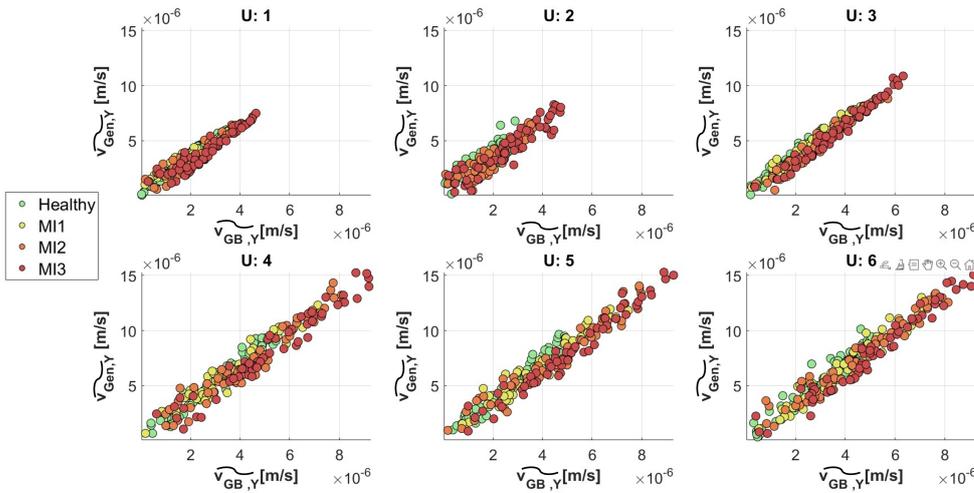


Figure 4.7: Mass imbalance - 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6

A particularly insightful analysis involves the correlation between the extracted peak amplitudes of rotor speed and the various CMS vibration signals. Figures 4.8 and 4.9 show the comparison between rotor speed and gearbox housing. A clear horizontal trend of the data point clusters can be observed. This result underscores the predominance of rotor speed in identifying the presence of mass imbalance among all other CMS signals.

In Figure 4.8, for instance, selecting a peak amplitude value from the gearbox vibration signal (x-axis) would not allow for differentiation between wind speeds. However, on the y-axis (representing rotor speed), visible discrimination is observed as the data cluster in distinct horizontal rows, each corresponding to different peak amplitude magnitudes. As expected, a decreasing trend in amplitude val-

ues is evident with increasing wind speed. This behavior aligns with the averaged amplitude analysis presented in the 3D plot in figure 4.3 earlier in this Chapter.

In Figure 4.9, a clear clustering is also observed, displaying an increasing trend within the same wind speed as the fault progresses, starting from the green healthy group to the red severe fault case (MI3).

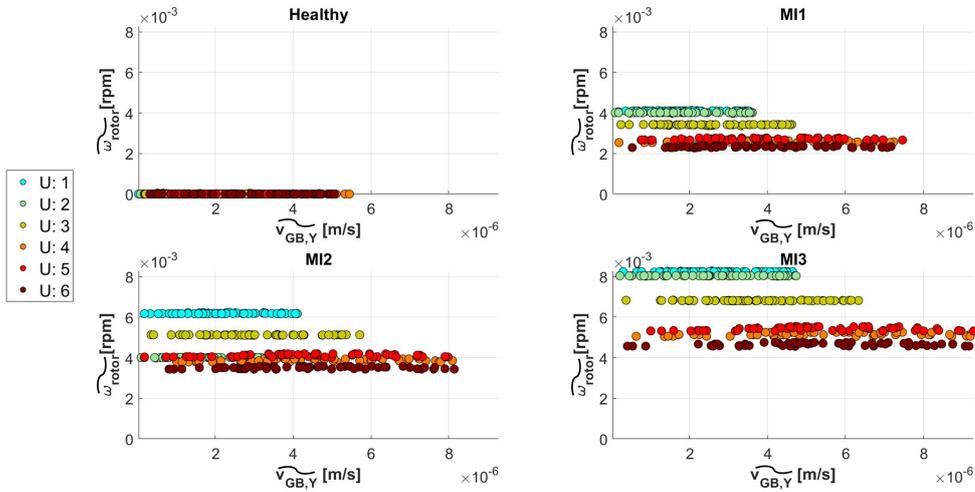


Figure 4.8: Mass imbalance - 1P amplitudes scatter plots of $v_{GB,Y}$ and ω_{rotor} signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) MI1; (c) MI2; (d) MI3

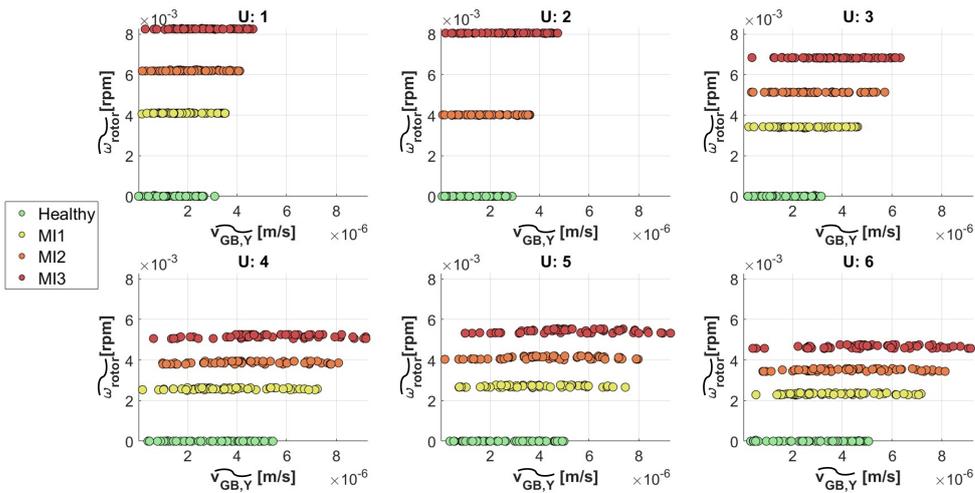


Figure 4.9: Mass imbalance - 1P amplitudes scatter plots of $v_{GB,Y}$ and ω_{rotor} signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6

Pitch misalignment stands out among the other studied imbalance cases due to its strong correlation between all analyzed signal pairs. This allows for clear clustering and a distinct, linearly increasing trend across both different environmental conditions and fault severity. In Figure 4.10, the linear trend along the plot diagonal, corresponding to increasing wind speeds, is clearly visible. The data clusters shift from the lower left, in the healthy condition subplot (Figure 4.10 (a)) with tightly clustered points and minimal variation across wind speeds, to the upper right, in PM3 fault severity, where the data points are more spread out and distinct.

The same pattern is observed in Figure 4.11, where, for each operating wind speed, a clear clustering trend is detectable across the various health conditions. For each wind speed, this indicates a proportional effect between the fault severity and the peak amplitude response of the signals. At higher

wind speeds, the effect of the fault is more pronounced, leading to greater separation between different health states in the plots. In contrast, at lower wind speeds, the signals are less sensitive to the fault, resulting in closer clustering and less distinguishable differences between health conditions.

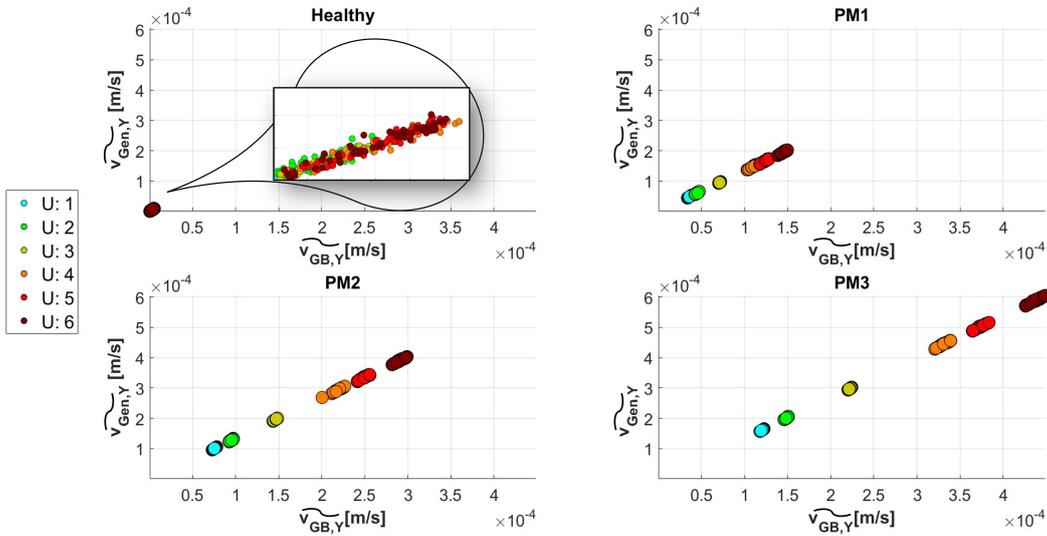


Figure 4.10: Pitch misalignment- 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) PM1; (c) PM2; (d) PM3

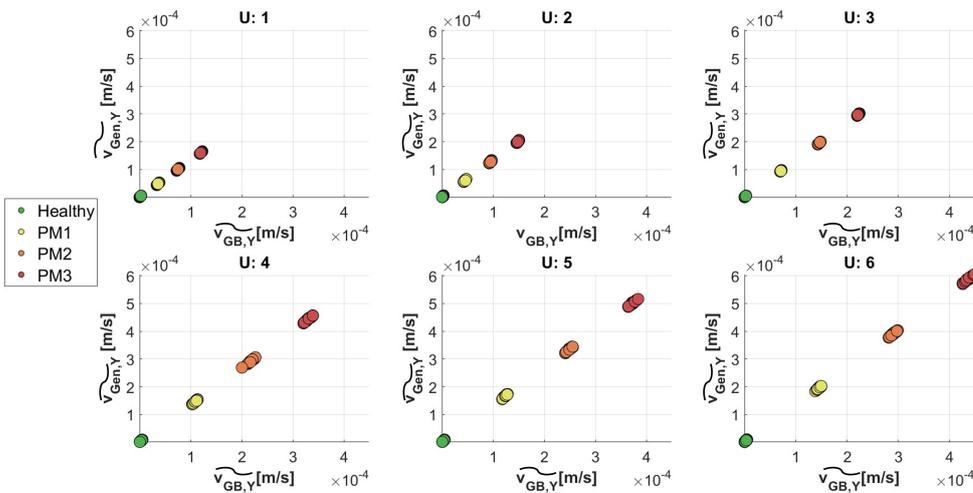


Figure 4.11: Pitch misalignment- 1P amplitudes scatter plots of $v_{GB, \gamma}$ and $v_{Gen, \gamma}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6

Lastly, the yaw misalignment scenario is analyzed through scatter plots. Linear correlation is also evident in Figures 4.12 and 4.13. However, in the former, tight clustering along the data trend line is observed in the first three subplots, making it difficult to distinguish between different wind speeds. The clusters spread out only in the last subplot, which corresponds to the highest fault severity level. Notably, as seen from the upper-right yellow points group and confirmed by the previous averaged analysis, wind speed $U_3 = 10\text{m/s}$ generates the highest amplitude response in the 3P harmonic extracted peaks.

When analyzing the system's health response at the same wind speed, Figure 4.13 shows the inability of the selected harmonic to identify differences in system vibration behavior at low cut-in wind speeds. The data points overlap without any clear distinction between health states. However, at higher wind

speeds, the data groups show more pronounced separation, particularly for higher fault severities (YM2 and YM3). As observed in the histogram analysis in Section 4.1.3, the method demonstrates limited sensitivity in detecting lower levels of yaw misalignment (YM1). The extracted amplitudes significantly overlap with those corresponding to the healthy state. Consequently, the green clusters representing the healthy condition are not distinguishable in this figure, indicating that the fault detection capability at this low severity level is not sufficiently pronounced.

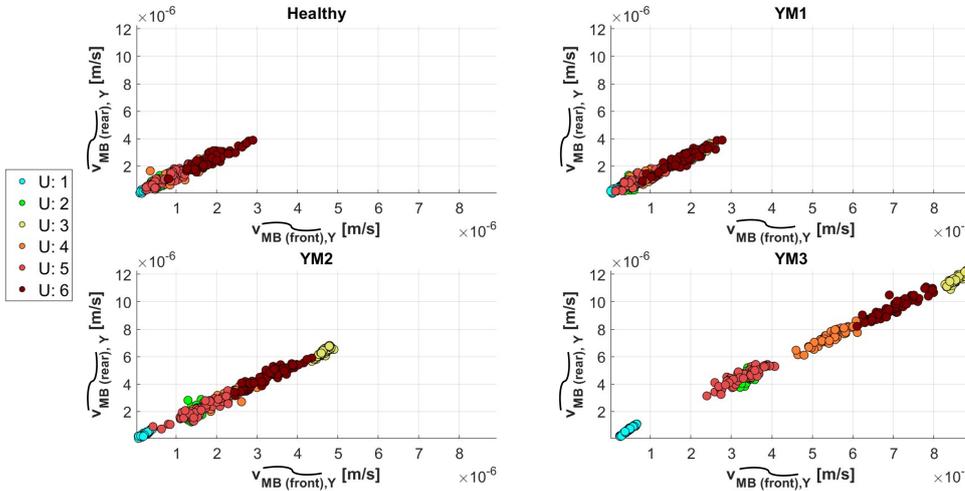


Figure 4.12: Yaw misalignment- 3P amplitudes scatter plots of $v_{MB(front),Y}$ and $v_{MB(rear),Y}$ signals. Each subplot refers to the relationship across all wind speeds for a particular health condition: (a) Healthy case; (b) YM1; (c) YM2; (d) YM3.

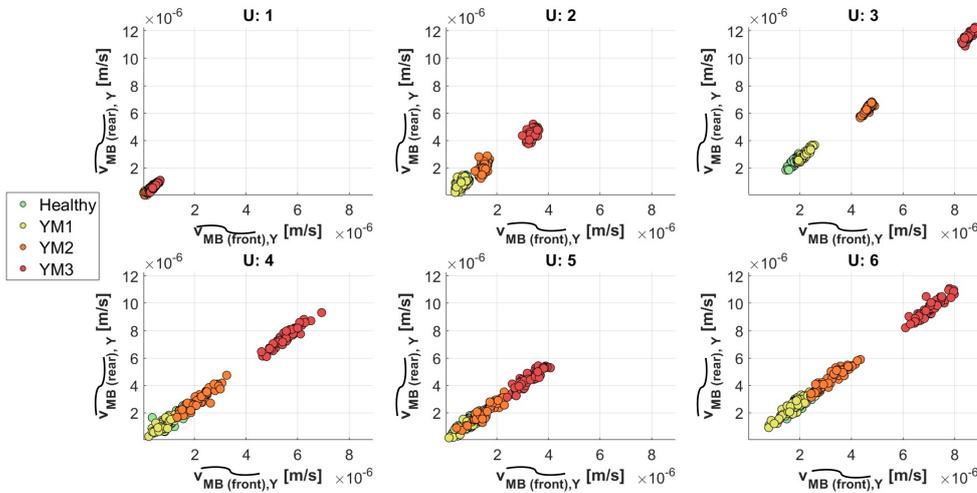


Figure 4.13: Yaw misalignment- 3P amplitudes scatter plots of $v_{MB(front),Y}$ and $v_{MB(rear),Y}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U_1 ; (b) U_2 ; (c) U_3 ; (d) U_4 ; (e) U_5 ; (f) U_6

Overall, while the correlation between CMS signals remains consistently strong, this analysis demonstrates that the ability to differentiate between fault severities improves as wind speed increases. At higher wind speeds, yaw misalignment becomes more detectable, with clearer separation between health conditions, whereas at lower wind speeds, it is more difficult to distinguish between fault levels. This suggests that the system's response to yaw misalignment is more sensitive under higher aerodynamic loading conditions.

4.1.5. Box Plots and Median Values Multi-comparison Analysis

As outlined in Chapter 3 (3.5), this section analyzes the extracted features by investigating several key aspects. The first step involves creating box plots to examine the distribution of the extracted peak harmonic amplitudes for each imbalance scenario under different wind speeds. The analysis revealed a consistent pattern in the data distribution across the various wind speeds. Given the extensive nature of the simulated dataset, this analysis aggregates the detected peak amplitudes for all three imbalance cases and selected signals across all wind speeds. This approach allows for a comprehensive assessment of how these amplitudes compare, providing deeper insight into the system's response. For completeness, the results for each individual wind speed are provided in Appendix D for the gearbox housing vibration signal as an example, the same can be achieved for the other CMS signals. A meaningful observation to point out about these individual speed plots is the clear subdivision without overlapping of the distribution boxes for pitch misalignment case, standing out again for its clear detectability also in terms of fault severity.

Three graphic representations are reported which show box plots for the three imbalance scenarios, each representing the four main CMS signals selected. Figure 4.14 refers to the mass imbalance case, Figure 4.15 to the pitch misalignment case, and finally Figure 4.16 to the yaw misalignment case. In all three cases, the figures confirm an increasing trend in the fault-related amplitude and variance ranges as the fault severity increases from healthy to more severe levels. Additionally, it can be observed that the amplitudes increase going from the turbine rotor side to the end of its power transmission system, with the generator signal exhibiting the highest vibration amplitude ranges. This strengthens the observations made with the clustering analysis described in the previous sub-section (4.1.4). Specifically, for mass imbalance the boxes for each of the four signals tend to overlap quite significantly for different health states, and the difference between their median values (shown by the red horizontal lines) is insignificant. In the case of pitch misalignment healthy and faulty states are clearly distinguishable, while yaw misalignment exhibits noticeable differences in quartile distributions mainly for the last two fault severity levels, which enables effective fault detection. However, as stated before, there is no distinction between the healthy state and the YM1 conditions.

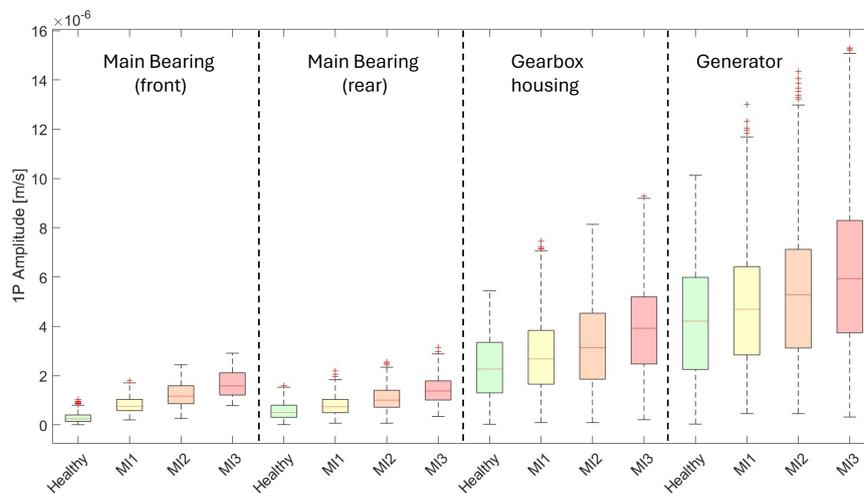


Figure 4.14: MI box plot- 1P amplitude distribution and variance of the selected monitoring signals aggregated for all wind speeds

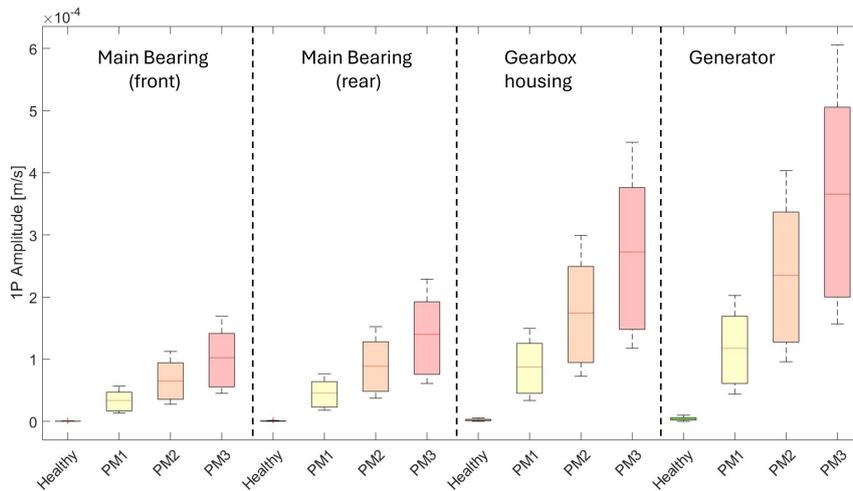


Figure 4.15: PM box plot - 1P amplitude distribution and variance of the selected monitoring signals aggregated for all wind speeds

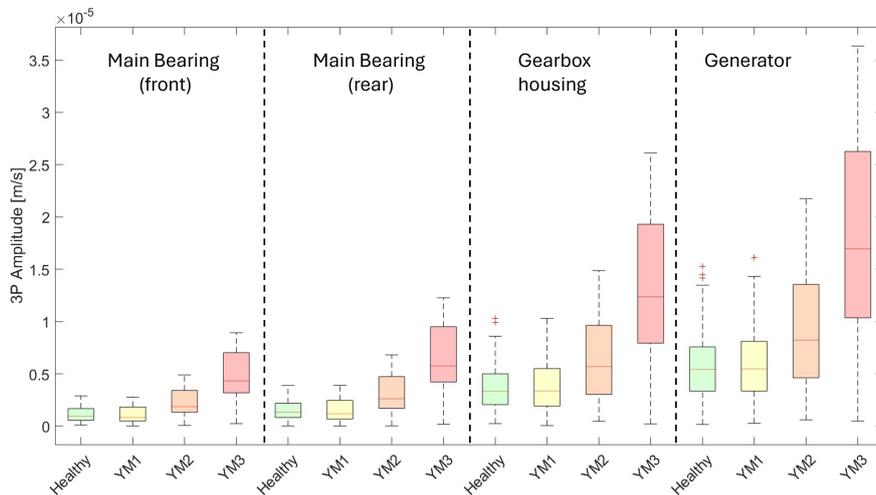


Figure 4.16: YM box plot - 3P amplitude distribution and variance of the selected monitoring signals aggregated for all wind speeds

Beyond providing an overview of the data distribution, this analysis highlights the potential of using the distribution median values of the features extracted for selected harmonics as a fault detection indicator. This is further studied by performing a statistical follow-up analysis where the Mann–Whitney U test is applied. As introduced in Section 3.5.1, the Mann–Whitney U test calculates the likelihood that two different groups have statistically significant differences in their median values. In this analysis, the test was applied in pairs, comparing each imbalance subset against the corresponding healthy condition. Also in this case, for presenting the results, the data points have been grouped among all the analyzed wind speeds. Nonetheless, the same results trend has been obtained when analyzed singularly, the figures are reported in Appendix D Section D.2.

The statistical comparison results corresponding to the three box plots from the previous figures are displayed in Figures 4.17, 4.18, and 4.19. In these plots, the x-axis represents the median amplitude magnitudes, while the y-axis lists the four CMS signals for each health condition. The blue markers indicate healthy state conditions, the red markers highlight groups with statistically significant differences compared to the healthy condition, and the grey markers represent groups where no significant difference was observed between the fault scenario and the corresponding healthy condition.

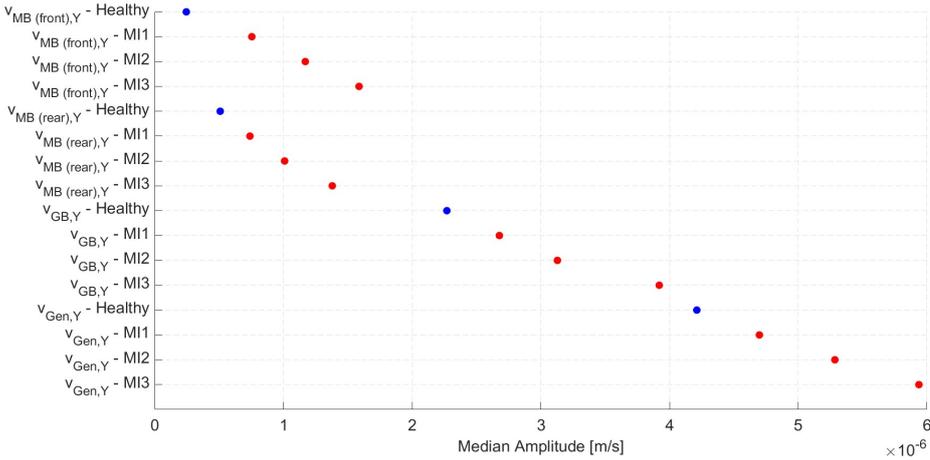


Figure 4.17: MI - Multiple comparison median values analysis for CMS signals

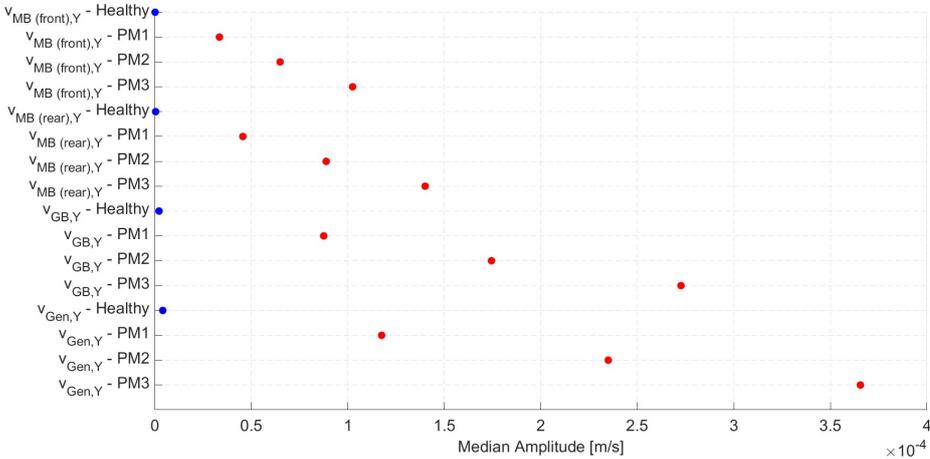


Figure 4.18: PM - Multiple comparison median values analysis for CMS signals

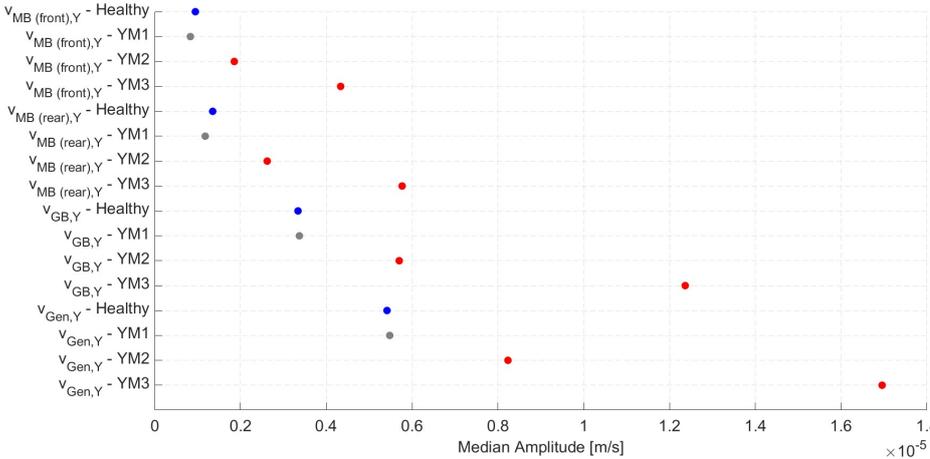


Figure 4.19: YM - Multiple comparison median values analysis for CMS signals

The considerations made in the previous box plot analysis for different imbalance scenarios also apply to the median value analysis across all CMS signals. Mass imbalance cases show only a small difference in magnitude range compared to healthy conditions. Pitch misalignment, however, demonstrates a high potential for distinguishing between different health states using the newly introduced criteria. On the other hand, yaw misalignment is confirmed to not show a significant statistical difference for low-severity cases compared to healthy conditions measurements.

4.1.6. Pitch Misalignment vs. Mass Imbalance

In the previous Section, it was observed that the three imbalance cases exhibit similar trends reflected in different magnitudes of median value ranges. However, it is important to note that although these trends are consistent, the sensitivity of the signals varies depending on the type of imbalance. Specifically, yaw misalignment stands out because it is characterized by the analysis of 3P peak amplitudes rather than 1P. In contrast, pitch misalignment and mass imbalance require further analysis and comparison to assess how, and if, they can be detected in real conditions. For this reason, distribution analysis through box plots, along with multiple comparison analyses of median values, were conducted for each signal to directly compare the two types of imbalance.

The analysis across all four CMS signals produced consistent results. However, for illustrative purposes, only the gearbox housing velocity vibration signal is presented here. Figure 4.20 shows box plots comparing the amplitude distributions of this signal under healthy conditions and different fault levels for both mass imbalance and pitch misalignment scenarios. This side-by-side comparison provides a clear visualization of how these two fault types differ.

After the Mann-Whitney statistical test is applied between the faulty groups and the corresponding healthy condition data, and Figure 4.21 is generated, displaying the median values of the distributions. It is evident from the figure that mass imbalance (MI1, MI2, and MI3, y-axis) does not show a statistically significant difference in median values (x-axis) which are represented in grey, compared to the blue marker for healthy condition. This lack of distinction limits its effectiveness in detecting mass imbalance.

In contrast, pitch misalignment (PM1, PM2, and PM3) shows a clear shift in median values, as highlighted by the red markers. The higher the fault severity, the more the median value shifts to the right, indicating higher amplitude magnitudes.

This confirms the vibration signal as a reliable indicator for detecting pitch misalignment. Similar behavior is observed across the other CMS signals analyzed, which demonstrate comparable effectiveness for pitch misalignment detection. The corresponding plots for these signals are presented in Appendix C, Section C.3.

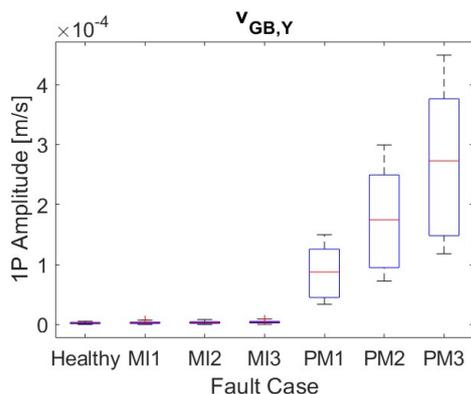


Figure 4.20: $v_{GB, \gamma}$ signal box plots peak amplitude distribution analysis for healthy case compared to MI1, MI2, MI3 and PM1, PM2, PM3 imbalance scenarios

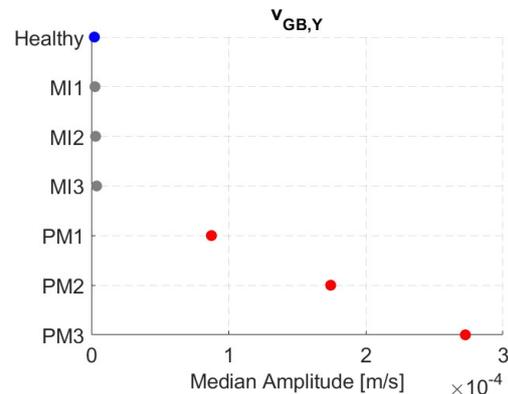


Figure 4.21: $v_{GB, \gamma}$ amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test)

The only signal that resulted to be robust enough for mass imbalance detection, giving peak amplitudes in the same range of magnitude of the pitch misalignment imbalance case, is the rotor speed. Results are shown in the box plot and multiple comparison analysis in figures 4.22 and 4.23, respectively.

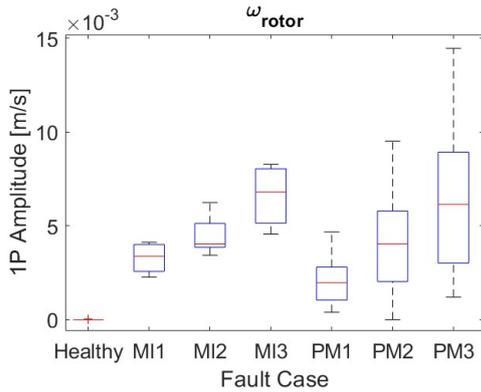


Figure 4.22: ω_{rotor} signal box plots peak amplitude distribution analysis for healthy case compared to MI1, MI2, MI3 and PM1, PM2, PM3 imbalance scenarios

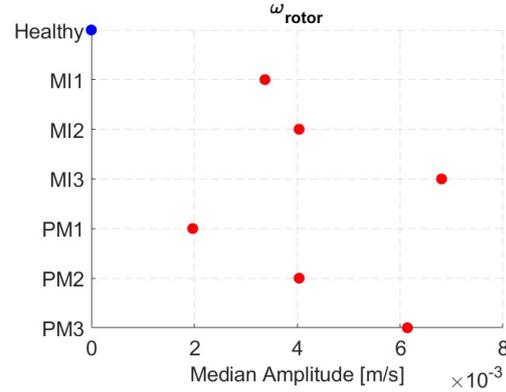


Figure 4.23: ω_{rotor} amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test)

4.2. Time Domain Analysis

As introduced in Chapter 3, two analysis domains have been selected for this thesis work. This section focuses on the results obtained from the analysis in the time domain. The system's torsional mode was identified as the primary mode of interest to be investigated (see Section 3.6). Based on the modal analysis simulations of the 10MW drivetrain model, a central frequency of 4.21 Hz has been selected and isolated as described in section 3.1. For the time-domain analysis, for each wind speed, the same signals under different health conditions are compared. After the initial post-processing steps, such as transient time removal and band-pass filtering, the selected statistical features are extracted.

The first two analyzed features are CI_{APDF} and CI_{DND} , presented in Equations 2.20 and 2.21. According to the literature, the evaluation of these two condition indicators begins with the assumption that filtered signals in healthy conditions follow a normal distribution. Any deviation from this ideal behavior is considered indicative of deterioration in the system's operating conditions. Due to the higher frequency range investigated for the torsional mode analysis compared to frequency domain features, acceleration vibration signals were selected from the CMS sensors in the side-to-side measurement direction. However, upon analyzing the model signal outputs for each imbalance case, only the main bearings signals (front and rear) showed behavior consistent with this hypothesis. Therefore, the proposed time domain analysis steps are applicable just to these two signals. In contrast, the nacelle and generator side-to-side acceleration signals exhibited healthy behavior distributions that already deviated from a normal distribution.

The first analysis step, as outlined in Section 3.5.2, involves calculating and comparing the normal probability density function (PDF) for the four signals. The first health condition indicator, CI_{APDF} , measures the amplitude difference of the PDF between the different health states. Since front and rear main bearings yielded similar results, only the front main bearing analysis outputs are reported in this section, but the same conclusions apply to the rear one. All three imbalance cases were analyzed. The results reported in this section, demonstrate a clear detection of deviations from the normal distribution for the pitch misalignment case scenarios at rated power conditions, i.e. from rated wind speed and above, namely U_4 , U_5 and U_6 . In contrast, mass imbalance and yaw misalignment scenarios have been proven to consistently match the normal distribution for the torsional mode analysis, indicating insufficient sensitivity for fault detection through torsional mode analysis. Examples of results obtained for MI and YM are reported in Section C.4 of Appendix C, for U_4 .

Figure 4.24 presents the front main bearing PDFs for four different pitch health conditions, comparing the healthy case to the three pitch misalignment fault levels, PM1, PM2, and PM3. Numerical results are collected in Table 4.2.

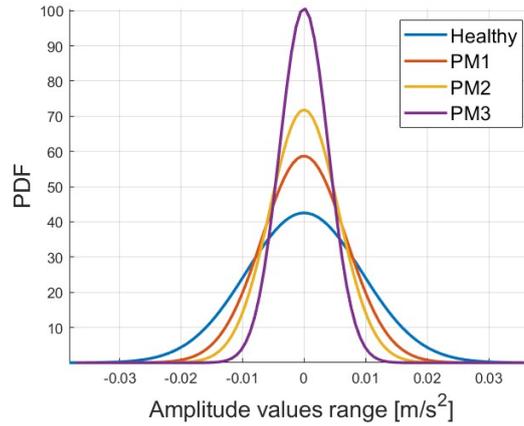


Figure 4.24: PDFs of $a_{MB, (front), Y}$ signals for healthy, PM1, PM2, PM3 conditions, at U_4

This analysis shows a noticeable increase in the peakedness of the PDF distributions as the fault severity increases, suggesting a narrower range of amplitude variations. Consequently, this leads to a rising trend in the CI_{APDF} condition indicator, reflecting worsening health conditions. On the other hand, the curves for low-level faults and healthy conditions exhibit reduced amplitude values.

To understand this behavior, we need to explore the aerodynamics behind the design of a wind turbine blade. The pitch angle refers to the angle between the chord line of the blade and the plane of rotation. By adjusting this angle, it is possible to control how much aerodynamic force is exerted on the blades. In this context blade stiffness refers to the resistance of the blade to deformation. Therefore, as we model higher severities of pitch misalignment, we also decrease the aerodynamic forces acting on the blade, leading to less flexing and bending. This causes the blade to behave in a more rigid manner, resulting in less severe vibration amplitudes, as observed in the probability density function for PM3.

When fitting the histogram representation of the signals across the four different health conditions to the theoretical normal probability density function, as shown in Figure 4.25), it becomes evident that the fit is nearly perfect for the healthy condition. However, the fit progressively deteriorates as fault severity increases. The degree of deviation from the normal distribution can be quantified using the second condition indicator, CI_{DND} , which measures the area difference between the actual measured distribution and the ideal normal distribution.

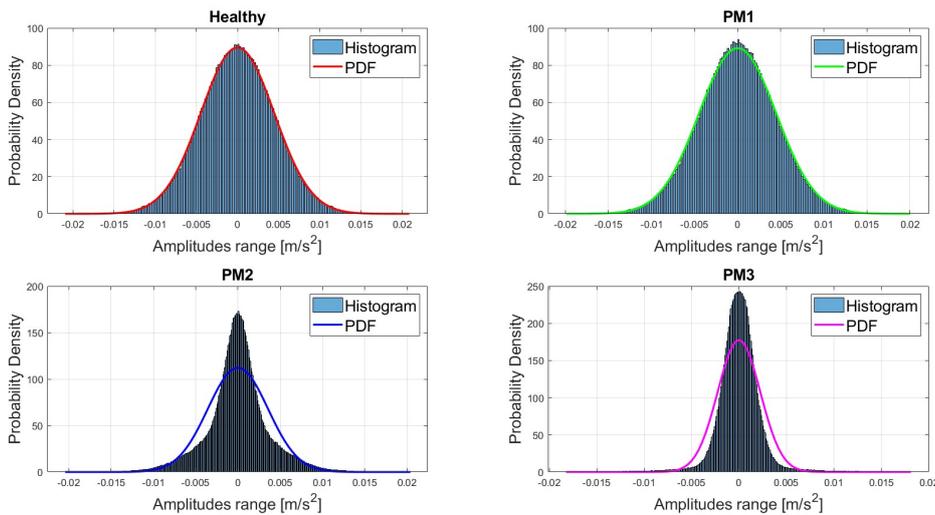


Figure 4.25: Healthy vs PM1/PM2/PM3 - $a_{MB, (front), Y}$, U_4 : histogram data fitting of the simulated dataset with respect to theoretical normal distribution.

This is visualized through the quantile comparison in the normal probability plots (NPP), as shown in Figure 4.26. The increase in area difference for the different pitch misalignment cases reflects the growing deviation in the tails of the distribution as fault severity worsens. This result matches the expectation acquired from other studies, such as [30], when dealing with health monitoring of mechanical components. Also in this case, output numerical results are collected in Table 4.2.

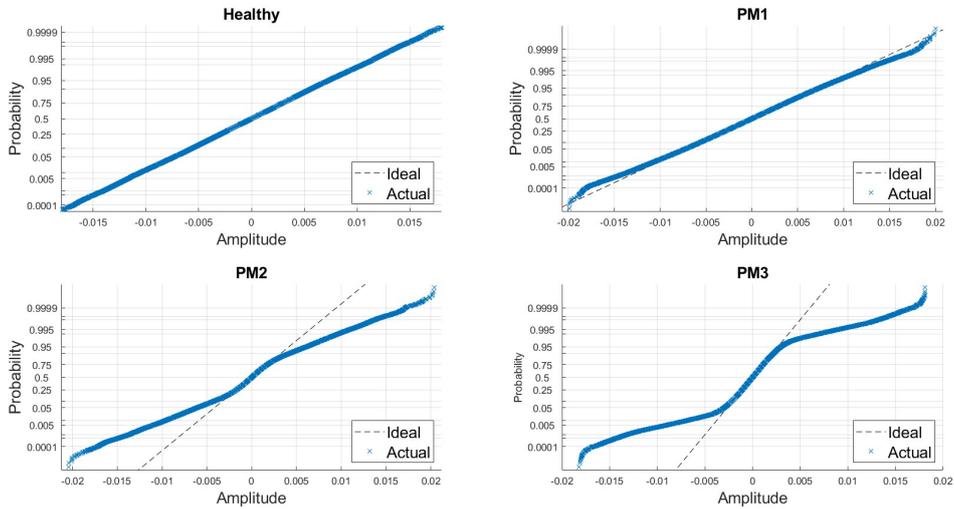


Figure 4.26: Healthy vs PM1/PM2/PM3 - a_{MB} , (*front*), γ , U_4 : NPPs for quantile comparison of simulated data with theoretical normal distribution.

To verify the goodness-of-fit between the healthy case distribution and the theoretical normal distribution, the Kolmogorov-Smirnov (KS) test is applied. The test returns $h = 1$ if the null hypothesis is rejected at the 5% significance level, indicating the data does not follow a normal distribution; otherwise, the output will be $h = 0$. The results are shown in the fourth column of table 4.2 labeled as "KS_h_value". As expected the data match a normal distribution just for healthy conditions.

Lastly, the three remaining condition indicators, the crest factor (CI_{CF}), kurtosis ($CI_{Kurtosis}$), and root-mean-square value (CI_{RMS}), are calculated according to Equations 2.18, 2.19, 2.17. Graphical representation is provided using bar plots in Figure 4.27 and numerical values reported in 4.2. The plots and the data shown reveal an increasing trend for the first two indicators, namely the crest factor and the kurtosis, and a decreasing trend for the root-mean-square value.

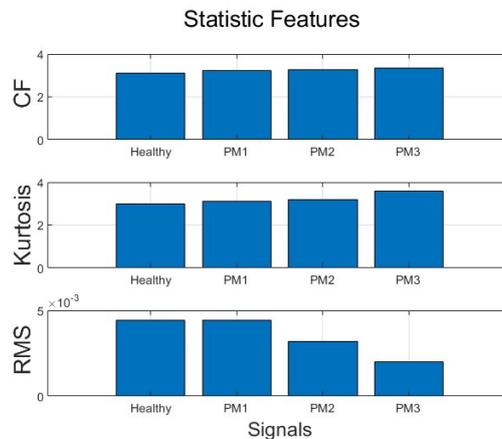


Figure 4.27: Bar plots for (CI_{CF}), ($CI_{Kurtosis}$) and (CI_{RMS}) for healthy vs PM1/PM2/PM3 faulty conditions

Table 4.2: Time domain condition indicators for $v_{MB(front),Y}$ comparing Healthy vs PM1/PM2/PM3 cases

Fault level	CI_APDF	CI_DND	KS_h_value	CI_CF	CI_kurtosis	CI_rms
Healthy	89.161	2.96E-05	0	3.122	2.979	0.0044
Low	89.468	6.87E-05	1	3.238	3.100	0.0044
Medium	111.987	0.000468	1	3.262	3.192	0.0032
High	177.157	0.000498	1	3.327	3.593	0.0019

The numerical values of all the collected condition indicators are summarised in Table 4.2, offering insights into their effectiveness as health indicators under rated wind conditions for different levels of pitch misalignment. An increasing trend is noticeable for all condition indicators, except for the last one (CI_{rms}), which displays a decreasing trend, consistent with the increased peakedness observed in the probability distribution function for the high-severity case (PM3).

4.3. Summary

In this chapter, the analysis method outlined in Chapter 3, has been applied to steady-state conditions, aiming at providing a robust system for fault detection and health condition estimation in wind turbines which present different imbalance types on the rotor. Vibration velocity signals from simulated drivetrain CMS sensor positions were selected as the basis for frequency domain fault detection. The analysis involved signals from both main bearings, the gearbox housing, and the generator coupling. To summarize the results, this work has explored the feasibility of using 1P and 3P peak frequency amplitudes in all the examined CMS measurements, combined with the investigation of the rotor speed signal. Positive results were achieved in differentiating between the different imbalance scenarios.

The extracted features from the selected signals displayed a consistent proportionally increasing trend as the fault severity increased for the same wind speed. The averaged amplitude values, extracted from the signal windows, were presented in 3D plots and quantified through relative change histograms, measured against healthy conditions. However, limitations were identified in detecting 3P peak amplitudes for yaw misalignment, which proved insufficient for cut-in wind speed and low-severity misalignments.

Signals correlation analysis revealed that the rotor speed signal is the most effective signal for mass imbalance detection, allowing for clear clustering between fault levels at all wind speeds. Pitch misalignment detection proved robust across all signals, with distinct clustering observed for each simulated input combination and fault level, supported by high correlations between signal pairs, approaching unity. Similarly, yaw misalignment showed a correlation approaching unity and clear clustering for higher fault severity levels. Moreover, the median value of peak amplitude distributions demonstrated potential for fault detection and system health evaluation.

As a conclusion, the results underscore that when investigating 1P frequency the selected CMS signals are robust for pitch misalignment detection but insufficient for mass imbalance, where rotor speed signals emerged as the leading indicator. On the other hand, the same CMS vibration measurements proved to be a valuable tool for the detection of yaw misalignment when looking at the 3P frequencies, although with few limitations. Additionally, time-domain analysis of the main bearings signals demonstrated promising results for assessing health conditions in the particular case of pitch misalignment for wind speeds falling into the rated power region, through the evaluation of the presented statistical condition indicators.

5

Results: Turbulent Conditions Analysis

In this chapter, the results of the analysis conducted under turbulent wind inflow conditions are presented. The primary objective is to validate the fault detection methodology introduced in the previous chapter, which was applied under steady-state conditions, and to compare the outcomes from both scenarios. This analysis is performed under realistic environmental conditions to assess whether the same trends and results hold. It is important to note that for all modeled imbalance scenarios and fault severities, the inflow conditions now incorporate two turbulence seeds, while the number of analyzed wind speeds has been reduced from six to four, resulting in a total of 80 simulations.

Section 5.1 starts by evaluating the sensitivity of the selected frequency domain features to variations in the number of turbulence seeds. Following this, Section 5.2 presents the frequency domain analysis, which is divided into three Subsections: 5.2.1, focuses on the averaged peak amplitudes across signal windows, 5.2.2 compares the correlation coefficients results between signal pairs of the two simulation environments, 5.2.3, examines the full distribution of extracted amplitudes for each signal. This is done using the methods outlined in Chapter 3, with a direct comparison with steady-state results. Lastly, Section 5.3 discusses the time-domain analysis and presents the results of the selected condition indicators. A summary of the key findings of the turbulent inflow analysis is provided in Section 5.4.

5.1. Number of Seeds Sensitivity Analysis

Before starting with the turbulent condition analysis, this section assesses the sensitivity of the analyzed peak amplitudes to the number of seeds used. to validate the decision to reduce the number of turbulence seeds used in the analysis compared to the reference study conducted by Mehlan et al. [23]. Specifically, a faulty case from the simulation dataset was selected for analysis using six different turbulence seeds. In the reference study, six seeds were used for the simulated turbulent conditions. The goal here is to assess how reliable the results of the analysis are when reducing the number of seeds.

The selected case is the high-severity pitch misalignment fault (PM3) in rated power conditions ($U_4 = 12\text{m/s}$). This case was chosen since pitch misalignment in the full load control region of the wind turbine has been identified as one of the most critical fault scenarios in steady-state vibration analyses. Therefore, this case is expected to show higher sensitivity to variations in turbulent external conditions if there are any.

This analysis examined how the peaks of the rotor's rotational frequency (1P) and its third harmonic (3P) behave when averaged across different numbers of seeds. The results are presented in Figure 5.1, where the peak amplitudes of the four selected signals are plotted for the six different turbulent input conditions.

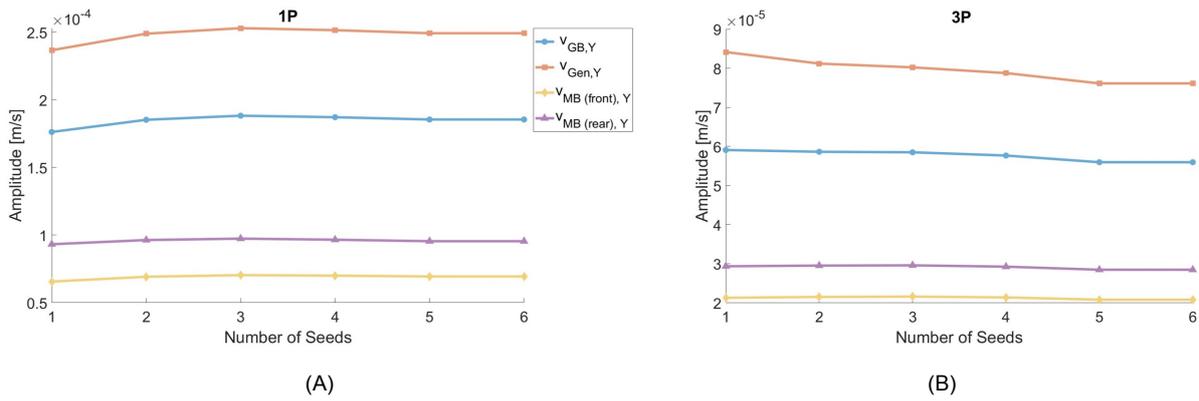


Figure 5.1: (A) 1P frequency amplitudes for CMS signals, (B) 3P frequency amplitudes for CMS signals

Figure 5.1 shows that the generator velocity vibration signal has the highest amplitudes in 1P and 3P, followed by the gearbox housing and rear main bearing signals, and lastly the lowest magnitudes in the front main bearing signal. An increase in peak amplitudes is evident when moving from the wind turbine rotor side toward the end of the transmission system.

Additionally, for the same vibration signal, the peak magnitudes do not significantly change when the number of input turbulence seeds increases. This shows that when performing ensemble averaging, the more seeds we consider, the more the turbulence variability is smoothed out. However, given the small magnitude of the studied values, a more detailed analysis is required. To achieve this, a further assessment was conducted by calculating the relative percentage change in peak amplitudes with respect to one turbulence seed. The aim is to assess the sensitivity of the extracted feature compared to the number of seeds. These results are presented in Figure 5.2.

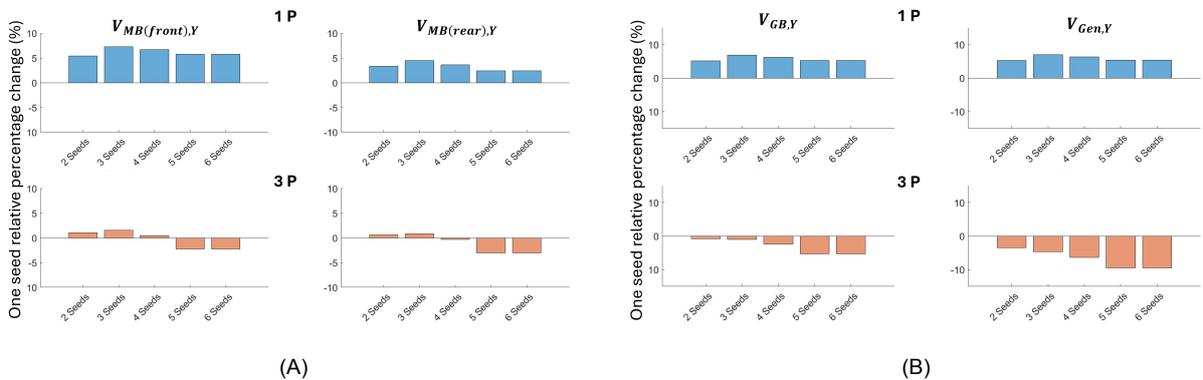


Figure 5.2: 1P and 3P amplitudes for PM3 at U4: relative percentage difference with respect to the one seed input turbulent conditions: (A) Main bearing velocities side- side direction, front and rear (B) Gearbox housing and generator velocities for side-side direction,

In the Figure, it can be appreciated that compared to the reference case with one turbulence seed, the 1P peak amplitudes (blue bars) show an average relative percentage increase of 5% across all signals for seeds equal to or greater than 2. For the 3P peaks (orange bars), $V_{MB(front),Y}$ and $V_{MB(rear),Y}$ display low variations ranging between 0% and 3%. However, the gearbox housing and generator signals show a decreasing trend, with the first one dropping by up to 5% and the second one by up to 9%. These percentage changes remain constant in both the five-seed and six-seed cases for all the signals, indicating a convergence towards a stable limit that ensures the validity of the results. Here, it can be noticed with stronger clarity that the high variability introduced by the turbulence seed is smoothed out with a higher number of seeds. Therefore, this analysis confirms that the reduced number of turbulence seeds does not compromise the accuracy or robustness of the results presented in this chapter.

5.2. Frequency Domain Analysis

In this case, due to the presence of turbulence, it was particularly important to apply the resampling COT algorithm to avoid shifts in peak positions within the frequency spectrum when analyzing wind speeds across different regions of the power curve. As discussed in Section 2.7.1, applying the resampling procedure involves switching from the frequency domain to the order domain. This ensures that peak positions remain fixed relative to the rotational speed of the main shaft, regardless of the wind speed simulated in the inflow conditions. The algorithm was applied to all signals prior to conducting the frequency domain analysis.

In the plots presented in Figure 5.3 just one representative signal for each imbalance case is presented, showing overlapping distributions for the four different inflow conditions (wind speeds: U1, U2, U3, and U4) once COT is performed. The example signal is the vibration velocity measured at the gearbox housing for a low-severity fault case. As expected, and in line with findings from other studies [58], Figure 5.3 demonstrates that the 1P and 3P frequency peaks overlap at the same order positions, with increasing amplitudes as the wind speed increases.

In the three subplots, the 1P and 3P frequencies are clearly identifiable in their respective imbalance cases. The higher magnitude of these frequencies overshadows other system frequency modes, making them less visible in the reported turbulence case.

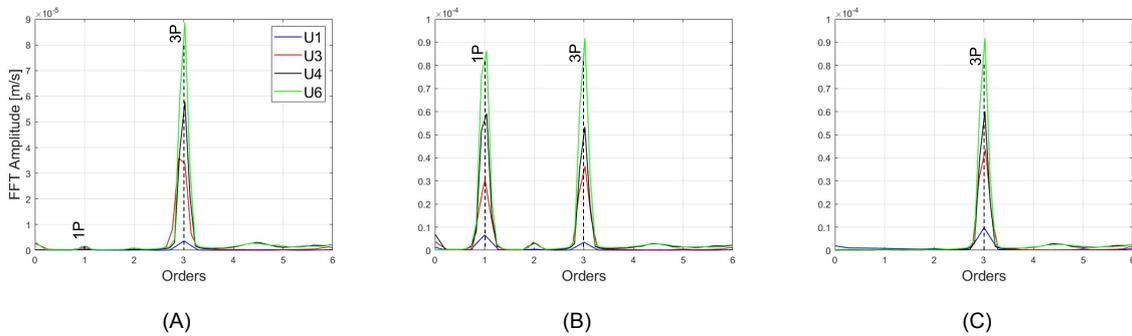


Figure 5.3: FFT plot analysis of $v_{GB,Y}$ signal resampled with COT algorithm for different wind speeds: (A) MI1, (B) PM1, (C) YM1

Additionally, it can be observed that under turbulent inflow conditions, the system experiences stronger dynamic effects, which contribute to higher amplitudes in the 3P harmonic frequency amplitude compared to what was observed in steady-state conditions. This effect is also evident in the healthy condition, reducing the gap between healthy and faulty conditions, as shown in Figure 5.4.

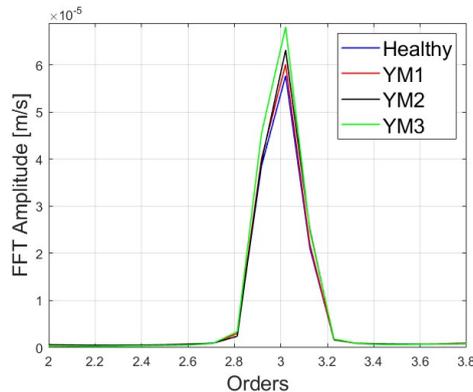


Figure 5.4: FFT plot Healthy vs YM1/YM2/YM3 fault conditions for $v_{GB,Y}$ signal in order domain

5.2.1. Averaged Peak Amplitudes Analysis

In this section, the analysis performed under steady-state conditions, as detailed in Section 4.1), is extended to turbulent inflow conditions. To highlight the impact of turbulence, shaded paths representing the steady-state trends have been overlaid on the 3D plot representations. The results for the gearbox housing vibration velocity signal are illustrated in Figure 5.5, where the averaged peak amplitudes from the turbulent inflow simulations are compared to the previously observed steady-state results shown in Figure 4.2. The whole set of 3D plots for all the analyzed signals is reported in Appendix E (Section E.1).

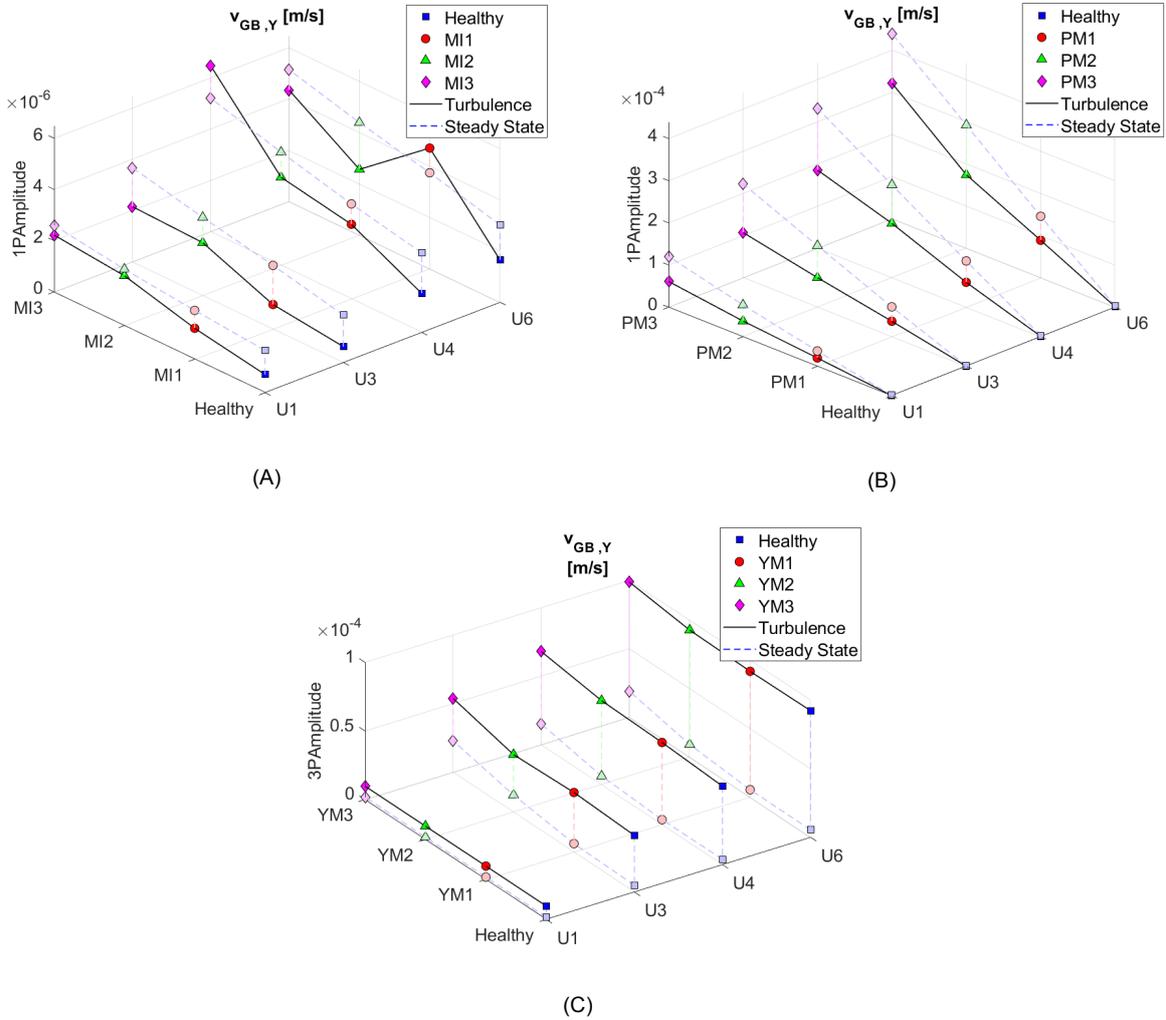


Figure 5.5: Averaged peak amplitudes comparison of $v_{GB,y}$ signal under varying health states and turbulent inflow conditions. (A) MI - 1P, (B) PM - 1P, and (C) YM - 3P

In the case of mass imbalance, shown in Figure 5.5 (A), the randomness introduced by the turbulence still allows for the detection of the imbalance, but the trend is no longer clearly monotonically increasing. This is due to the low sensitivity range of the 1P peak amplitudes. A similar behavior was observed in other CMS velocity vibration signals, with the exception of the front main bearing, where the trend still shows to be upward increasing. In contrast, the rotor speed 1P amplitudes follow a trend similar to the steady-state one but shifted downward along the z-axis due to a lower overall amplitude magnitude (Figure 5.6 (A)).

For pitch misalignment, shown in Figure 5.5 (B), the upward monotonic trend is preserved for both increasing wind speeds and fault severity levels, as previously observed in steady-state conditions. However, it is noticeable that the averaged amplitudes are lower in the case of transient conditions, and this applies to all CMS signals, as well as to the rotor speed (5.6 (B)).

Conversely, yaw misalignment, shown in Figure 5.5 (C), exhibits higher magnitudes, as previously observed in Section 5.2. Due to the stronger healthy system response at the 3P frequency, the overall trend shows a flattening effect, significantly reducing the steepness that was observed under steady-state conditions.

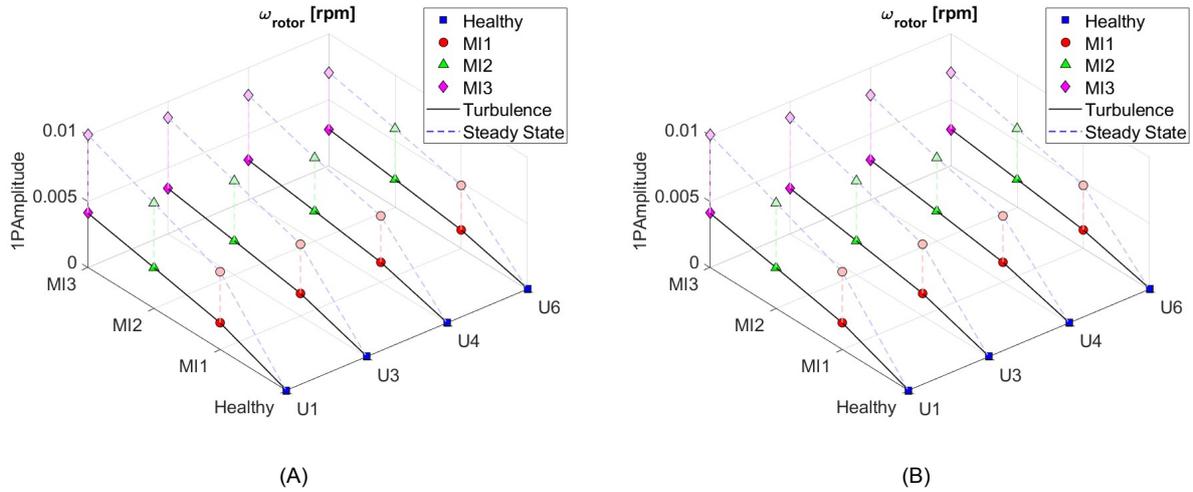


Figure 5.6: Averaged 1P peak amplitudes comparison of ω_{rotor} signal: (A) MI, (B) PM

To visually compare the relative change detection of the signals shown in the 3D plot with those calculated under steady-state conditions, histograms are provided. Error bars are used to highlight the differences between the two sets of data. In this case, the error bars are centered on the turbulent relative change values from the histograms and extend in both directions—above and below. The error bars were calculated by estimating the absolute *difference* between steady-state and turbulent relative change values. Therefore in Figures 5.7, 5.8 and 5.9, they are going to be referenced in the legend as the steady-state error (SS Error). In the graphical representation, an upper error (positive difference) occurs when the steady-state relative change value is higher than the turbulence value, causing the error bar to extend above the turbulent histogram bar. Conversely, a lower error (negative difference) indicates that the turbulent value exceeds the steady-state value, so the error bar extends below the turbulent case bar.

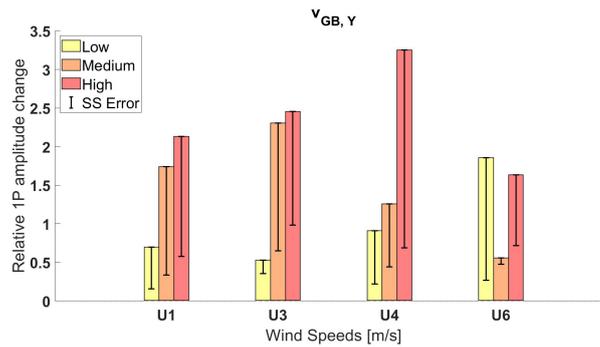


Figure 5.7: MI - 1P: Normalised relative change of the averaged amplitudes, comparing turbulent and steady-state conditions through SS error bars

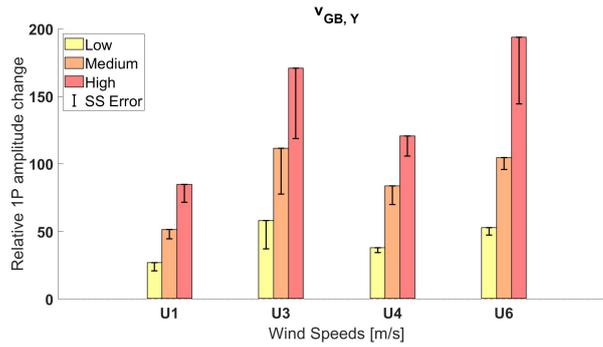


Figure 5.8: PM - 1P: Normalised relative change of the averaged amplitudes, comparing turbulent and steady-state conditions through SS error bars

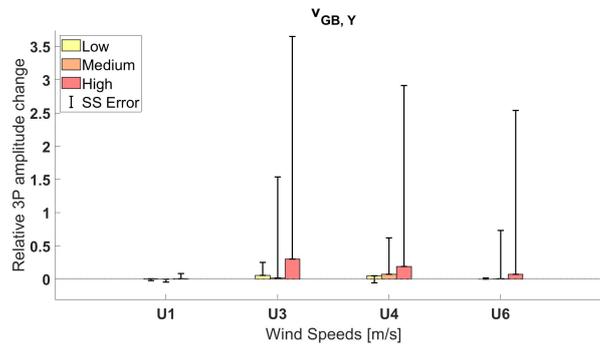


Figure 5.9: YM - 3P: Normalised relative change of the averaged amplitudes, comparing turbulent and steady-state conditions through SS error bars

In Figure 5.7, the mass imbalance case is represented. Despite the randomness introduced by turbulence, discussed earlier, the relative change in peak amplitudes compared to healthy conditions shows higher ranges and improved detection for the gearbox vibration velocity sensor under turbulent conditions. The error bars show that the steady-state results exhibited a maximum increase of up to twice the magnitude compared to healthy conditions for this signal. This is indicated by the lower ends of the error bars, which stop at a maximum value of 1 on the y-axis. A y-axis value of 1 signifies that the relative change detection value for steady-state conditions is one time greater than its respective healthy baseline. A similar pattern is observed for the front main bearing velocity vibration signal, where the overall relative change detection magnitude decreases with increasing wind speed compared to steady-state values. In contrast, other CMS sensor measurements show more significant fluctuations, leading to a loss of clear fault detection trends and worsening health condition indicators. The corresponding plots can be found in Appendix E (Section E.2). These results highlight the limitations of CMS sensitivity for robust mass imbalance detection, as the aerodynamic effects of varying inflow conditions tend to overshadow the imbalance signals. However, the rotor speed signal remains the most effective indicator for mass imbalance detection due to its higher sensitivity compared to the other analyzed signals (Section E.2, Appendix E).

Similarly, for pitch misalignment, a general downward shift is observed compared to steady-state conditions in the corresponding 3D plot visualizations. As shown in Figure 5.8, despite this shift, the steepness of the trend increases for the gearbox housing vibration measurements, resulting in higher relative change detection under turbulent conditions. In contrast, other CMS signals exhibit a decrease in the detected change between peak amplitudes of faulty and healthy conditions across all wind speeds, except for cut-in conditions, where the trends remain more stable. The corresponding plots are provided in Appendix E (Section E.2). Overall, the analysis of 1P peak amplitudes from CMS sensor measurements continues to demonstrate robustness for detecting pitch misalignment faults, support-

ing effective health condition monitoring of the system. In the case of misalignment of the yaw, Figure 5.9 illustrates the effects of the upward shift in peak amplitudes observed in the previous plots in Figure 5.5 (C). The flattening of the amplitude trend results in a lower relative change detection sensitivity compared to healthy conditions and lies within a consistently reduced magnitude range compared to steady-state conditions. To clarify the comparison, the detected relative change in magnitude for high-severity faults now shows an increase of 10% to 30% compared to healthy conditions, corresponding to a 0.1–0.3 range on the y-axis in the graph. Previously, this increase was three to four times higher, as indicated by the longer error bars above the histogram bars. Despite the lower sensitivity under turbulent conditions, a closer examination still confirms the effectiveness of 3P peak amplitudes for detecting yaw misalignment, albeit with certain limitations. The difficulty in detecting imbalance at cut-in wind speed, which was also observed under steady-state conditions, persists in this case. Furthermore, from the histograms, it can be observed that in the wind speed regions below rated conditions, the relative change in frequency amplitudes no longer exhibits an upward trend with increasing fault severity, reducing the accuracy of health monitoring. In contrast, beyond the rated wind speed, a clearer trend emerges, with change detection proving robust for medium to high fault severity levels. Similar patterns are evident across all other CMS signals, as detailed in Section E.2 of Appendix E.

5.2.2. Correlation Analysis

Estimating the correlation between peak amplitudes extracted from different signals allows to determine whether the imbalance exerts a uniform influence across various measurement locations in the drivetrain, or if one sensor location demonstrates greater sensitivity to fault detection. The bar plots in Figures 5.10, 5.11, and 5.12, illustrate the differences in correlation coefficient values between steady state and turbulent conditions, calculated from correlation matrices similar to those presented in the steady-state analysis in Section 4.5. The x-axes display the different signal pairings from the correlation matrices, starting with gearbox housing side-side vibration velocity, paired first with the generator sensor, followed by front and rear main bearings, and finally, rotor speed, continuing with other combinations.

It can be observed that the presence of turbulence induces fluctuations in the correlation coefficients between CMS signals, while still preserving an overall coherent output across all three imbalance cases with respect to steady-state conditions. For mass imbalance, shown in Figure 5.10, the coefficients range between 0.65 and 0.97 for CMS signal pairs, highlighting a correlation approaching unity. However, as discussed in Chapter 4, the features extracted from CMS signals for this specific imbalance case tend to overlap, showing little distinction from healthy condition data points due to the low magnitude of the frequency peak amplitudes. On the other hand, a notable impact on those signal pairs compared to the rotor speed signal is visible. These coefficients are closer to zero and this indicates a horizontal clustering trend in the scatter plots of amplitude points, which aligns with the findings from the steady-state analysis shown in Figures 4.8 and 4.9. This result confirms the dominant influence of the rotor speed signal in detecting mass imbalance faults, proving enhanced clustering potential.

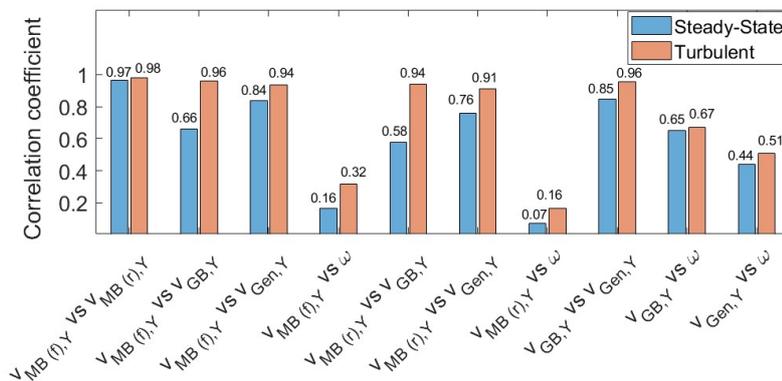


Figure 5.10: Correlation coefficients values for mass imbalance scenario signal pairs comparing steady-state and turbulent input conditions

In general, for mass imbalance scenario the same trends described in Section 4.1.4 have been observed in this second set of simulations referring to turbulent conditions. Although the scatter plots of peak amplitude distributions exhibit slightly more noise due to turbulence, the overall results remain consistent and validated under these turbulent conditions. Therefore, the graphic presentations of these results are omitted from this section, examples can be found in Appendix E, Section E.3.

Similarly, for pitch misalignment, correlation coefficients approaching unity are maintained across all the signal pairs, with a slight reduction in their values around 10% as shown in Figure 5.11. The selected CMS signals under analysis still demonstrate potential for fault detection, as was observed in Figures 4.10 and 4.11. However, the scatter plots resulted in being more affected by the noise introduced by turbulence leading to the presence of more spread outliers not allowing for a clear visual clustering through full peak amplitudes data set distribution. As such, the scatter plots between signal pairs for this case are also omitted and an example is reported in Section E.3, Appendix E. To investigate if the possibility of reliable health monitoring is preserved follow-up analyses are reported in Section 5.2.3.

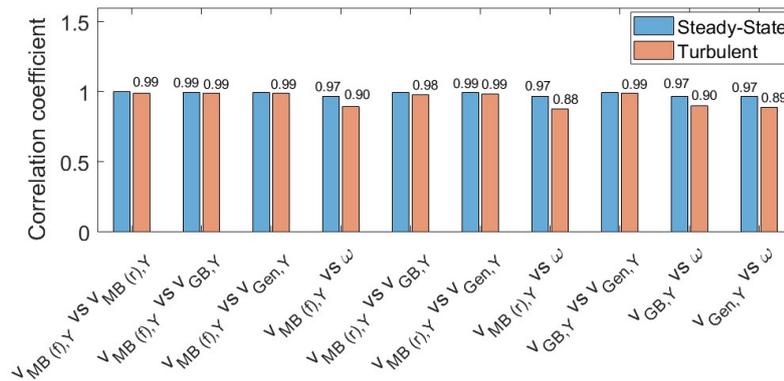


Figure 5.11: Correlation coefficients values for pitch misalignment scenario signal pairs comparing steady-state and turbulent input conditions

For yaw misalignment, the correlation matrix reveals nearly uniform and consistently high values, indicating a strong positive correlation approaching unity, as illustrated in the bar plot comparison in Figures 5.12. The scatter plot analysis of 3P amplitude distributions further supports these findings, demonstrating the strong fault detection and clustering at higher fault levels. Also in this case the results are consistent with those obtained in the steady-state analysis, as shown in Figures 4.12 and 4.13, therefore just one example is reported in Section E.3 of Appendix E. It is worth it to note that also for this new analysis, the rotor speed signal has not been taken into account for the yaw misalignment case due to the absence of any detectable difference compared to healthy conditions.

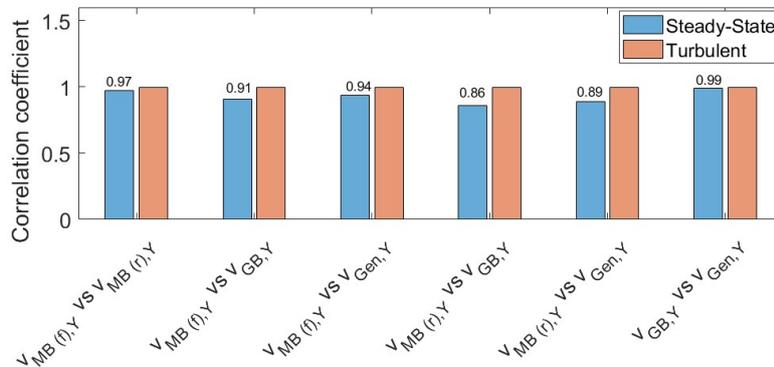


Figure 5.12: Correlation coefficients values for yaw misalignment scenario signal pairs comparing steady-state and turbulent input conditions

5.2.3. Box Plots and Median Values Multi-comparison Analysis

Following the correlation analysis, the same steps used to study the extracted feature distributions for the three imbalance scenarios under steady-state conditions were consistently applied to the second dataset of simulations involving turbulence.

Box plot distribution analysis is shown in Figures 5.13, 5.14, and 5.15. The alignment of the faulty distributions relative to their corresponding healthy state distributions remains consistent with the trends observed under steady-state conditions. However, differences in the data distribution can be noted, particularly in the skewness, as reflected by the asymmetry of the whiskers extending from the boxplots in most cases. Additionally, an increase in the number of outliers spread farther from the boxplot positions, is observed in the 1P distributions for the CMS signals of mass imbalance and pitch misalignment. This is attributed to the added randomness and impulsiveness introduced by turbulence. It is important to recall that outliers are identified as values lying outside the interquartile range (IQR), and they represent 0-5% of the data points in the distribution, reaching up to 10%-15% in higher fault severity cases. This indicates that the system becomes more sensitive to severe faults when turbulence induces additional excitation.

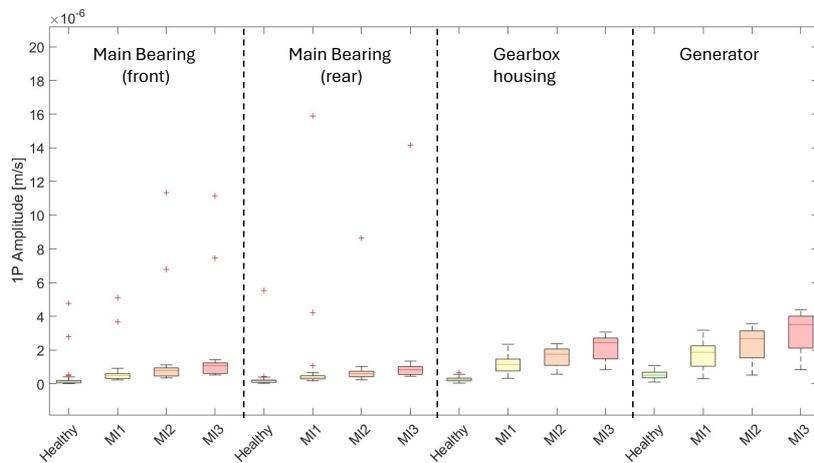


Figure 5.13: MI: 1P amplitudes distribution and variance analysis aggregated for all wind speeds

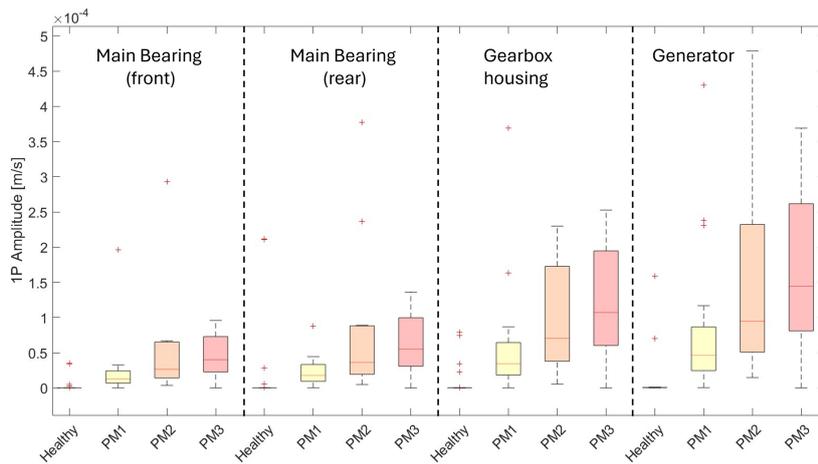


Figure 5.14: PM: 1P amplitudes distribution and variance analysis aggregated for all wind speeds

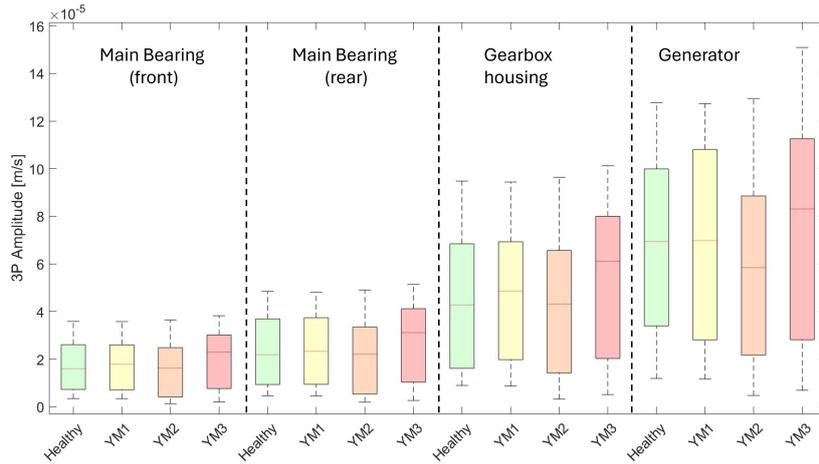


Figure 5.15: YM: 3P amplitudes distribution and variance analysis aggregated for all wind speeds

Following the examination of the distributions, a statistical analysis based on changes in the median values is conducted using the Mann-Whitney U test, as described in Chapter 3. The comparison of peak amplitude distributions between healthy conditions and the three fault severity levels is performed for each wind speed individually. The results confirm the trends previously observed under steady-state conditions. The following figures present results for the $v_{GB,Y}$ signal, as representative examples of the different CMS signals, along with the ω_{rotor} signal across wind speeds U_1 , U_3 , U_4 , and U_6 . In Figure 5.16, a comparison between healthy conditions and the MI1/MI2/MI3 fault levels is shown for all the specified wind speeds. Figure 5.17 illustrates the comparison between mass imbalance and pitch misalignment fault severities against their respective healthy conditions, following the approach used in Section 4.1.6. Lastly, Figure 5.18 applies the same analysis to the yaw misalignment scenario.”

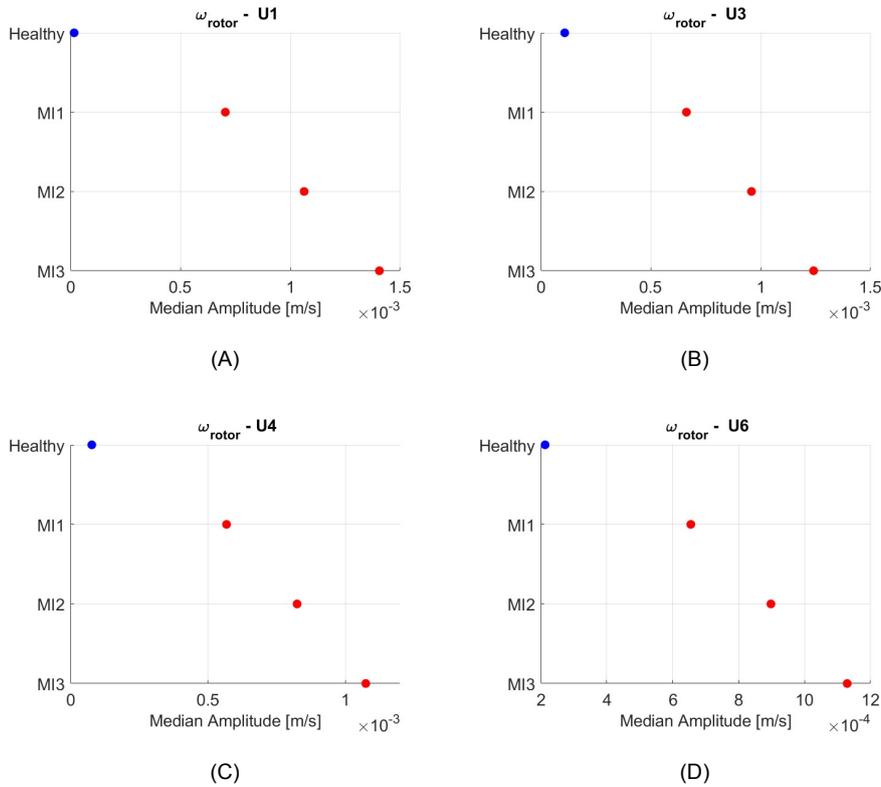


Figure 5.16: Healthy condition vs MI1/MI2/MI3 - ω_{rotor} amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test)

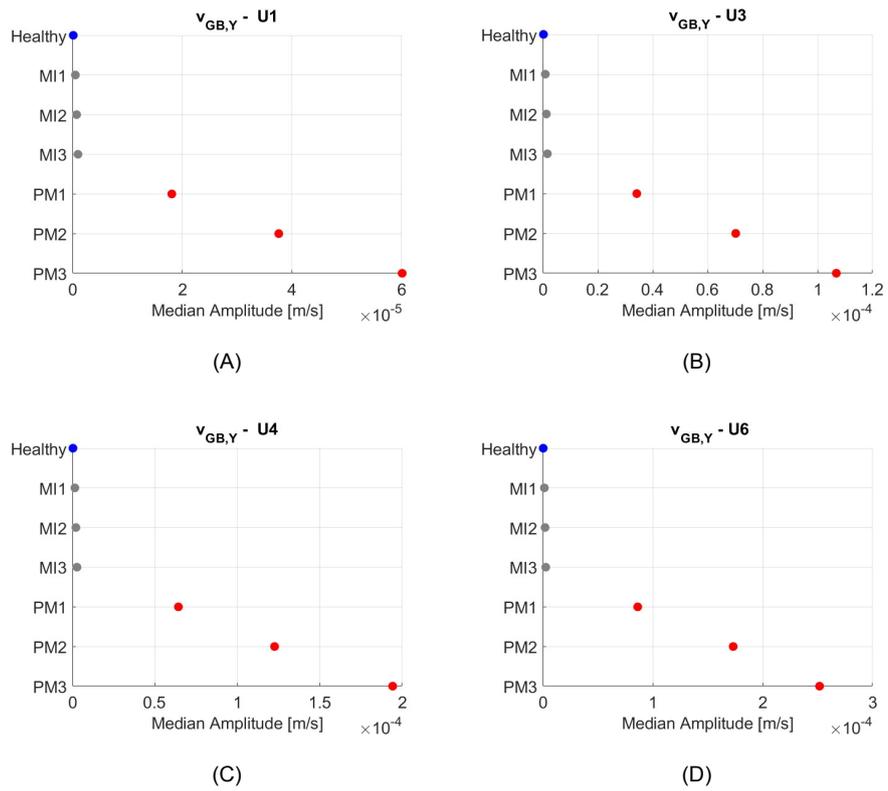


Figure 5.17: Healthy condition vs MI1/MI2/MI3 and PM1/PM2/PM3 - $v_{GB,\gamma}$ amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test) - wind speeds: U_1, U_3, U_4, U_6

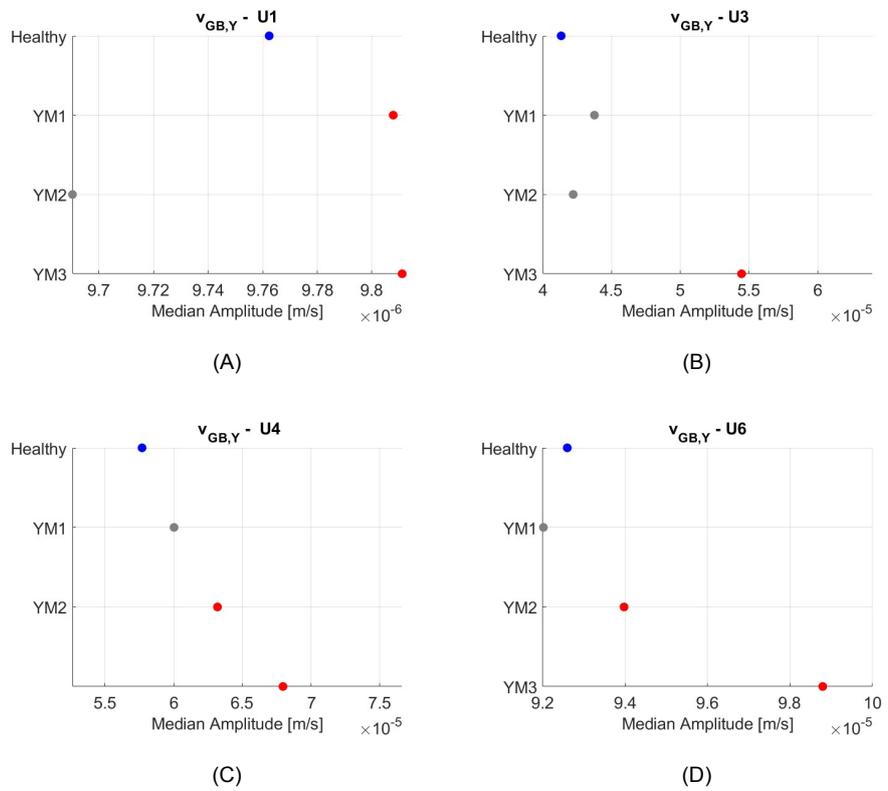


Figure 5.18: Healthy condition vs YM1/YM2/YM3 - $v_{GB,\gamma}$ amplitude distribution median values comparison analysis after significant statistic difference detection (Mann-Whitney U test), wind speeds: U_1, U_3, U_4, U_6

The presented graphical results are obtained through the same procedure described in Section 4.1.5. Specifically, the one shown in Figures 5.17 and 5.18, are consistent across all other CMS signals. For pitch misalignment, robust detection, and health condition indication are consistently observed across all wind speeds and power curve regions for all CMS signals. Median values for fault conditions are one to two orders of magnitude higher than those for healthy conditions, with distinct clustering between different health states. In comparison, mass imbalance shows lower sensibility in detecting 1P peak amplitudes, as the amplitude mismatch for mass imbalance falls within a range similar to that of healthy conditions. This leads to higher uncertainty and an increased likelihood of false alarms, as highlighted in Section 4.1.6. Similarly, the detection reliability achieved for pitch misalignment across all CMS signals was successfully replicated for mass imbalance by applying the same analysis to the rotor speed signal.

For yaw misalignment, the proposed method does not provide robust health monitoring at cut-in wind speeds, as noted in previous analyses. However, robust detection is achievable for medium to high fault severity at wind speeds within the rated power region. Below rated wind speeds, detection remains reliable only for higher fault severity, with reduced sensitivity to low and medium fault levels.

5.3. Time Domain

As outlined in Chapter 3, the second part of the analysis focuses on the time domain, specifically examining the filtered drivetrain torsional mode to assess its response to faulty system dynamics. It was detailed in Section 3.5.2, that the central frequency of interest from the drivetrain model is 4.21 Hz, and it is isolated using a band-pass filter. This section presents the results of the five chosen condition indicators, defined in Equations 2.17, 2.18, 2.19, 2.20, and 2.21, and evaluates their response under turbulent inflow conditions. It is important to highlight that, as discussed in Chapter 4, the observed result trends consistently hold for wind speeds within rated power conditions. However, velocities in the partial control region, i.e., below-rated wind speed, did not exhibit a clear trend suitable for fault detection. Additionally, pitch misalignment was identified as the only detectable scenario via torsional mode analysis, while mass imbalance and yaw misalignment showed no significant difference between healthy and faulty conditions. Therefore, this section reports accordingly, simulation results for varying pitch misalignment degrees which apply under rated power conditions.

Similar to the findings under steady-state conditions, the vibration signals from both the front and rear main bearings conform to a normal distribution in healthy conditions. This validates the assumptions for applying the condition indicators CI_{APDF} and CI_{DND} for fault detection analysis. As the results are consistent across both bearings, only the front main bearing signal is presented in this section, with the same trends applicable to the rear bearing. Figure 5.19 illustrates the normal probability distributions and compares healthy conditions to the three different levels of pitch misalignment.

However, unlike the trends observed in Chapter 4, a different behavior emerges here. Specifically, as the fault level increases, the distribution's tails widen, driven by the higher impulsivity in the signal caused by turbulence. The presence of turbulence leads to more intense spikes being registered in the vibration signals. This results in an inverse trend compared to the steady-state case, where the peak values of the PDF decrease as fault severity rises. This happens because the area under the PDF curve of a normally distributed signal must always equal 1, and the distribution broadens as a result of these larger spikes.

In Figure 5.19, it can be seen that the PDF peak for the healthy condition is higher and the distribution is more concentrated around zero, in a range similar to the steady-state case. This indicates that under turbulent inflow, the system remains relatively stable in healthy conditions, with only a minor broadening in the signal's amplitude range. However, with increasing severity of pitch misalignment, the tails of the distribution broaden and the peaks lower, indicating increased signal variance due to the combined effects of misalignment and turbulence.

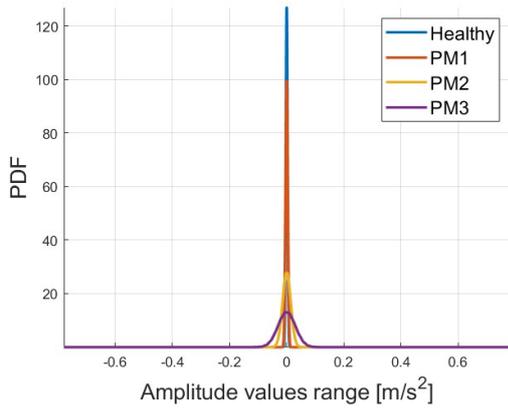


Figure 5.19: PDF for $a_{MB(front),Y}$ signal comparing healthy conditions with multiple pitch misalignment degrees ($U_4 = 12$ m/s)

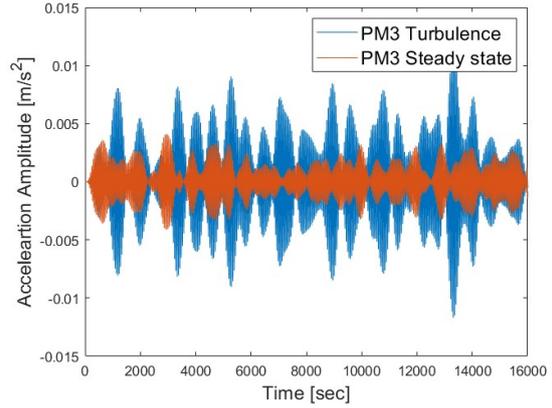


Figure 5.20: Time domain torsional mode comparison of PM3 signal in steady-state compared to turbulent inflow conditions

This higher variance is further evidenced in Figure 5.21, particularly in the fourth histogram fitting subplot (PM3 fault severity case), where the impact of higher pitch misalignment is clearly visible. As the average amplitude range of the signal narrows with increased pitch misalignment, the effect of turbulence becomes more pronounced. This observation aligns with the findings in Section 4.2, where the absence of turbulence reduced signal variability. The large outliers caused by turbulence significantly widen the tails of the PDF, as these spikes introduce values far from the mean. The system's lower initial amplitude range amplifies the impact of each spike, resulting in a more dramatic stretching of the PDF tails. The difference in time domain filtered torsional mode signals for the high severity pitch misalignment (PM3) is compared between steady-state and turbulent conditions in Figure 5.20 to highlight the effects described. As a result, it can be concluded that in real turbulent inflow conditions the higher the pitch misalignment severity the lower the PDF peak amplitude, implying therefore a decreasing trend.

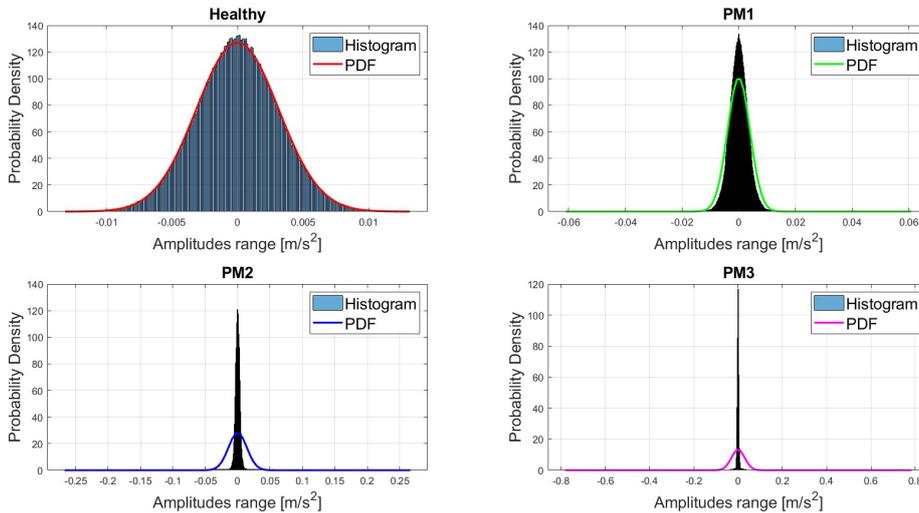


Figure 5.21: Histogram data fitting of the simulated dataset with respect to theoretical normal distribution. Study case: pitch misalignment imbalance for front main bearing acceleration signal

When comparing the quantiles of the ideal normal distribution with the simulated dataset in Figure 5.22, a similar trend to that observed in Figure 4.26 is evident. The data confirms a larger deviation from ideal conditions as misalignment increases, with greater emphasis on deviations in the distribution tails.

This results in a larger area difference, as captured by the CI_{DND} values, increasing the potential of this parameter as a health indicator. Normal distribution fitting is further quantified also in this case through the Kruskal-Wallis test, and numerical results are reported in the fourth column of Table 5.1.

The indicator numerical values for CI_{APDF} and CI_{DND} from the previous analysis are also summarized in Table 5.1. Additionally, the calculated values for CI_{CF} , $CI_{Kurtosis}$, and CI_{RMS} from Equations 2.18, 2.19, and 2.17 are also reported. All these values are presented across the three different misalignment severity levels analyzed. This summary allows for a clearer visualization of potential health monitoring trends.

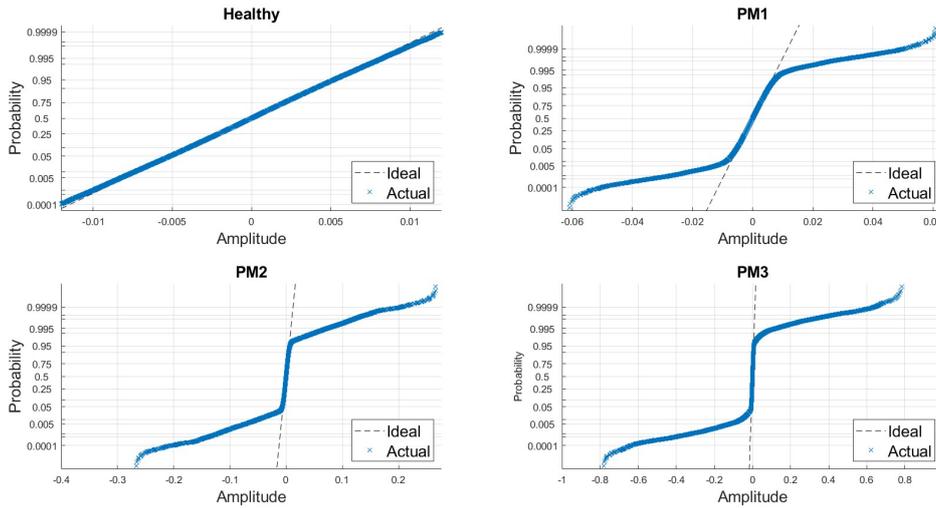


Figure 5.22: NPP for quantile comparison of simulated data with theoretical normal distribution. Healthy conditions for wind speed $U_4 = 12m/s$ compared with multiple degrees of pitch misalignment (PM1, PM2, PM3), for $v_{MB(front),Y}$ signal

Table 5.1: Time domain analysis in turbulent conditions. Condition indicator results are compared between healthy conditions and three pitch misalignment degrees PM1/PM2/PM3 for $v_{MB(front),Y}$ signal

Fault level	CI_apdf	CI_DND	KS_h_value	CI_CF	CI_kurtosis	CI_rms
Healthy	126.92	3.37E-05	0	3.180	3.038	0.0031
Low	99.74	0.00067	1	3.25	3.21	0.0034
Medium	27.82	0.00081	1	3.29	3.22	0.0065
High	13.14	0.0202	1	3.327	3.24	0.0106

As seen in the table, all the time domain condition indicators prove to be effective tools for health monitoring in the specific case of pitch misalignment under rated power conditions. They demonstrate higher sensitivity and robustness in turbulent conditions, particularly in terms of amplitude detection of the probability density function, area difference estimations between normal distributions and actual data, and root-mean-square values. Although crest factor and kurtosis also exhibit increasing trends, they display less sensitivity, with values closer across different health conditions, therefore, they are not considered robust indicators for system health discrimination in this study case.

5.4. Results Findings Summary

In conclusion, the key findings of this chapter are summarized as follows. The proposed analysis methods demonstrated robustness, and the detection trends were validated even with turbulence included in the wind inflow conditions. Starting with the frequency domain analysis, the sensitivity of the extracted features to the number of turbulence seeds was first assessed, concluding that the reduction to two seeds did not significantly affect the drivetrain's measured signals. The analysis of 1P peak amplitudes

confirmed the method's robustness for detecting mass imbalance and pitch misalignment faults. It highlighted the rotor speed signal as the most effective for the detection of mass imbalance, while all the considered CMS sensor signals are sensible to pitch misalignment showing high detection reliability. For the analysis of the 3P peak amplitudes, yaw misalignment detection capability was reduced under real turbulent conditions because of the system's higher response in the same frequency range under healthy conditions. However, medium to high degrees of misalignment were still accurately detected, with limitations dependent on the region of the power curve corresponding to the current wind speed.

Overall, the analysis of the median values using the Mann-Whitney U test has been confirmed as a valuable tool for identifying different operating states of a wind turbine rotor by calculating statistically significant differences between groups. As a result, this thesis proposes the median values of peak amplitude distributions at 1P and 3P frequencies as key features for a data fusion framework, specifically at the decision fusion level. This approach lays the groundwork for implementing a decision tree to detect and diagnose the three investigated imbalance scenarios, as well as monitor the health state.

Lastly, the time domain analysis of the drivetrain's torsional mode proved to be a powerful tool for detecting pitch misalignment and assessing health conditions, using the main bearing acceleration signals under rated power conditions through three key condition indicators: CI_{APDF} , CI_{DND} , and CI_{RMS} .

6

Discussion, Recommendations, and Limitations

In the previous Chapters (4 and 5), the features extracted from CMS signals under steady-state and turbulent inflow conditions were thoroughly examined, with a focus on detecting and classifying three types of wind turbine rotor imbalances. This Chapter aims to synthesize the findings and provide a comprehensive discussion of how these feature data can be fused into a systematic fault detection framework. Specifically, Section 6.1 outlines a decision tree model for fault classification, which lays the groundwork for future implementation of detection algorithms.

Moreover, Section 6.2 identifies potential health monitoring indicators and their applicability to the three main imbalance scenarios: mass imbalance, pitch misalignment, and yaw misalignment. These health indicators are evaluated on the basis of their ability to track the fault level evolution and their diagnostic capabilities, under varying wind conditions, using both frequency and time-domain analyses. Particular attention is paid to the development of signal-based thresholds and decision-making criteria that can be adapted to different operating regimes of the wind turbine.

Finally, Section 6.3 outlines the key limitations of the current methodology and provides specific recommendations for future research and further analysis.

6.1. Decision Tree for Fault Detection and Diagnosis

The work presented in this thesis aims to contribute to the development of more efficient predictive maintenance techniques by proposing novel approaches to analyze CMS signals. In the context of wind turbines, rotor imbalances are critical to address to prevent severe failures and minimize production losses due to aerodynamic misalignment. This study focuses on three primary types of imbalance: mass imbalance, pitch misalignment, and yaw misalignment, each evaluated at multiple wind speeds, under both steady and turbulent conditions, and for three different severity levels. The results of various combinations of these fault conditions have been analyzed and compared with healthy system conditions to extract robust features for fault detection and classification.

This section summarizes the key findings from the previous two chapters and integrates them into a unified fault detection and classification flowchart. Building on the decision fusion level approach discussed in Section 2.9, individual signals are processed and analyzed separately. The measurement features extracted from each source are then combined to diagnose rotor imbalance scenarios. The resulting flowchart is structured as a binary decision tree, where each node addresses a specific diagnostic task, progressively guiding the detection and classification of various fault types.

Let's review the steps required to implement the flowchart shown in Figure 6.1. The process begins with the selection of signals from the OpenFAST and SIMPACK drivetrain models, which are representative of the SCADA and CMS signals typically collected by modern wind turbines. These signals are used to capture the dynamic response of the system. For the frequency domain analysis, velocity vibration

measurements are chosen, specifically focusing on the side-to-side direction (y-axis). These signals are denoted as: $v_{MB_{front}}$, $v_{MB_{rear}}$, v_{GB} , and v_{GEN} . Additionally, the rotor speed signal, ω_{rotor} , is also included in the analysis. The vibration signals, simulated for 4,000 seconds each and sampled at 200 Hz, are first processed by removing the transient phase (the first 300 seconds). The remaining 3,600 seconds are then resampled in the order domain to mitigate the effects of peak smearing and shifting inherent in the variable rotational speed of the turbine. A Hann window is applied to prevent aliasing and spectral leakage, and the signals are divided into 60 windows of 60 seconds each. FFT is performed on each window, and the amplitude peaks at 1P and 3P frequencies are extracted, yielding 60 data points for each frequency and each signal. A similar procedure is followed for the SCADA rotor speed signal, taking into account its lower sampling frequency of 40 Hz. After pre-processing, the median values of the distributions for 1P and 3P frequency amplitudes are calculated for each health condition, serving as the basis for fault presence detection and diagnosis (differentiating between imbalance types). With this foundation, a decision tree is constructed using statistical change detection based on the median amplitude values. The proposed classification tree is shown in Figure 6.1.

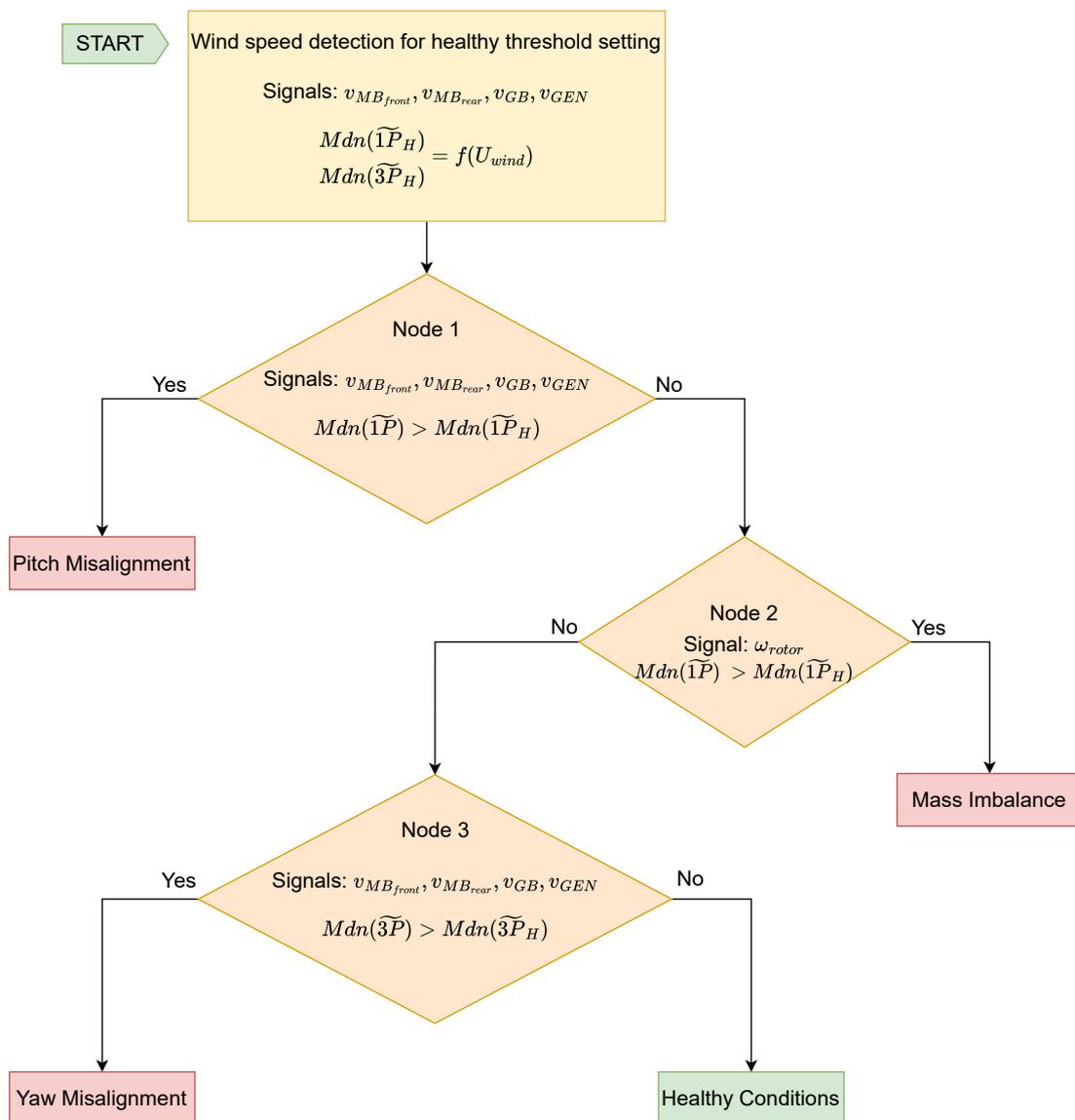


Figure 6.1: Detection and diagnosis decision tree for three different imbalance scenarios under study (MI, PM, YM)

The first step involves determining the healthy thresholds for the median values of the 1P and 3P peak amplitude distributions, $Mdn(1\tilde{P}_H)$ and $Mdn(3\tilde{P}_H)$, as these thresholds depend on environmental factors, specifically they are a function of the current wind speed ($f(U_{wind})$). To accurately differentiate between healthy and faulty conditions, these thresholds must be established through system training based on healthy operational data. This process is repeated for each signal under consideration.

Once the healthy thresholds are established, Node 1 compares the median values of the four CMS signals from different drivetrain positions, $Mdn(1\tilde{P})$, against their corresponding healthy thresholds, $Mdn(1\tilde{P}_H)$. If the median values of all signals exceed their respective thresholds, a pitch misalignment fault is detected. This criterion is based on the strong correlation observed between CMS signals, which consistently reflects the influence of pitch misalignment on 1P frequency amplitude peaks (refer to Section 5.2.2). This multisignal approach enhances the robustness of fault detection and minimizes the risk of false alarms.

If pitch misalignment is not detected, Node 2 evaluates the rotor speed signal, ω_{rotor} , by comparing its $Mdn(1\tilde{P})$ value against the corresponding healthy threshold $Mdn(1\tilde{P}_H)$. This signal has been demonstrated to be the primary indicator for detecting mass imbalances (Sections 4.1.6 and 5.2.3). If significant deviations are observed, a mass imbalance fault is identified. Although pitch misalignment can also affect the rotor speed signal's 1P amplitude, the robust detection of pitch misalignment at Node 1 ensures that mass imbalances are clearly diagnosed at this second stage.

If neither pitch misalignment nor mass imbalance is detected, the system proceeds to Node 3, where it calculates the median values of the 3P peak amplitudes ($Mdn(3\tilde{P})$) for all CMS vibration signals and compares them to their respective healthy thresholds ($Mdn(3\tilde{P}_H)$). If no conditions are met at any of the three nodes, the system is classified as healthy.

This approach forms the foundation for a potential data-driven model capable of robustly detecting and classifying rotor imbalances using existing CMS signals already integrated into modern wind turbines. If properly implemented, this method could significantly reduce operations and maintenance (O&M) costs by enabling more effective predictive maintenance scheduling. This methodology aligns its scope with the research conducted by Mehlan et al. [23], sharing common principles while offering an alternative to the proposed knowledge-based expert system. In particular, their approach relies on the detection of 1P frequency peak amplitudes. The study highlighted the use of these frequency peaks in the turbine's nacelle side-to-side acceleration signal as a universal recognition parameter for rotor imbalances, and drivetrain side-to-side vibration velocity as an indicator of pitch misalignment. However, robust detection and distinction between mass imbalance and yaw misalignment were only observed at cut-in wind speeds through rotor speed signal analysis, with no clear insights provided for partial and full load operational ranges. The study also proposed a method for identifying the location of the faulty blades.

The key distinction between this thesis and the reference study lies in the use of alternative CMS signals and the extraction of different features for fault detection and diagnosis. This approach enables the monitoring of various health conditions by tracking the evolution of fault severity. Specifically, while the reference study selects signal mean values and 1P peak amplitudes as features for imbalance identification, this thesis incorporates both 1P and 3P-related peak frequencies in the analysis. The median values of their corresponding amplitude distributions are proposed as the primary features in the detection and diagnosis framework. Additionally, unlike the reference study, which takes into account acceleration measurements, this work focuses solely on velocity measurements. Moreover, for the specific case of pitch misalignment detection, the findings in this thesis are validated through the results from the reference study concerning the gearbox housing's sensitivity to pitch misalignment. However, this study further extends the detection capabilities by incorporating vibration measurements from the main bearings and generator sensors, offering new detection paths. Additionally, the proposed framework demonstrates the potential for reliability and accuracy across multiple wind speeds, covering different power regions of the wind turbine.

6.2. Health Monitoring

After faults have been detected and diagnosed using the flowchart in Figure 6.1, the next step is to quantify and track the progression of fault severity in order to monitor the system's overall health. This section will focus on identifying potential health indicators for each of the three imbalance scenarios—mass imbalance, pitch misalignment, and yaw misalignment—individually. The respective health classification trees are outlined in Figures 6.4, 6.6, and 6.8.

Figure 6.4 presents a fault level classification tree for mass imbalance case using the ω_{rotor} signal. The median value of the 1P frequency amplitudes distribution ($Mdn(1\dot{P})$) is proposed as a potential health indicator. This approach is based on identifying which range the detected median values fall into (low, medium, or high), with threshold margins represented by $Mdn(1\dot{P}_{T_1})$ and $Mdn(1\dot{P}_{T_2})$. These thresholds define the boundaries for different fault levels. To better illustrate the application of this health indicator, Figure 6.2 presents an example result from Chapter 5, Section 5.2.3. In this example, three distinct fault level ranges are identified, with threshold values that depend on the input wind speed. Specifically, this case focuses on the rated power region, corresponding to wind speed U_4 . The faulty median amplitude values, $Mdn(1\dot{P})$, are depicted as red markers ('x' values) and are shown to fall within visually distinct regions. Furthermore, the corresponding median values amplitude magnitudes are reported in Table 6.3. The threshold boundaries are defined graphically in this example in Figure 6.2, the results demonstrate the feasibility of implementing them numerically for a precise fault level classification.

Since manually setting the fault severity threshold margins is not feasible in real conditions as a scalable solution, advanced methods such as machine learning are required. These methods would allow for the automatic definition of threshold values, pre-trained on previously collected faulty data sets. Specifically, supervised machine learning algorithms, such as Support Vector Machines (SVM), could serve as an effective tool for establishing the boundary values for health conditions assessment. In general, SVM works by identifying an optimal decision boundary, or hyperplane, that maximally separates data points from different classes (an optimization problem). Maximizing the margin, i.e. the distance between the decision boundary and the nearest data points, ensures robustness when classifying new data points. The decision function of an SVM is determined by a subset of training samples known as support vectors. For a more detailed exploration of machine learning techniques, including SVM, numerous open-source books, such as the one by James et al. [59], offer comprehensive insights. However, due to time constraints, implementing a supervised data-driven machine learning algorithm was beyond the scope of this thesis. Therefore, this chapter is intended to provide an overview of the potential practical applications and uses of the findings collected in this study.

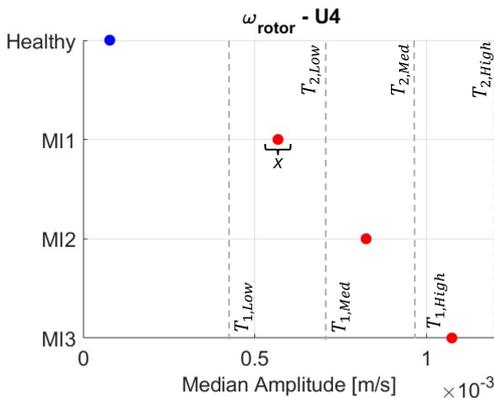


Figure 6.2: ω_{rotor} amplitude distribution median values comparison after significant statistic difference detection

Fault level	x_{values}
Healthy	7.75e-05
Low	5.65e-04
Medium	8.25e-04
High	1.1e-03

Figure 6.3: ω_{rotor} amplitude distribution median values magnitude for different health states: Healthy vs MI1/MI2/MI3

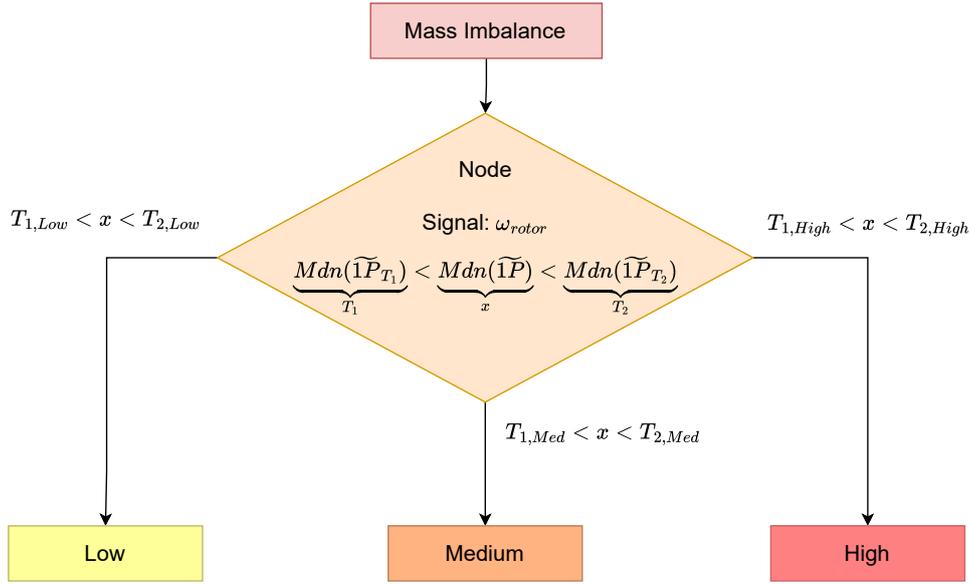


Figure 6.4: Mass Imbalance fault severity assessment tree applied to ω_{rotor} signal

Figure 6.6 presents two potential methods for monitoring the severity of pitch misalignment. Based on the findings from Chapter 4 and Chapter 5, some particular fault-related features of the signals analyzed demonstrate potential for pitch misalignment detection.

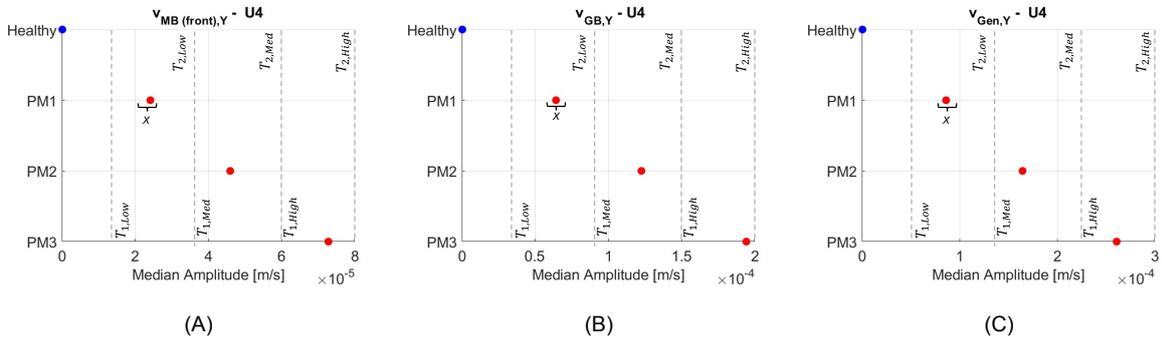


Figure 6.5: 1P - Amplitude distribution median values magnitude for different health states: Healthy vs PM1/PM2/PM3: $v_{MB(front), Y}$, $v_{GB, Y}$ and $v_{Gen, Y}$ signals

The first method presented builds a health monitoring approach similar to what was proposed for the case of mass imbalance. It can be applied within the entire turbine operational wind speed range and it uses the median value of the 1P peak amplitude distributions ($Mdn(1P)$) as a health indicator. However, in this scenario, the threshold values, $Mdn(1P_{T_1})$ and $Mdn(1P_{T_2})$, must be established for both different incoming wind speed and the specific CMS vibration signal under analysis. As it can be deduced from the three plots presented in Figure 6.5, different signals have different magnitude ranges for their respective health states. This also applies in the context of different wind speeds, in this case, U_4 wind speed analysis is reported for median value comparison of $v_{MB(front), Y}$, $v_{GB, Y}$ and $v_{Gen, Y}$ signals. Similarly to the previous imbalance scenario, boundary thresholds are extracted graphically for representative purposes, in real-world conditions appropriate training of the system through collected faulty data sets for each sensor position would be necessary. This study's findings indicate that all three levels of misalignment severity—low, medium, and high—are clearly distinguishable.

The second method proposed relies on time-domain analysis (3.5.2), according to the findings of Section 5.3. It includes the evaluation of the drivetrain torsional mode filtered in the main bearings accel-

eration signals ($v_{MB_{front}}$ and $v_{MB_{rear}}$). However, this method is limited to wind speeds in the rated power region; below-rated conditions, detection has proven to be not possible (Section 4.2). When the wind speed condition is met, the system's health can be assessed by differentiating between the three misalignment severity levels using three primary condition indicators, which showed robustness in the analysis. These are CI_{RMS} (Equation 2.17), CI_{APDF} (Equation 2.20), and CI_{DND} (Equation 2.21).

Similarly, for each severity level, threshold margins $CI(1\tilde{P}_{T_1})$ and $CI(1\tilde{P}_{T_2})$ must be established, considering both the specific condition indicator used and the corresponding wind speed. The results from turbulent conditions, as shown in Table 5.1, reveal significant gaps between the value ranges of the three condition indicators, allowing for clear differentiation between health states. The same considerations for defining health state thresholds, as discussed in the previous methods, apply here as well. It is important to emphasize that the purpose of the presented decision trees is to demonstrate that, based on this study's findings, such results are achievable and provide a solid foundation for future machine learning-based implementations.

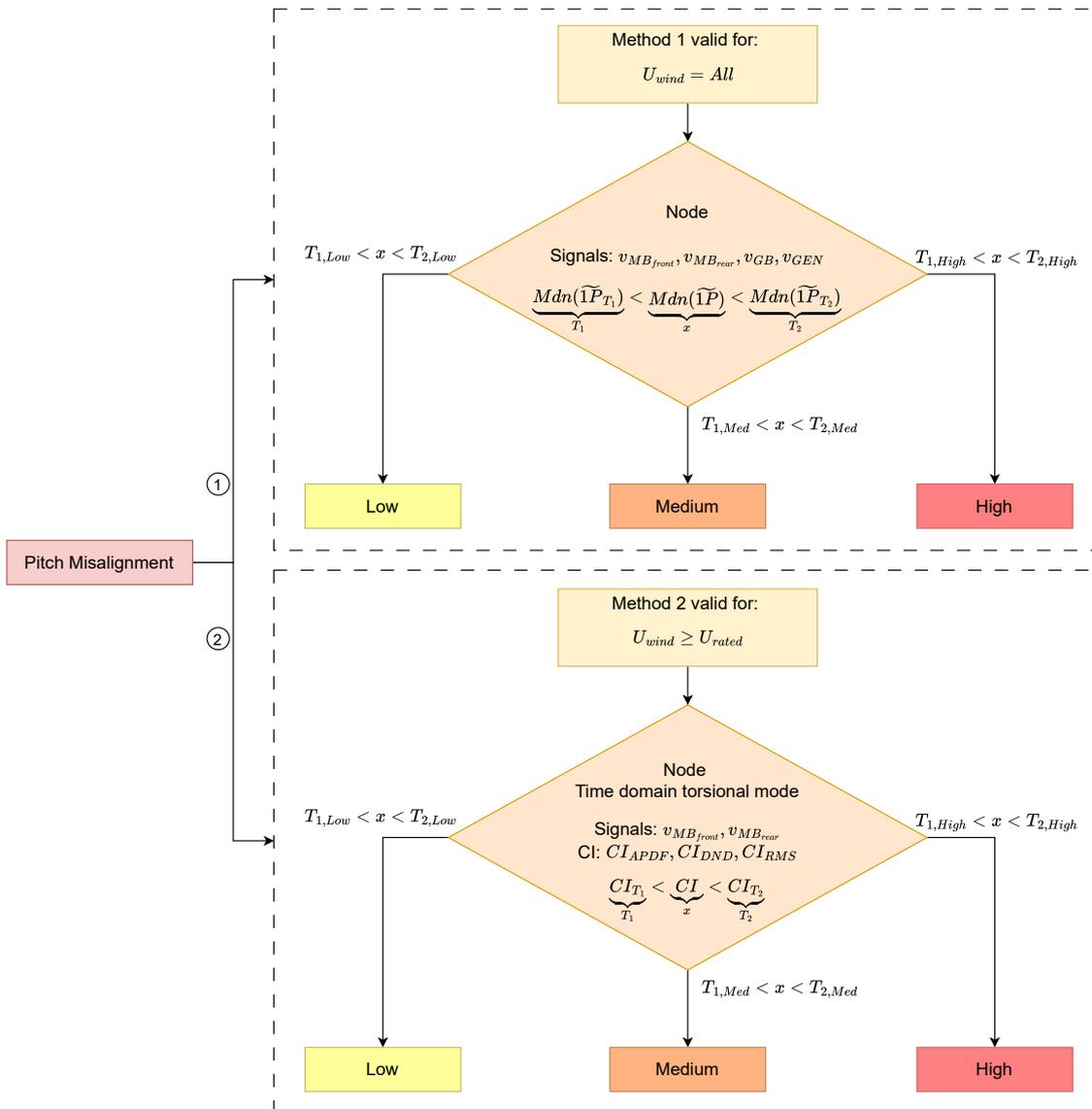


Figure 6.6: Pitch misalignment fault severity assessment tree

Lastly, the health monitoring approach for yaw misalignment is illustrated in Figure 6.8. This method is similar to those used for mass imbalance and pitch misalignment, but it selects the 3P-related median values from amplitude distributions as health indicators ($Mdn(3\tilde{P})$). As in previous cases, the median values would have to be estimated for all four CMS signals and compared to pre-trained thresholds, $Mdn(3\tilde{P}_{T_1})$ and $Mdn(3\tilde{P}_{T_2})$, which are determined for all the different incoming wind speeds.

However, a key distinction in this case is the presence of two possible pathways in the flowchart, depending on the wind speed. If the wind speed is within the rated power region of the curve ($U_{rated} < U_{wind} < U_{cut-in}$), thresholds can be robustly defined, for medium to high misalignment severities, as it is shown in the example reported in Figure 6.7 (B). Conversely, in the transition region, where the wind speed is between U_{cut-in} and U_{rated} , only high misalignment severity ranges can be reliably identified (see Figure 6.7 (A)). Therefore, this approach presents certain limitations in terms of detectability, particularly when assessing the presence and severity of yaw misalignment under varying wind conditions.

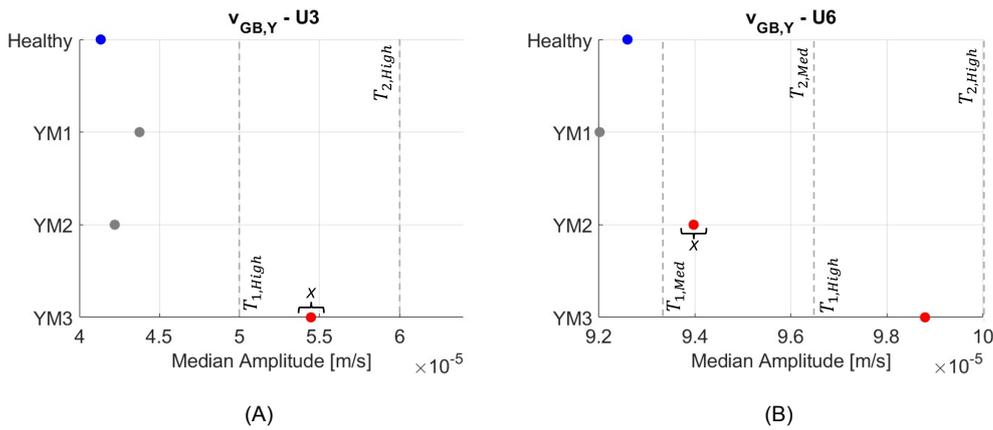


Figure 6.7: 3P - Amplitude distribution median values magnitude for different health states: Healthy vs YM1/YM2/YM3: $v_{GB, Y}$ signal: (A) U_3 ; (B) U_6

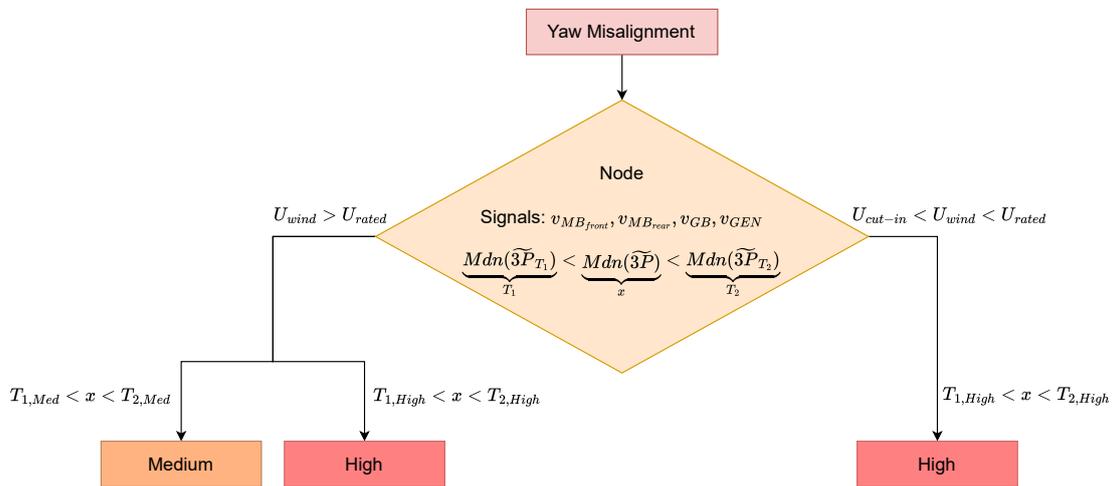


Figure 6.8: Yaw misalignment fault severity assessment tree through $Mdn(3\tilde{P})$ values and for all CMS signals

6.3. Recommendations and Limitations

Despite the promising results presented in this thesis, several recommendations and limitations need to be carefully addressed when considering the real-world implementation of the proposed methodology.

One significant aspect to be considered is the absence of noise in the simulations, which does not reflect the complexity of real-world conditions where noise is inherently present and can influence the measurements. Furthermore, frequency domain signal analysis is proved to be strongly dependent on measurement resolution, therefore, it might be beneficial to investigate the detection capabilities when going towards higher sampling frequencies with respect to the one used in this case study.

Additionally, the work presented in this thesis refers to a four-point-support, geared drivetrain model, which features simplified housing models for the main gearbox and generator. This may introduce inaccuracies when simulating vibration measurements, as the flexibility of the housing is not fully captured.

The input fault models used in the simulations are constrained, with either mass imbalance or pitch misalignment being applied to only one blade or as single fault cases. In reality, combined fault scenarios, such as faults affecting multiple blades, or the joint presence of aerodynamic imbalances and mass imbalance are plausible and should be explored further, as their behavior could significantly differ from the isolated cases studied here.

Moreover, the simulations employed a limited number of seeds for turbulent inflow conditions due to the expensive computational time for the simulation in SIMPACK. More comprehensive studies that include additional seeds and better representation of complete turbulent atmospheric conditions should be considered for more realistic inflow modeling, as suggested in the TurbSim manual.

It is also important to note that the current time-domain analysis primarily targets the torsional mode, as it is the dominant mode that transmits the effects of rotor torque loading conditions throughout the drivetrain and power transmission system. To develop a more comprehensive understanding, it is essential to investigate also other system modes that reflect the impact of non-torque loads such as thrust and shear forces. Such loading conditions are mainly reflected through axial modes response in the drivetrain since they induce blades out of plane motions. Additionally, further analysis is required in the x and z-axis measurement directions, as this study primarily focused on the y-axis direction.

Frequency domain analysis primarily focused on velocity measurements; however, due to the interest in lower frequency ranges, it may be beneficial to explore displacement sensor measurements as well, as suggested in Table 3.7. This would provide a broader perspective and potentially more robust detection criteria.

Yaw misalignment detection through 3P frequency analysis has been found to lack robustness at cut-in wind speeds and for small misalignment angles, limiting its reliability. To overcome these limitations, integrating SCADA data, specifically by using the electrical power output of the wind turbine, offers a promising solution. Jing et al. [60] explored this approach, focusing on the partial load region of the power curve, particularly for wind speeds near the cut-in range, as an effective method for yaw misalignment detection.

Another crucial factor to consider is the precise calibration of thresholds for the proposed health monitoring indicators, alongside careful investigation for management of the false alarm rate. Both aspects are vital to ensuring the system's reliability in real-world applications. To guarantee robust performance, a thorough validation of the proposed system's false positive and false negative detection rates should be conducted.

Finally, the proposed framework for the system health monitoring and assessment requires in real-world conditions a data-driven approach (see Section 2.5), which requires large amounts of data collection for thresholds pre-training across both healthy and faulty operating conditions. However, such comprehensive datasets, specifically for faulty conditions, are often difficult to obtain in practice, which presents a significant challenge for accurate model implementation and validation.

7

Conclusions

Rotor imbalance represents a major risk in wind turbines, being a primary contributor to both increased downtime and diminished energy output. Conventional maintenance approaches, which rely on frequent inspections, are labor-intensive and time-consuming. Nevertheless, these inspections remain common because turbines often do not have reliable automatic detection systems integrated as part of their routine operations. Shifting from a time-based to a condition-based maintenance strategy shows promising results in reducing system failures, costs, and unexpected downtime, to enhance and optimize energy production.

In this chapter, the primary research question and subquestions posed in the introduction (Section 1.1) are addressed.

Main Research Question: *Can rotor imbalance be efficiently detected and diagnosed using existing condition monitoring sensors on the wind turbine drivetrain to prevent severe component failures and ensure reliable power production?*

The findings of this research confirm that rotor imbalance can be effectively identified and diagnosed by utilizing existing operational data collected from wind turbines. This is achieved by integrating drivetrain CMS vibration signals with SCADA system data. By exploiting these combined signals, the detection of rotor imbalances is significantly enhanced, eliminating the need for direct blade measurements—a common shortcoming in current turbine monitoring systems. The developed methodology underscores the feasibility of the proposed framework as a practical and scalable solution for fault detection and diagnosis. Additionally, it provides a foundation for a new approach to assess fault severity, thereby supporting the optimization of maintenance strategies in wind turbine operations.

This leads to answering the four subquestions posed:

1. *Which signal analysis techniques are most suitable for detecting the presence of rotor imbalances? Which drivetrain signals are the most effective and reliable for their detection and diagnosis?*

Firstly, suitable signal analysis techniques for detecting rotor imbalances were explored. The tools implemented in this study encompass fundamental and robust approaches, specifically frequency domain analysis using FFT and time domain analysis based on statistical methods that focus specifically on drivetrain torsional modes. CMS signals derived from different sensors positioned along the drivetrain have been analyzed. Front and rear main bearings, gearbox housing, and generator were selected according to the international ISO 10816-21 [18] standards. Based on the literature review and simulation results, the side-side (y-axis) direction measurements proved to be the most sensitive, offering enhanced resolution for detecting imbalances. Due to the frequency ranges under study, velocity vibration CMS signals were selected for frequency domain analysis, complemented by rotor speed signals from the SCADA system. For time domain analysis, acceleration signals from the CMS sensors were investigated.

2. *Which signal features ensure robust fault detection?*

The proposed approach focused on extracting two key frequency domain features that highlight dynamic impacts on the signals at the rotor rotational frequency and its third harmonic (1P and 3P). The 1P frequency, associated with harmonic periodic loads, is particularly useful for identifying mass imbalance and pitch misalignment, where the center of gravitational or thrust force shifts from the hub to the faulty blade. Meanwhile, the 3P frequency, which captures non-harmonic periodic loads 2.5, is proved to be effective for detecting dynamic effects such as yaw misalignment. After applying pre-processing techniques such as transient time removal, signals resampling in order domain, and windowing, these features were extracted from each signal and for each window. The final condition indicator for fault detection was defined as the median value of the distribution of peak amplitudes for both 1P and 3P frequencies. In the time domain analysis, five statistical condition indicators were applied to the filtered CMS signals, isolating drivetrain torsional mode. These indicators were effective only for detecting pitch misalignment under rated power conditions. Overall the listed features proved to be relevant methods for distinguishing normal rotor behavior from imbalanced conditions.

3. *Can a decision tree algorithm be developed for fault detection and classification?*

The results obtained from various input combinations in the simulation setup support the development of a fault detection and diagnosis classification tree through the implementation of a decision fusion level framework. The proposed method relies on the detection of statistical changes in the median values of peak amplitude distributions of all four CMS and rotor speed signals with respect to a threshold derived under healthy conditions. This approach, which is wind-speed dependent, enabled clear classification and distinction between the different imbalance scenarios. A binary decision tree was thus established, laying the groundwork for a data-driven approach that could be implemented in real-world conditions for automatic fault detection.

4. *Can health indicators be constructed for the specific imbalance scenarios under investigation? To what extent can they identify fault severity levels?*

The study found that fault severities can be distinguished by clustering the median value feature for both 1P and 3P frequencies ($Mdn(\tilde{1P})$ and $Mdn(\tilde{3P})$) within predefined upper and lower threshold ranges. These ranges can be determined for real-world implementation through machine learning techniques like SVM, which involve system training based on faulty datasets. This poses some limitations in real practical implementation since rarely these data are available. For mass imbalance, this health indicator is achieved using the rotor speed signal, while for pitch misalignment, all CMS signals are involved. In both cases, through the use of representative results examples obtained in Chapter 5, the fault level distinction was proved effective across the three imbalance severity levels studied. For yaw misalignment, 3P-related median values enabled health condition assessment, though robust detection was observed only for higher severity levels. Additionally, in the specific case of pitch misalignment under rated power conditions, three of the five statistical condition indicators in the time domain proved effective for health condition monitoring based on the main bearings' measured signals.

As previously outlined, several recommendations emerge from the findings of this study, suggesting directions for future work and further improvements. First, addressing the absence of noise in the analyzed simulated signals is crucial, as real-world conditions inherently introduce noise that can affect measurements. In parallel, exploring higher sampling frequencies would improve signal resolution and enhance the effectiveness of frequency domain analysis. Additionally, refining the drivetrain model to incorporate more realistic housing flexibility, and expanding the analysis to combined fault scenarios—such as simultaneous mass and aerodynamic imbalances—would increase the model's accuracy. Moreover, extending the analysis to cover additional system modes, particularly those influenced by non-torque loads, is recommended to capture a broader range of operational dynamics. Furthermore, integrating displacement sensors, as well as SCADA data like electrical power output, could improve the robustness of fault detection. Finally, precise calibration of health monitoring thresholds, combined with the collection of more extensive faulty condition datasets, is crucial for ensuring the model's applicability in real-world environments.

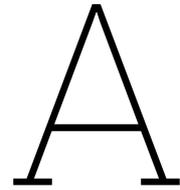
References

- [1] International Renewable Energy Agency. *World Energy Transitions Outlook 2023: 1.5°C Pathway, Volume 1*. Abu Dhabi, 2023. URL: https://mc-cd8320d4-36a1-40ac-83cc-3389-cdn-endpoint.azureedge.net/-/media/Files/IRENA/Agency/Publication/2023/Jun/IRENA_World_energy_transitions_outlook_2023.pdf?rev=db3ca01ecb4a4ef8accb31d017934e97.
- [2] Mark Hutchinson et al. *Global Wind Report 2023*. Brussels, Belgium: Global Wind Energy Council, 2023. URL: <http://www.gwec.net>.
- [3] Dayton A Griffin. *THE CHALLENGES OF WIND TURBINE BLADE DURABILITY*. URL: <https://www.dnv.com/publications/the-challenges-of-wind-turbine-blade-durability-243601/>.
- [4] European Agency for Safety and Health at Work. *Occupational safety and health in the wind energy sector – European Risk Observatory – Report*. Publications Office, 2013. DOI: doi/10.2802/86555.
- [5] Liliana Cunha, Daniel Silva, and Mariana Macedo. “Different Shades of Green: An Analysis of the Occupational Health and Safety Risks Faced by Wind Farm Workers”. In: *Sustainability* 16.7 (2024), p. 3012.
- [6] Jenny Niebsch and Ronny Ramlau. “Simultaneous estimation of mass and aerodynamic rotor imbalances for wind turbines”. In: *Journal of Mathematics in Industry* 4 (2014), pp. 1–19.
- [7] Christopher J. Crabtree, Donatella Zappalá, and Simon I. Hogg. “Wind energy: UK experiences and offshore operational challenges”. In: vol. 229. SAGE Publications Ltd, Nov. 2015, pp. 727–746. DOI: 10.1177/0957650915597560.
- [8] Bin Lu et al. “A review of recent advances in wind turbine condition monitoring and fault diagnosis”. In: *2009 IEEE power electronics and machines in wind applications* (2009), pp. 1–7.
- [9] Pierre Tchakoua et al. “New trends and future challenges for wind turbines condition monitoring”. In: *2013 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE. 2013, pp. 238–245.
- [10] Tomasz Barszcz. *Vibration-based condition monitoring of wind turbines*. Vol. 14. Springer, 2019.
- [11] Fausto Pedro García Márquez et al. “Condition monitoring of wind turbines: Techniques and methods”. In: *Renewable energy* 46 (2012), pp. 169–178.
- [12] “Wind turbine drivetrains: State-of-the-art technologies and future development trends”. In: *Wind Energy Science* 7 (1 Feb. 2022), pp. 387–411. ISSN: 23667451. DOI: 10.5194/wes-7-387-2022.
- [13] Mohamad Hazwan Mohd Ghazali and Wan Rahiman. “Vibration analysis for machine monitoring and diagnosis: a systematic review”. In: *Shock and Vibration* 2021 (2021), pp. 1–25.
- [14] R Bond Randall and Vibration-based Condition Monitoring. “Industrial, aerospace and automotive applications”. In: *VIBRATION-BASED CONDITION MONITORING*. West Sussex (2011), pp. 13–20.
- [15] Donatella Zappalá. “Advanced algorithms for automatic wind turbine condition monitoring”. PhD thesis. Durham University, 2014.
- [16] Z Hameed et al. “Condition monitoring and fault detection of wind turbines and related algorithms: A review”. In: *Renewable and Sustainable energy reviews* 13.1 (2009), pp. 1–39.
- [17] Gram & Juhl A/S. *TCM® Turbine Condition Monitoring*. Last accessed: 29th May 2024. 2010. URL: <http://www.rotomech.com/pdf/brochure-eng.pdf>.
- [18] International Organization for Standardization. *ISO 10816-21: Mechanical Vibration — Evaluation of Machine Vibration by Measurements on Non-Rotating Parts — Part 21: Horizontal Axis Wind Turbines with Gearbox*. Available at: <https://www.iso.org/standard/60328.html>. 2015.

- [19] Y. Guo et al. "A systems engineering analysis of three-point and four-point wind turbine drivetrain configurations". In: *Wind Energy* 20.3 (2017), pp. 537–550. DOI: <https://doi.org/10.1002/we.2022>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.2022>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2022>.
- [20] Leon Mishnaevsky Jr. and Kenneth Thomsen. "Costs of repair of wind turbine blades: Influence of technology aspects". In: *Wind Energy* 23.12 (2020), pp. 2247–2255. DOI: <https://doi.org/10.1002/we.2552>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.2552>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2552>.
- [21] Dimitris Al Katsaprakakis, Nikos Papadakis, and Ioannis Ntintakis. "A comprehensive analysis of wind turbine blade damage". In: *Energies* 14.18 (2021), p. 5974.
- [22] Jijian Lian, Huan Zhou, and Xiaofeng Dong. "A theoretical approach for resonance analysis of wind turbines under 1P/3P loads". In: *Energies* 15.16 (2022), p. 5787.
- [23] Felix C Mehlan and Amir R Nejad. "Rotor imbalance detection and diagnosis in floating wind turbines by means of drivetrain condition monitoring". In: *Renewable Energy* 212 (2023), pp. 70–81.
- [24] Jin Xu et al. "Rotor imbalance detection and quantification in wind turbines via vibration analysis". In: *Wind Engineering* 46.1 (2022), pp. 3–11.
- [25] Michiel Zaaijer and Axelle Viré. *Introduction to Wind Turbines: Physics and Technology*. Ed. by Andres Leiro Fonseca Ricardo Balbino Dos Santos Pereira Amir Daneshbodi. Version 20/09/2023, 2023/2024. 2023. URL: <https://mytimetable.tudelft.nl/>.
- [26] Glidden S. Doman. "Structural Dynamic Considerations in Wind Turbine Design". In: *Wind Turbine Technology: Fundamental Concepts in Wind Turbine Engineering, Second Edition*. ASME Press, Jan. 2009. ISBN: 9780791802601. DOI: 10.1115/1.802601.ch10. eprint: https://asmedigitalcollection.asme.org/ebooks/book/chapter-pdf/2802290/802601_ch10.pdf. URL: <https://doi.org/10.1115/1.802601.ch10>.
- [27] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [28] KR Fyfe and EDS Munck. "Analysis of computed order tracking". In: *Mechanical systems and signal processing* 11.2 (1997), pp. 187–205.
- [29] P Večeř, Marcel Kreidl, and R Šmíd. "Condition indicators for gearbox condition monitoring systems". In: *Acta Polytechnica* 45.6 (2005).
- [30] Pawel Jakub Rzeszucinski. *Development of reliable vibration-based condition indicators and their data fusion for the robust health diagnosis of gearboxes*. The University of Manchester (United Kingdom), 2012.
- [31] David L Hall and James Llinas. "An introduction to multisensor data fusion". In: *Proceedings of the IEEE* 85.1 (1997), pp. 6–23.
- [32] Federico Castanedo et al. "A review of data fusion techniques". In: *The scientific world journal* 2013 (2013).
- [33] Gang Niu, Bo-Suk Yang, and Michael Pecht. "Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance". In: *Reliability engineering & system safety* 95.7 (2010), pp. 786–796. DOI: 10.1016/j.res.2010.02.016.
- [34] Dassault Systèmes. *SIMPACT Multibody Simulation Software*. [Accessed: 30-Sep-2024]. 2024. URL: <https://www.3ds.com/products/simulia/simpack>.
- [35] W Dheelibun Remigius and Anand Natarajan. "A review of wind turbine drivetrain loads and load effects for fixed and floating wind turbines". In: *Wiley Interdisciplinary Reviews: Energy and Environment* 11.1 (2022), e417.
- [36] Shuaishuai Wang, Amir R Nejad, and Torgeir Moan. "On design, modeling, and analysis of a 10-MW medium-speed drivetrain for offshore wind turbines". In: *Wind Energy* 23.4 (2020), pp. 1099–1117.
- [37] C. Bak et al. *The DTU 10-MW Reference Wind Turbine*. Sound/Visual production (digital). 2013.

- [38] Francisco Oyague. *Gearbox modeling and load simulation of a baseline 750-kW wind turbine using state-of-the-art simulation codes*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2009.
- [39] Amir Rasekhi Nejad et al. “Development of a 5 MW reference gearbox for offshore wind turbines”. In: *Wind Energy* 19.6 (2016), pp. 1089–1106.
- [40] Bonnie Jonkman and Jason Jonkman. “FAST v8. 16.00 a-bjj”. In: *National Renewable Energy Laboratory* 1355 (2016).
- [41] International Electrotechnical Commission et al. “IEC 61400-13 Wind Turbine Generator Systems—Part 13: Measurement of Mechanical Loads”. In: *International Electrotechnical Commission: Geneva, Switzerland* (2015).
- [42] JM Jonkman. “FAST User’s Guide”. In: *National Renewable Energy Laboratory Technical Report* (2005).
- [43] Hatto Schneider. *Rotor Balancing: Fundamentals for Systematic Processes*. Springer Nature, 2023.
- [44] M. Saathoff et al. “Effect of individual blade pitch angle misalignment on the remaining useful life of wind turbines”. In: *Wind Energy Science* 6.5 (2021), pp. 1079–1087. DOI: 10.5194/wes-6-1079-2021. URL: <https://wes.copernicus.org/articles/6/1079/2021/>.
- [45] CARLO LUIGI Bottasso and CED Riboldi. “Estimation of wind misalignment and vertical shear from blade loads”. In: *Renewable energy* 62 (2014), pp. 293–302.
- [46] Martin Cardaun et al. “Analysis of wind-turbine main bearing loads due to constant yaw misalignments over a 20 years timespan”. In: *Energies* 12.9 (2019), p. 1768.
- [47] Bonnie J Jonkman. *TurbSim user’s guide*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2006.
- [48] John H McDonald. *Handbook of biological statistics*. Vol. 2. sparky house publishing Baltimore, MD, 2009.
- [49] Michael Steer. *Microwave and RF design*. NC State University, 2019.
- [50] EMWorks. *Understanding RLC Bandpass Filters: Simulation and Optimization*. Accessed: 2024-10-01. n.d. URL: <https://www.emworks.com/application/understanding-rlc-bandpass-filters-simulation-and-optimization>.
- [51] MathWorks. *Condition Indicators for Monitoring, Fault Detection, and Prediction*. Accessed: 2024-10-01. n.d. URL: <https://www.mathworks.com/help/predmaint/ug/condition-indicators-for-monitoring-fault-detection-and-prediction.html>.
- [52] Yadolah Dodge. “Kolmogorov–Smirnov Test”. In: *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 283–287. ISBN: 978-0-387-32833-1. DOI: 10.1007/978-0-387-32833-1_214. URL: https://doi.org/10.1007/978-0-387-32833-1_214.
- [53] Peter Ran Li Crabtree Christopher Tavner. *5.12 Wear and Debris Instrumentation*. 2020. URL: <https://app.knovel.com/hotlink/khtml/id:kt01294V71/condition-monitoring/wear-debris-instrumentation>.
- [54] A. Kahraman and G. W. Blankenship. “Experiments on Nonlinear Dynamic Behavior of an Oscillator With Clearance and Periodically Time-Varying Parameters”. In: *Journal of Applied Mechanics* 64.1 (Mar. 1997), pp. 217–226. ISSN: 0021-8936. DOI: 10.1115/1.2787276. eprint: https://asmedigitalcollection.asme.org/appliedmechanics/article-pdf/64/1/217/5464787/217_1.pdf. URL: <https://doi.org/10.1115/1.2787276>.
- [55] Zhanwei Li et al. “Dynamic modeling and analysis of wind turbine drivetrain considering the effects of non-torque loads”. In: *Applied Mathematical Modelling* 83 (2020), pp. 146–168. ISSN: 0307-904X. DOI: <https://doi.org/10.1016/j.apm.2020.02.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0307904X20301037>.

- [56] Cheng-Han Chuang et al. "Modal analysis, control of frequency, and torsional modes of microgrid with doubly fed induction generator wind turbines providing frequency support". In: *IET Renewable Power Generation* 17.9 (2023), pp. 2315–2334. DOI: <https://doi.org/10.1049/rpg2.12761>. eprint: <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/rpg2.12761>. URL: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/rpg2.12761>.
- [57] Jianxiang Yang et al. "Torsional vibration characteristics analysis and adaptive fixed-time control of wind turbine drivetrain". In: *Energy Science & Engineering* 11.12 (2023), pp. 4666–4686.
- [58] Sabrina Milani et al. "Automatic Detection and Intensity Classification of Pitch Misalignment of Wind Turbine Blades: a Learning-based Approach". In: *Journal of Physics: Conference Series*. Vol. 2767. 3. IOP Publishing, 2024, p. 032010.
- [59] Gareth James et al. *An introduction to statistical learning: With applications in python*. Springer Nature, 2023.
- [60] Bo Jing et al. "Improving wind turbine efficiency through detection and calibration of yaw misalignment". In: *Renewable Energy* 160 (2020), pp. 1217–1227. ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2020.07.063>. URL: <https://www.sciencedirect.com/science/article/pii/S0960148120311393>.



10 MW DTU Design Specifics

Parameter	Value
Rated power (MW)	10
IEC Class	IA
Rotor orientation, configuration	Upwind, 3 blades
Control	Variable speed, collective pitch
Drivetrain	medium-speed, multi-stage gearbox
Cut in wind speed (m/s)	4
Rated wind speed (m/s)	11.4
Cut out wind speed (m/s)	25
Rotor diameter (m)	178.3
Hub height (m)	119
Hub diameter (m)	5.6
Hub overhang (m)	7.1
Shaft tilt angle (deg)	5
Rotor mass (t)	229
Nacelle mass (t)	446
Tower mass (t)	605

Table A.1: 10 MW DTU reference turbine specifics [37]

Parameter	Value
Gearbox ratio	1:50
Minimum rotor speed (rpm)	6.0
Rated rotor speed (rpm)	9.6
Rated generator speed (rpm)	480.0
Electrical generator efficiency	94
Generator inertia about high-speed shaft (kgm ²)	1500.5
Equivalent drive-shaft torsional-spring constant (Nm/rad)	2317025352
Equivalent Drive-Shaft torsional-damping constant (Nm/(rad/s))	9240560
Natural frequency of free-free rigid shaft torsion mode	4.003
Natural frequency of free-fixed rigid shaft torsion mode	0.612

Table A.2: Drivetrain properties proposed by DTU[36]

B

COT Resampling algorithm

B.1. Function Code

```
1 function [x_alpha_vector, delta_t_alpha, t_alpha, fs_resampled, PHIs] =
2     COT(vibration, fs_vibration, fs_speed, speed_rpm)
3
4     % Create time vectors
5     t_vibration = (0:length(vibration)-1) / fs_vibration;
6     t_speed = (0:length(speed_rpm)-1) / fs_speed;
7
8     % Convert speed from RPM to RPS
9     speed_rps = speed_rpm / 60;
10
11    % Interpolate rotational speed vector to match the vibration signal's time vector
12    speed_rps_interp = interp1(t_speed, speed_rps, t_vibration, 'linear', 'extrap');
13
14    % Calculate the phase angle by integrating the rotational speed
15    phase_angle = cumsum(speed_rps_interp) * 2 * pi / fs_vibration; % phase angle in radians
16    (2*pi)
17
18    % Identify time instants corresponding to integer rotations (phase angle wraps for
19    constant angular increments)
20    num_rotations = floor(phase_angle(end) / (2 * pi));
21
22    Tk = zeros(1, num_rotations + 1); %searching for time instants idx which corresponds
23    to phase wrap points (rotational periods)
24    for k = 1:num_rotations + 1
25        [~, idx] = min(abs(phase_angle - (k-1) * 2 * pi));
26        Tk(k) = t_vibration(idx);
27    end
28
29    % Define angular resolution (phase angle domain sampling)
30
31    N = length(vibration); % number of samples in the angular domain
32    K = num_rotations + 1; % number of integer rotations
33    alpha_i = 2 * pi * (0:N-1) * K / N; % angular rotation vector
34
35    % Interpolate to obtain time instants corresponding to alpha_i
36    t_alpha = interp1((0:num_rotations) * 2 * pi, Tk, alpha_i, 'linear', 'extrap');
37
38    % Calculate the new sampling frequency
39    PHIs = length(vibration)/phase_angle(end);
40
41    % Interpolate to obtain the angular resampled signal
42    x_alpha = interp1(t_vibration, vibration, t_alpha, 'linear', 'extrap');
43
44    % Return the resampled signal as a column vector
45    x_alpha_vector = x_alpha';
46 end
```

B.2. Algorithm flow-diagram

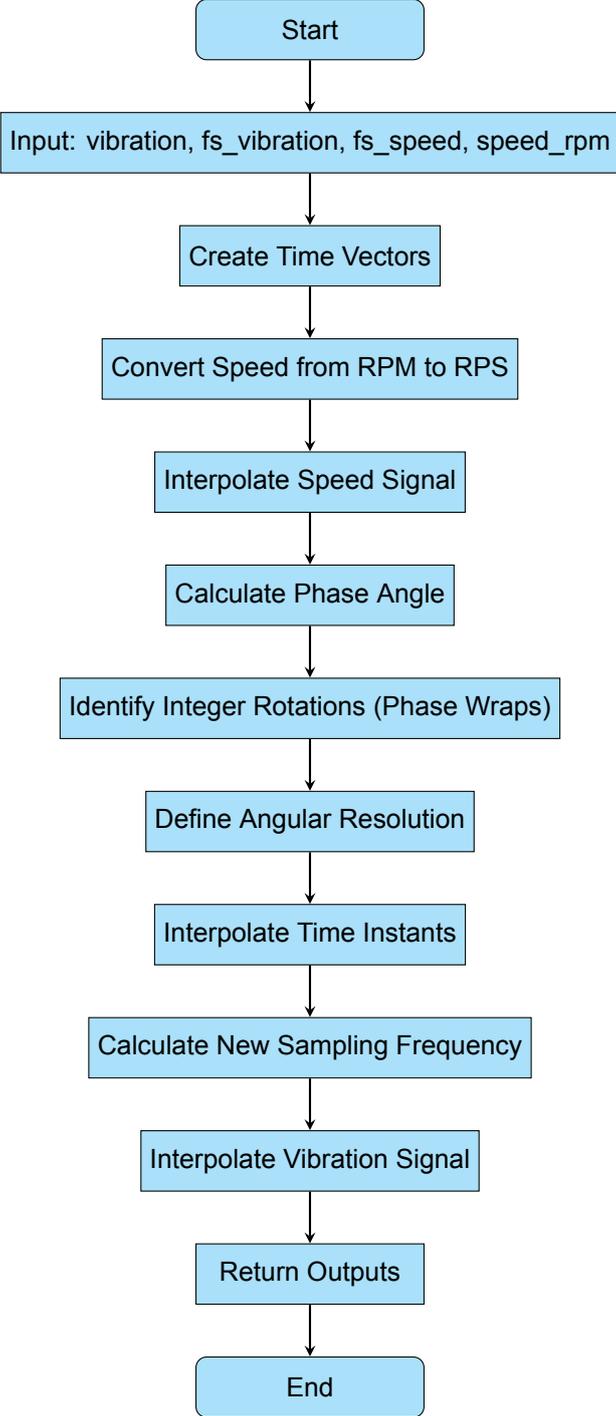
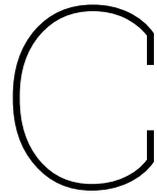
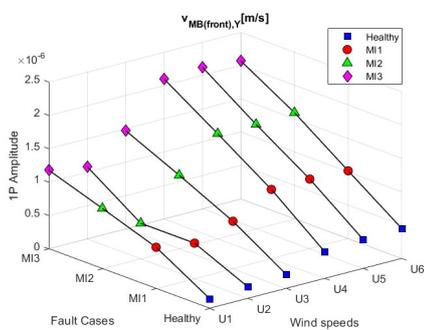


Figure B.1: Flow-diagram for COT function code (B.1)

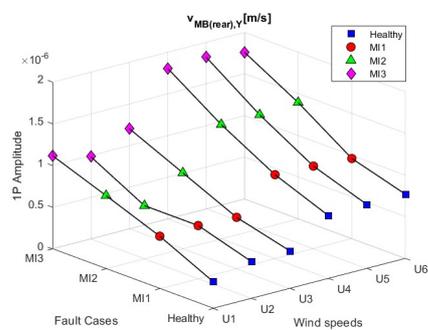


Steady State Analysis

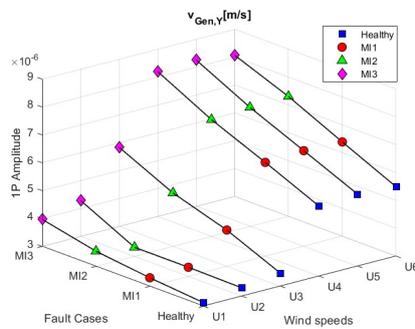
C.1. 3D averaged peak amplitudes plot



(A)



(B)



(C)

Figure C.1: MI - 1P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB}(\text{front}), \gamma$; (B) $v_{MB}(\text{rear}), \gamma$; (C) v_{Gen}, γ

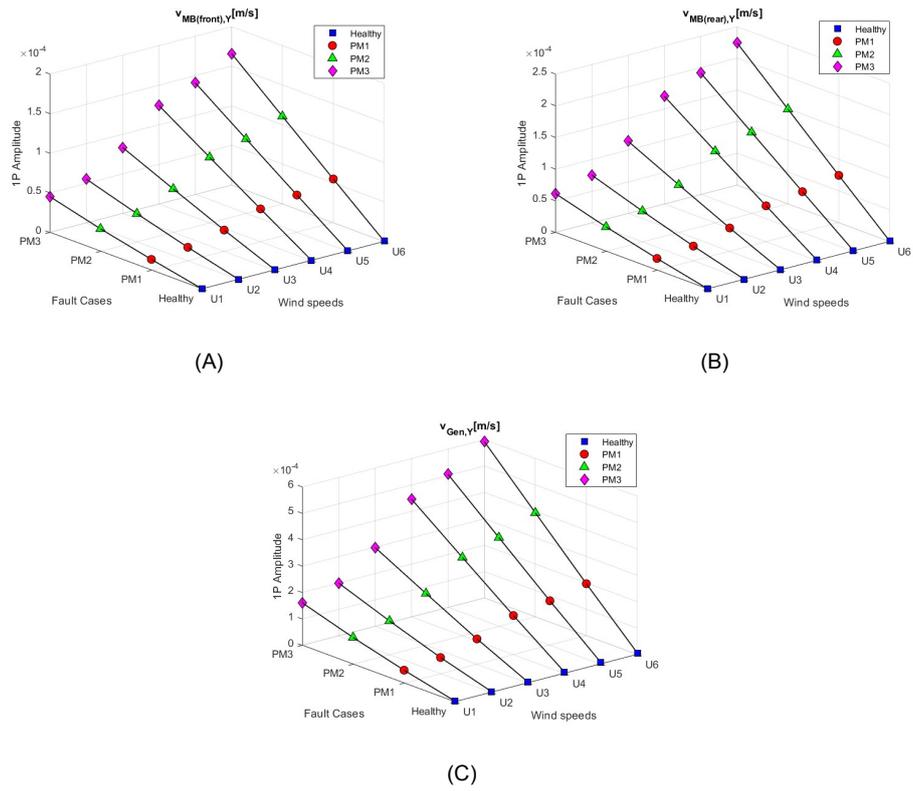


Figure C.2: PM - 1P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB} (front), \gamma$; (B) $v_{MB} (rear), \gamma$; (C) v_{Gen}, γ

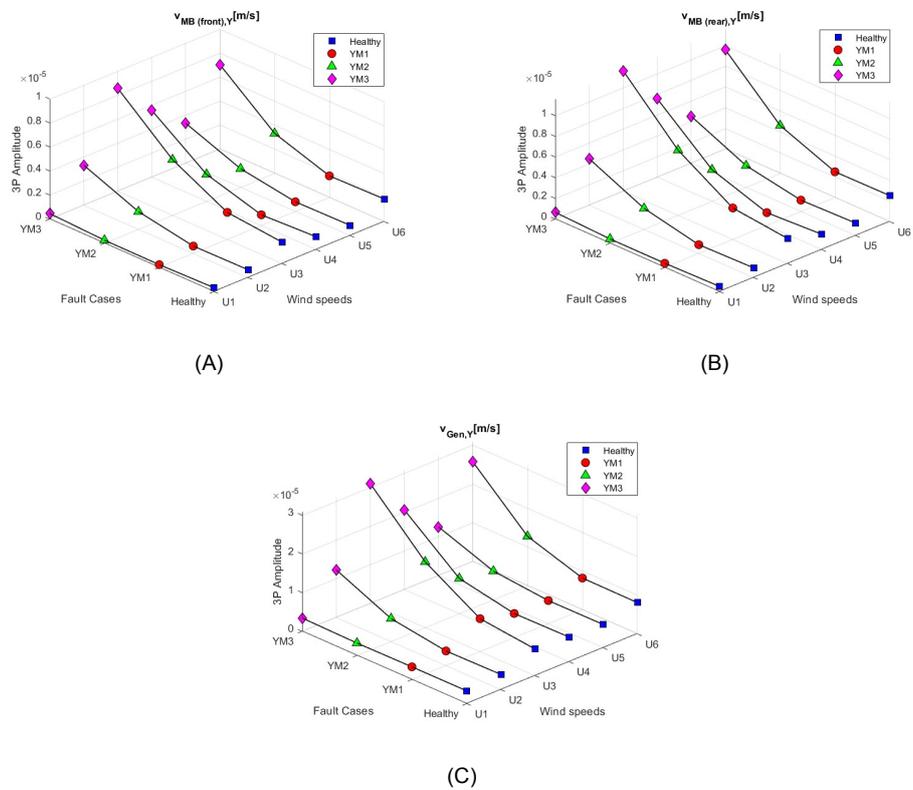


Figure C.3: YM - 3P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB} (front), \gamma$; (B) $v_{MB} (rear), \gamma$; (C) v_{Gen}, γ

C.2. Relative change detection histograms

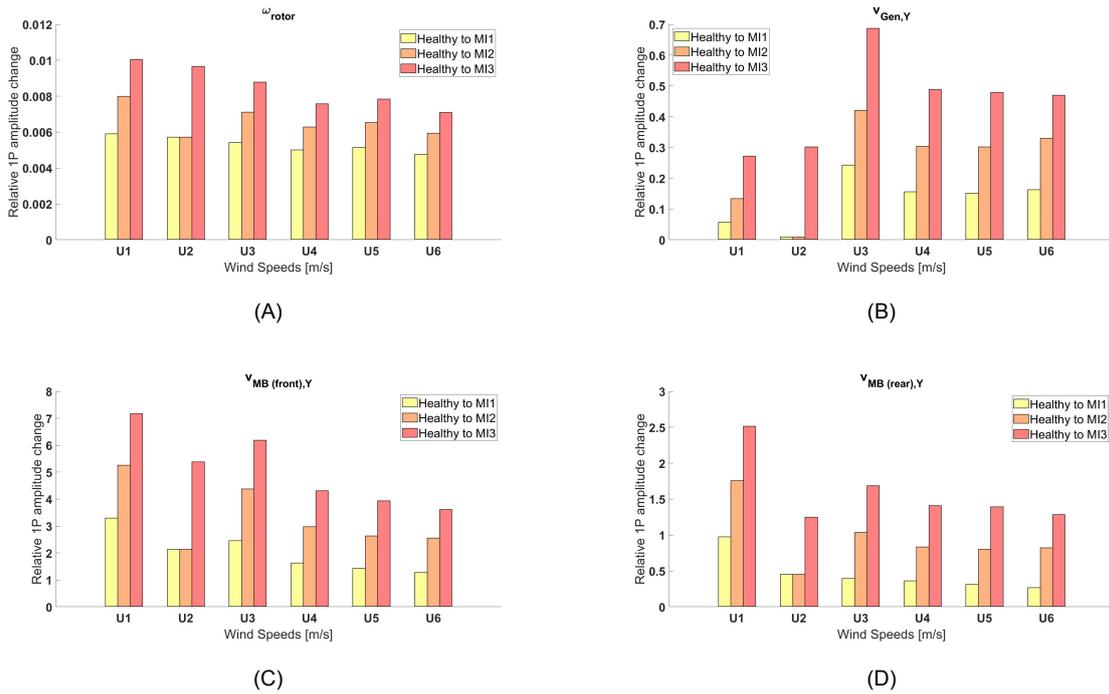


Figure C.4: MI - 1P: relative change averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $v_{Gen, \gamma}$; (C) $v_{MB (front), \gamma}$; (D) $v_{MB (rear), \gamma}$

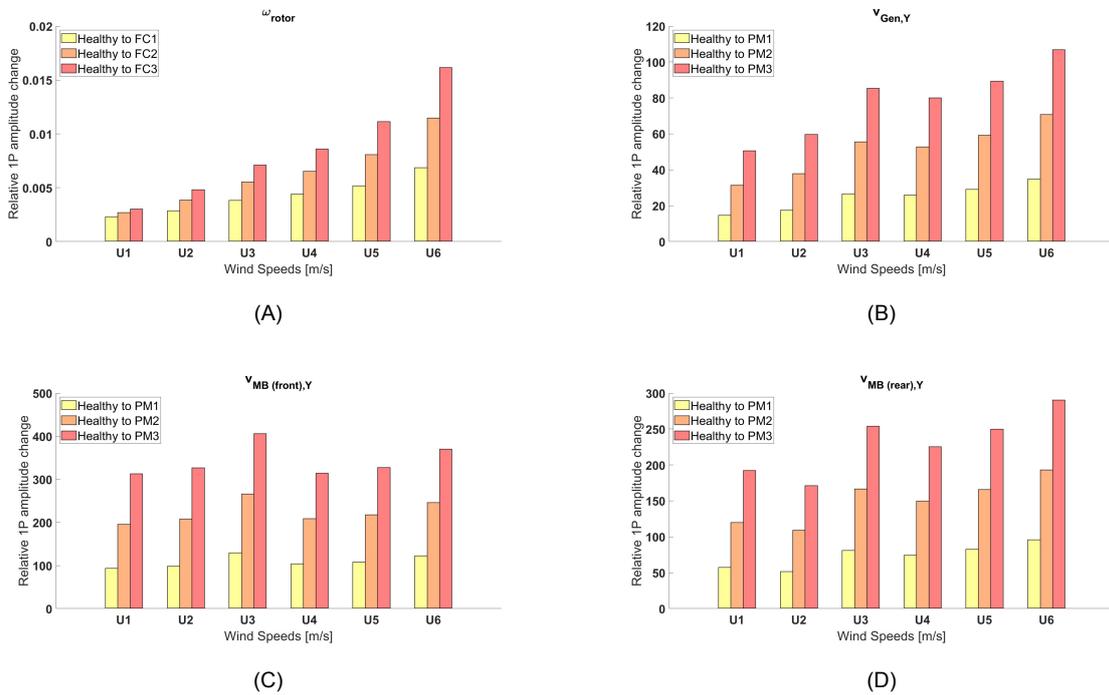


Figure C.5: PM - 1P: relative change averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $v_{Gen, \gamma}$; (C) $v_{MB (front), \gamma}$; (D) $v_{MB (rear), \gamma}$

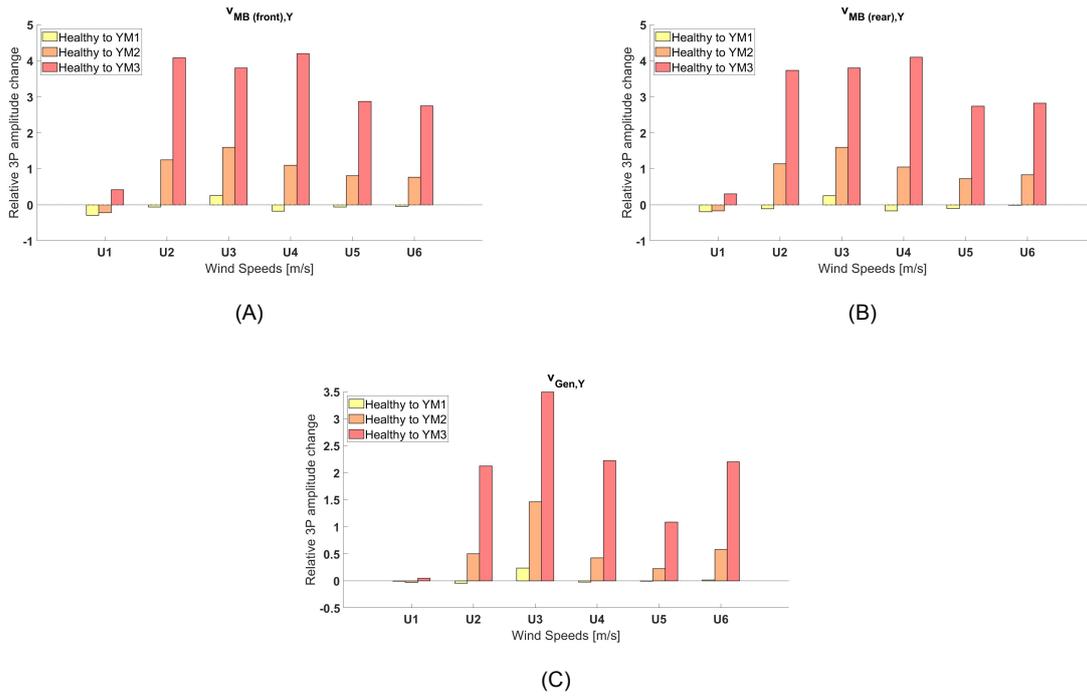


Figure C.6: YM - 3P: relative change averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) $V_{MB} (front), Y$; (B) $V_{MB} (rear), Y$; (C) V_{Gen}, Y

C.3. Box plots and Multiple Comparison analysis for MI vs PM imbalance cases

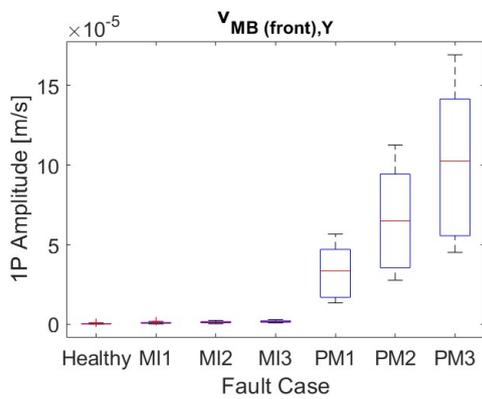


Figure C.7: Front main bearing boxplot

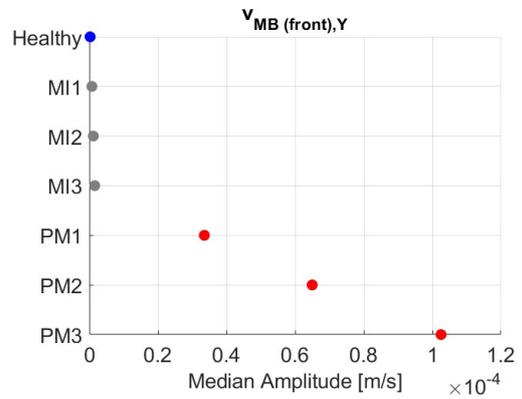


Figure C.8: Front main bearing multiple comparison test

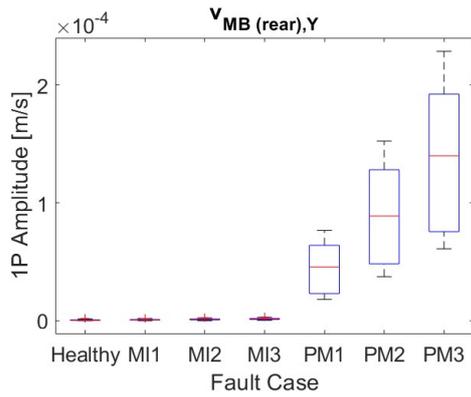


Figure C.9: Front main bearing boxplot

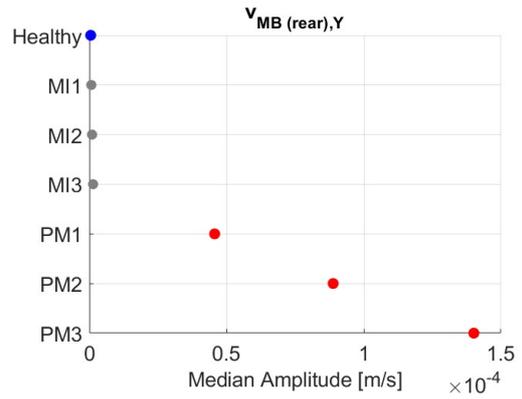


Figure C.10: Front main bearing multiple comparison test

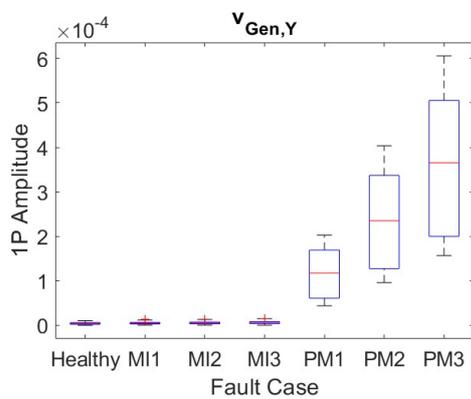


Figure C.11: Generator signal boxplot

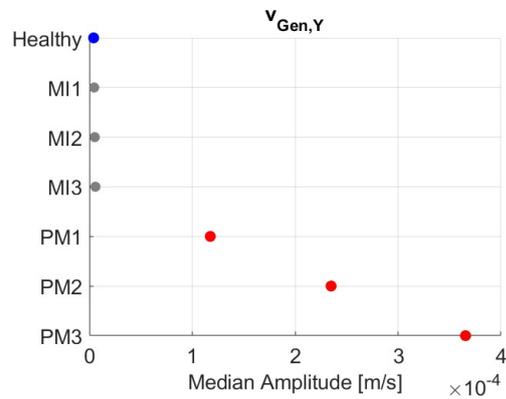


Figure C.12: Generator signal multiple comparison test

C.4. Time domain analysis for MI and YM

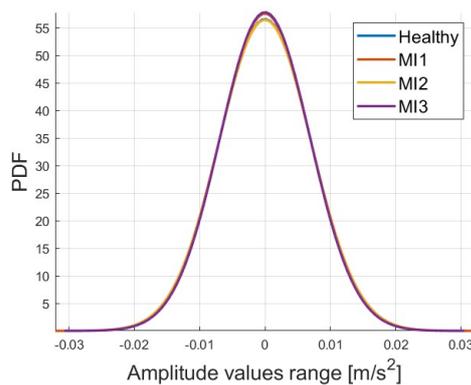


Figure C.13: PDFs of $a_{MB (front), Y}$ signals for healthy, MI1, MI2, MI3 conditions, at U_4

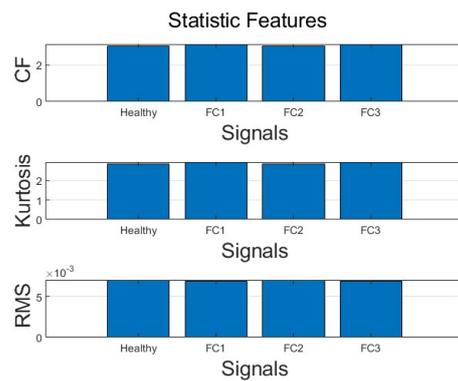


Figure C.14: Bar plots for (CI_{CF}) , $(CI_{Kurtosis})$ and (CI_{RMS}) for healthy vs MI1/MI2/MI3 faulty conditions

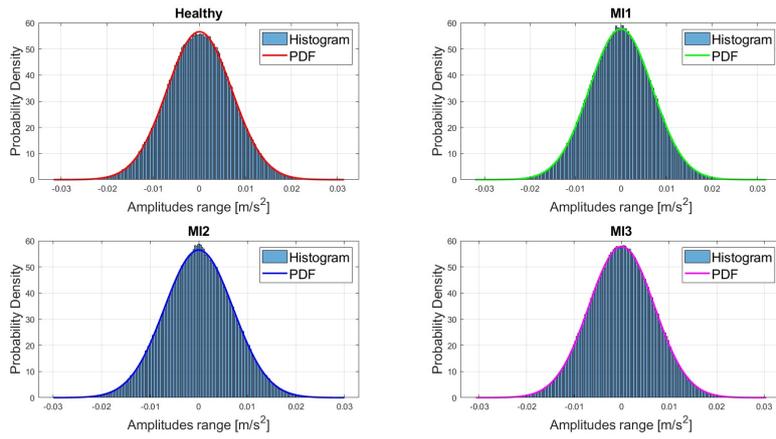


Figure C.15: Healthy vs MI1/MI2/MI3 - $a_{MB, (front), \gamma, U_4}$: histogram data fitting of the simulated dataset with respect to theoretical normal distribution.

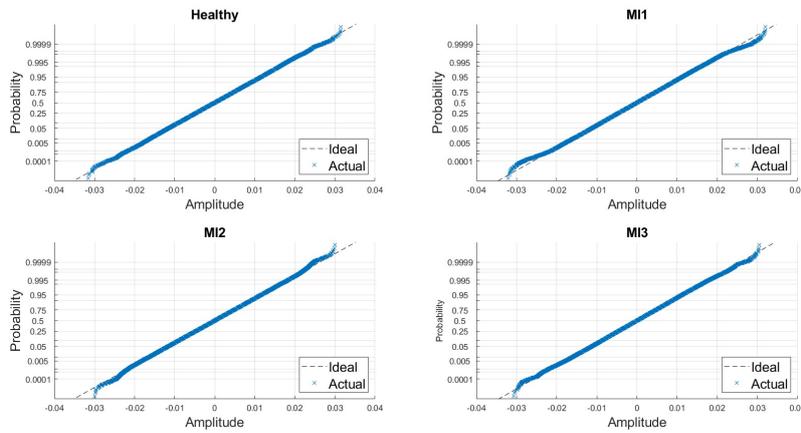


Figure C.16: Healthy vs MI1/MI2/MI3 - $a_{MB, (front), \gamma, U_4}$: NPPs for quantile comparison of simulated data with theoretical normal distribution.

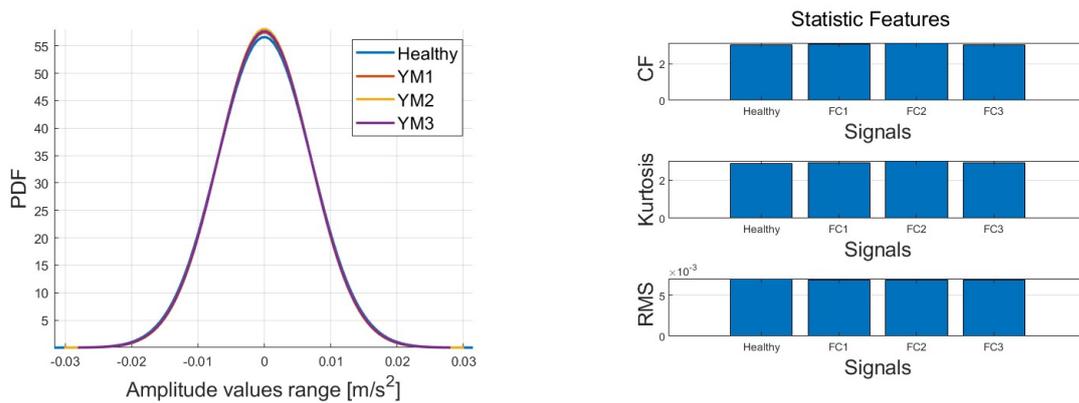


Figure C.17: PDFs of $a_{MB, (front), \gamma}$ signals for healthy, MI1, MI2, MI3 conditions, at U_4

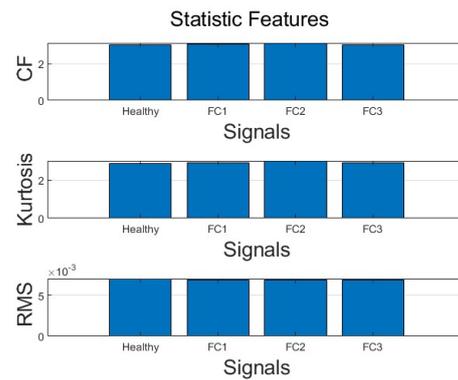


Figure C.18: Bar plots for (CI_{CF}) , $(CI_{Kurtosis})$ and (CI_{RMS}) for healthy vs YM1/YM2/YM3 faulty conditions

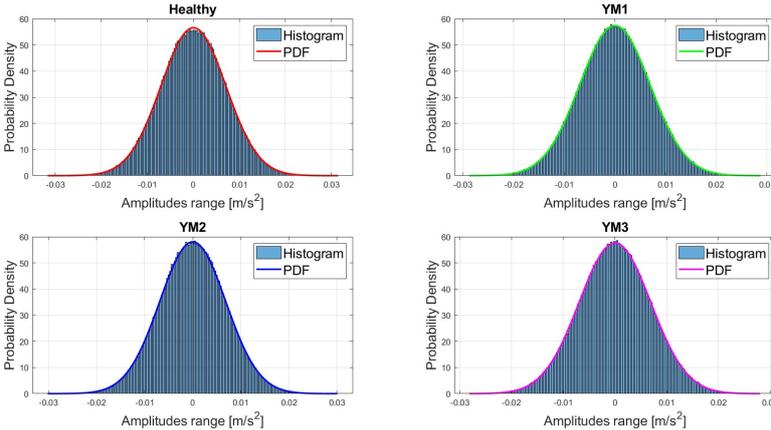


Figure C.19: Healthy vs YM1/YM2/YM3 - $a_{MB, (front)}$, γ , U_4 : histogram data fitting of the simulated dataset with respect to theoretical normal distribution.

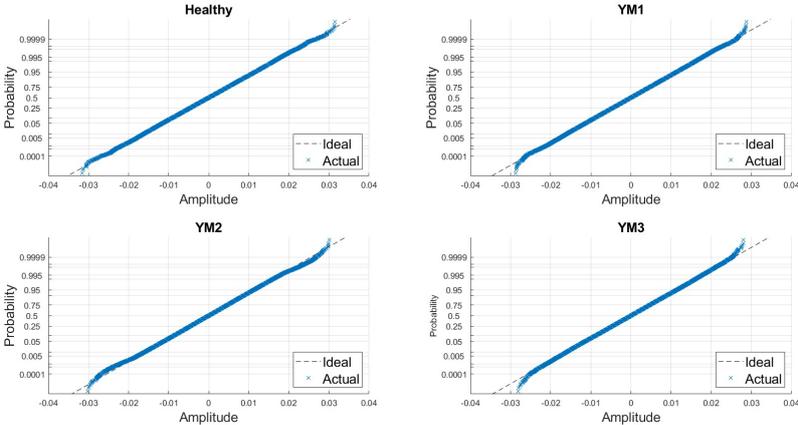


Figure C.20: Healthy vs YM1/YM2/YM3 - $a_{MB, (front)}$, γ , U_4 : NPPs for quantile comparison of simulated data with theoretical normal distribution.

D

Steady-state data distribution and statistical median value analysis for all wind speeds

D.1. Box plots amplitude distribution analysis

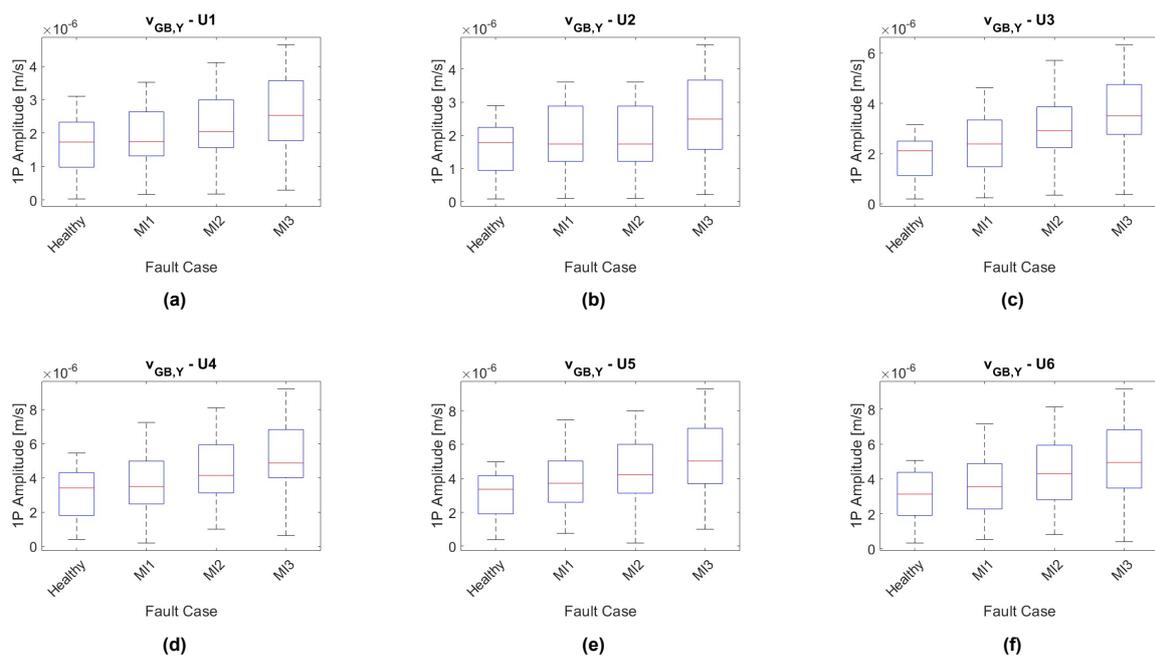


Figure D.1: Mass Imbalance 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds

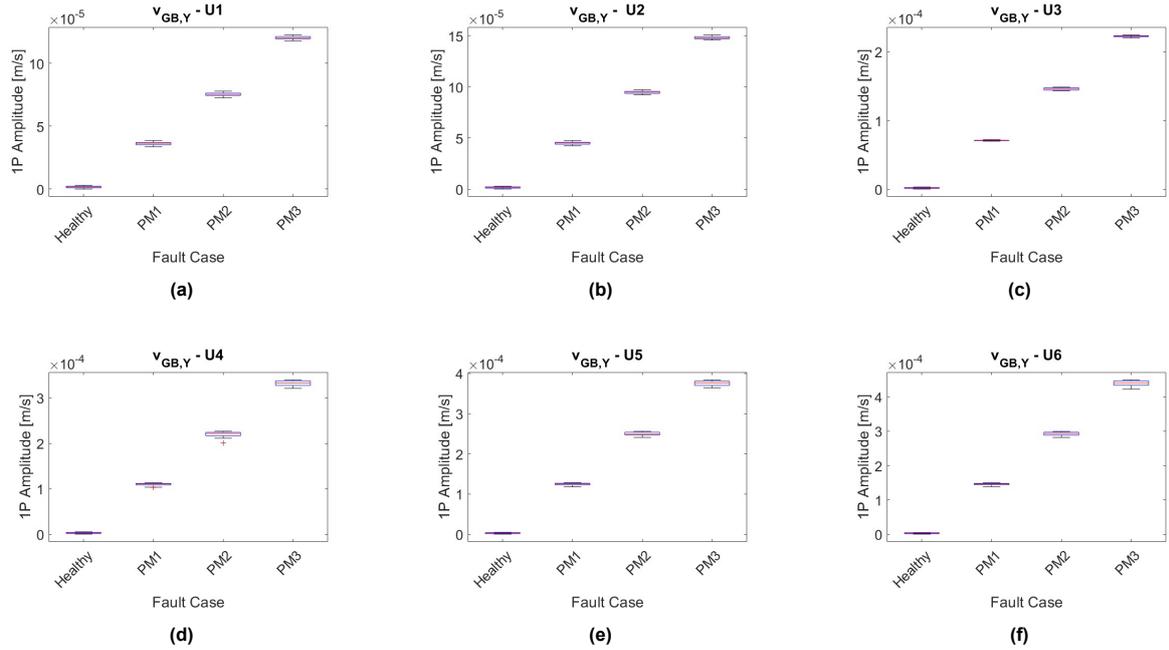


Figure D.2: Pitch misalignment 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds

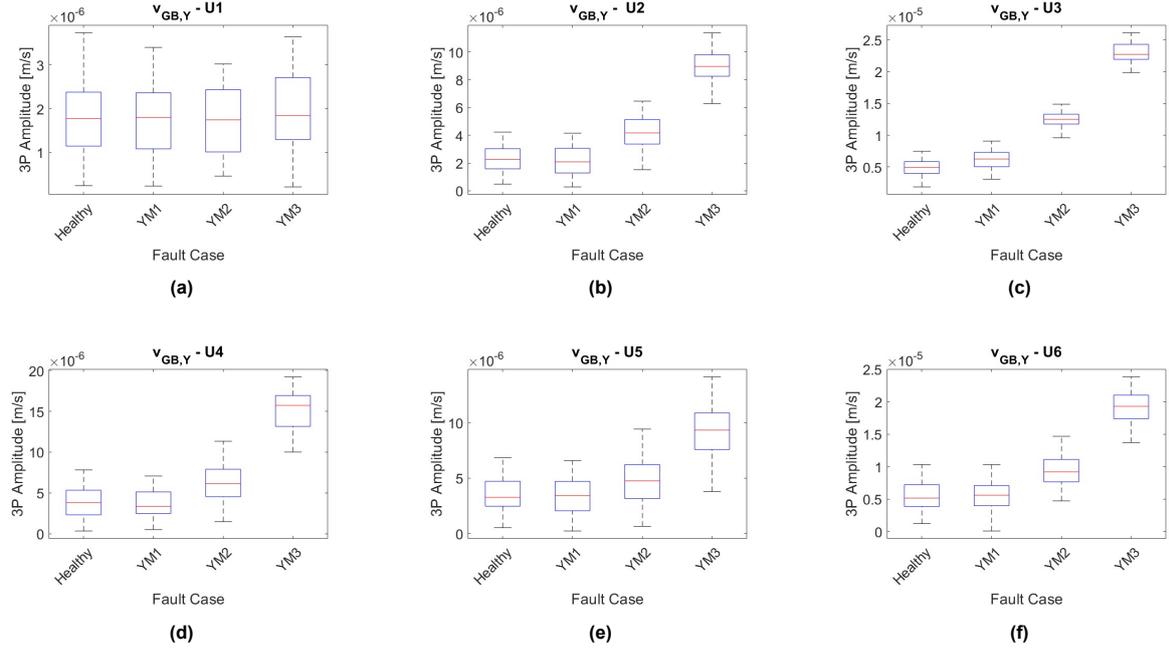


Figure D.3: Yaw misalignment 3P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds

D.2. Median values comparison analysis

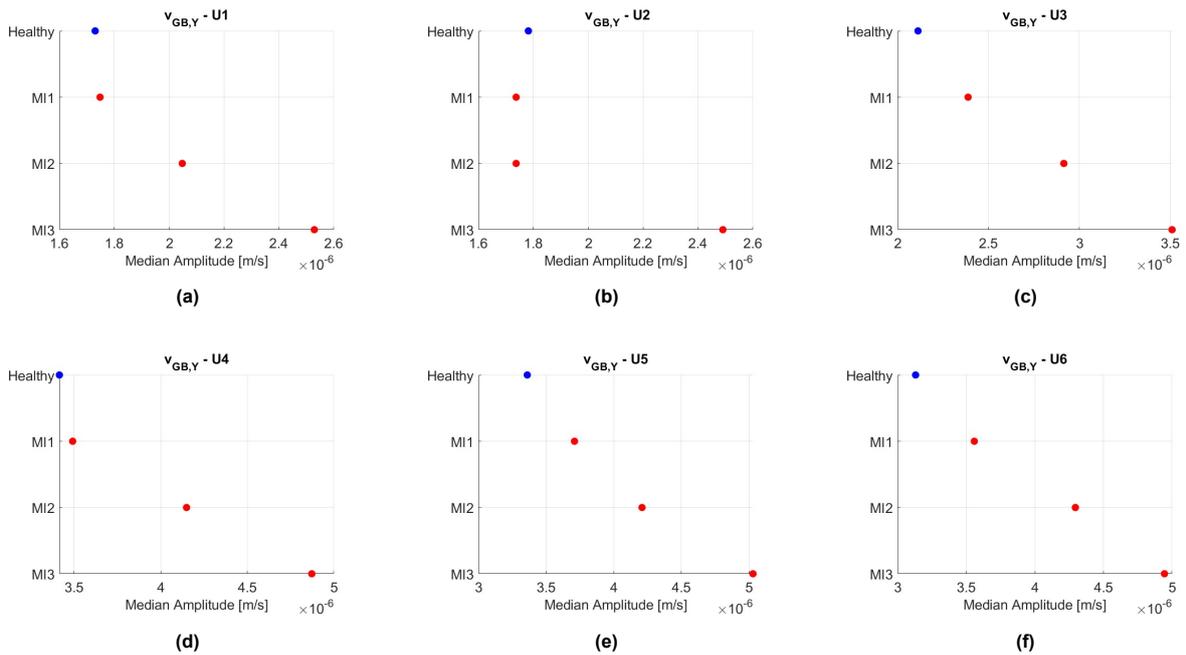


Figure D.4: Mass Imbalance 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds

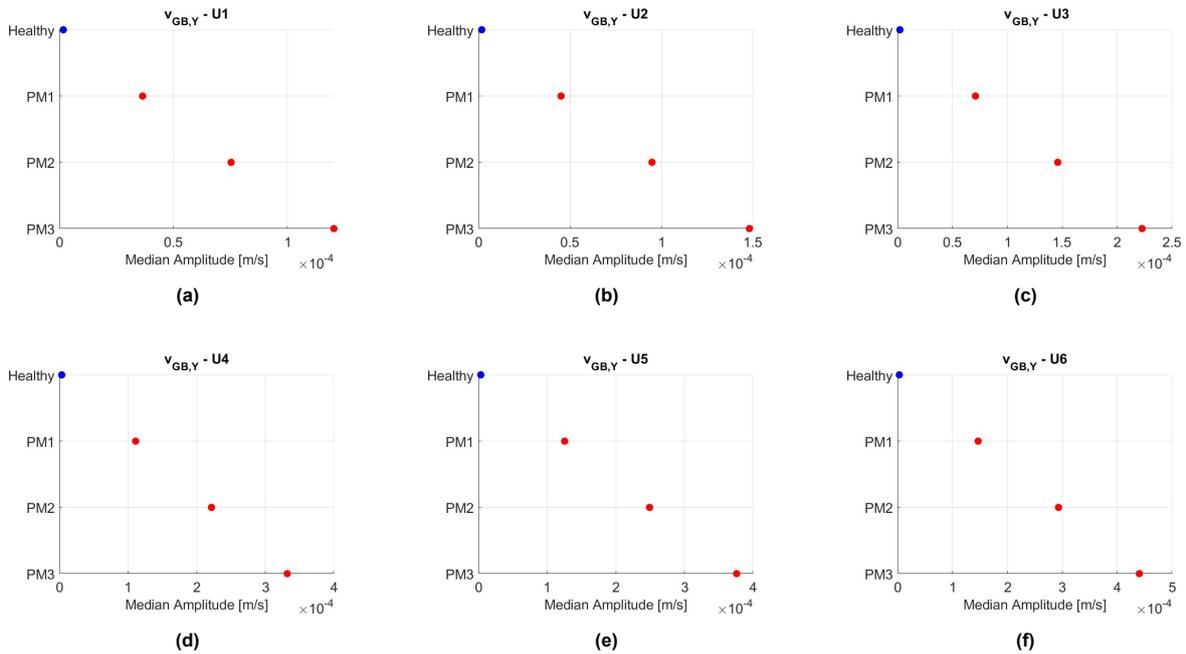


Figure D.5: Pitch misalignment 1P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds

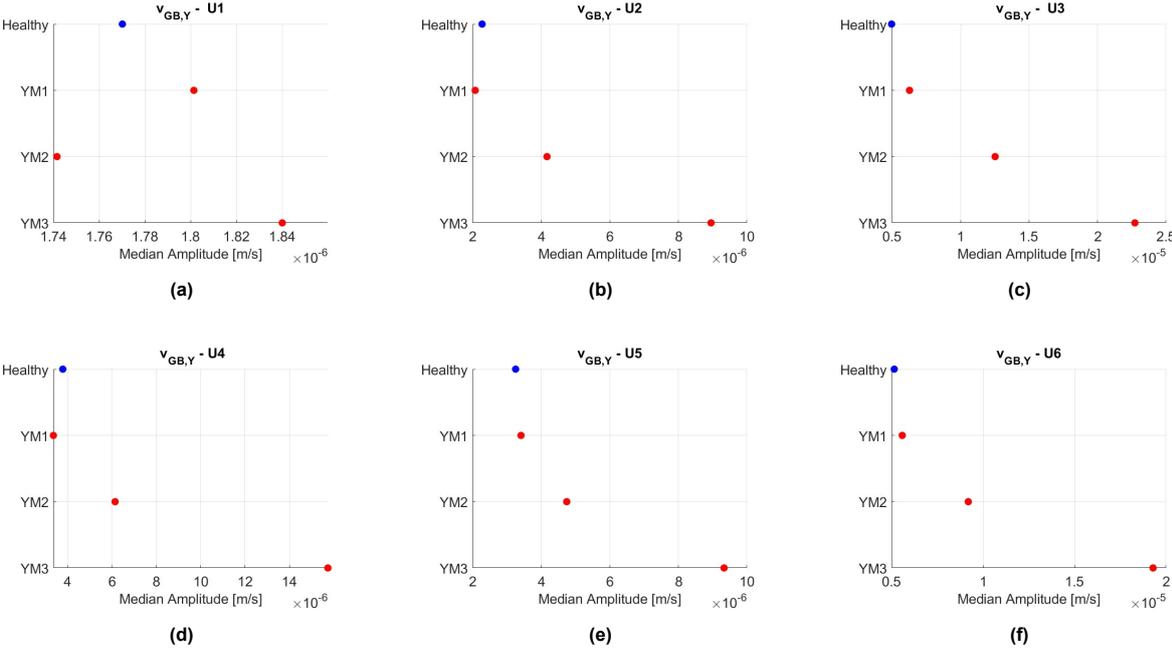


Figure D.6: Yaw misalignment 3P peak amplitudes distribution for gearbox housing velocity vibration signal across six selected wind speeds

E

Turbulence Analysis

E.1. 3D averaged peak amplitude comparison plot

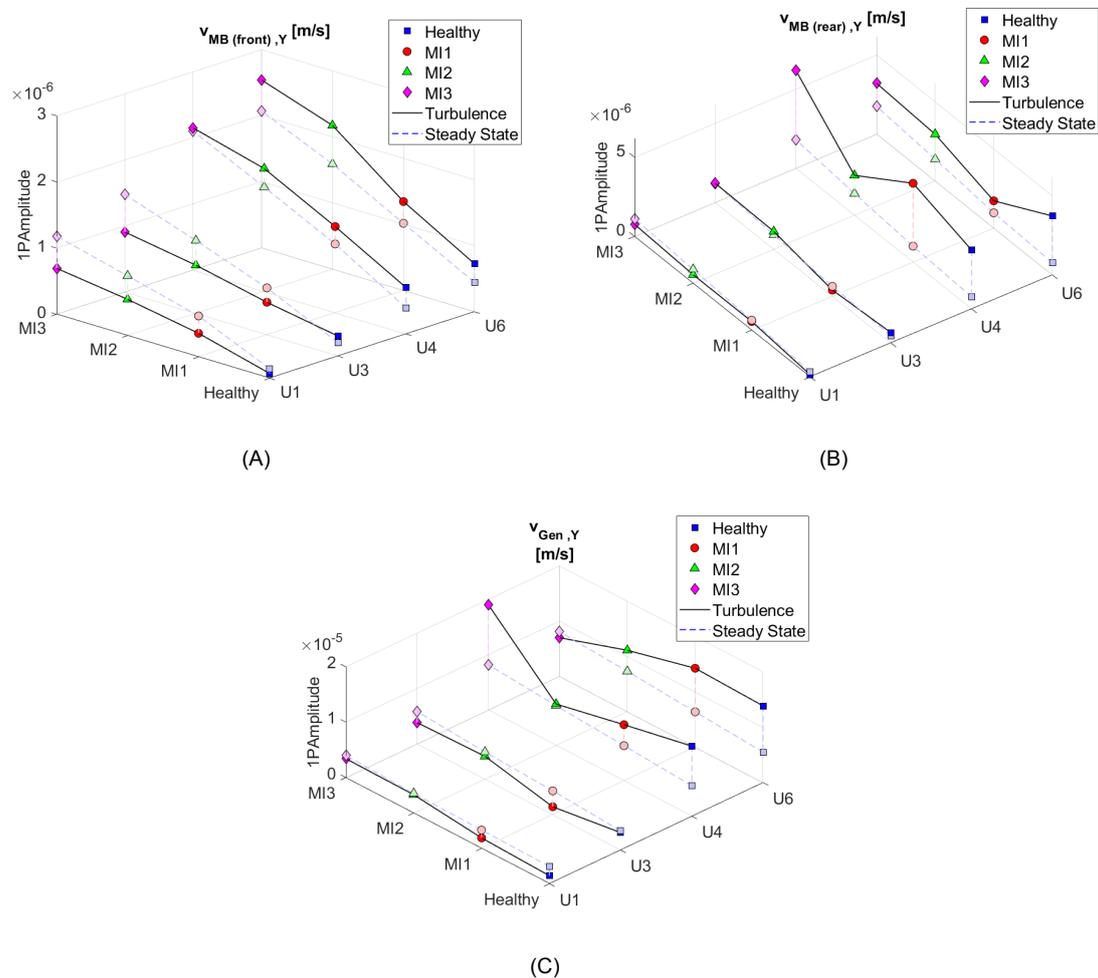


Figure E.1: MI - 1P: Averaged peak amplitudes comparison comparison under varying health states and wind speed conditions (A) $v_{MB} (front), Y$; (B) $v_{MB} (rear), Y$; (C) v_{Gen}, Y

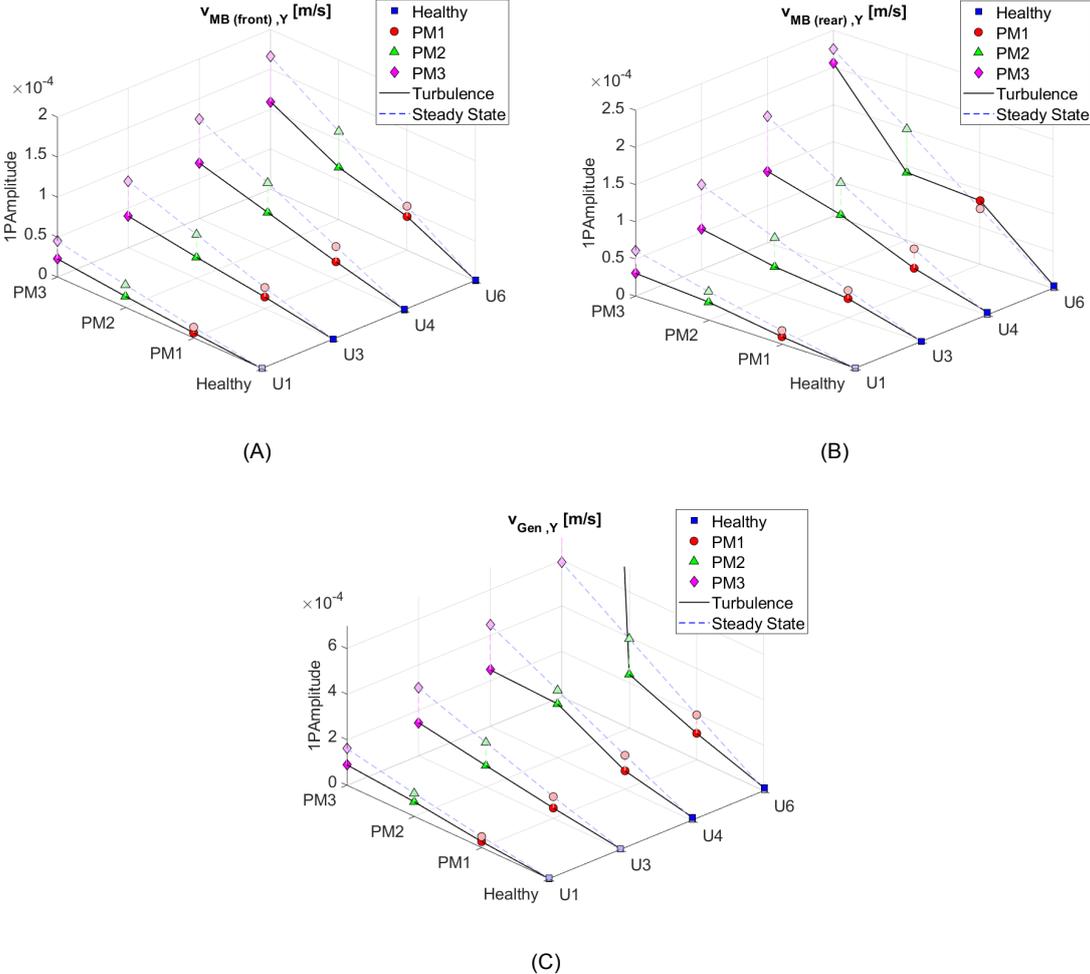


Figure E.2: PM - 1P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$

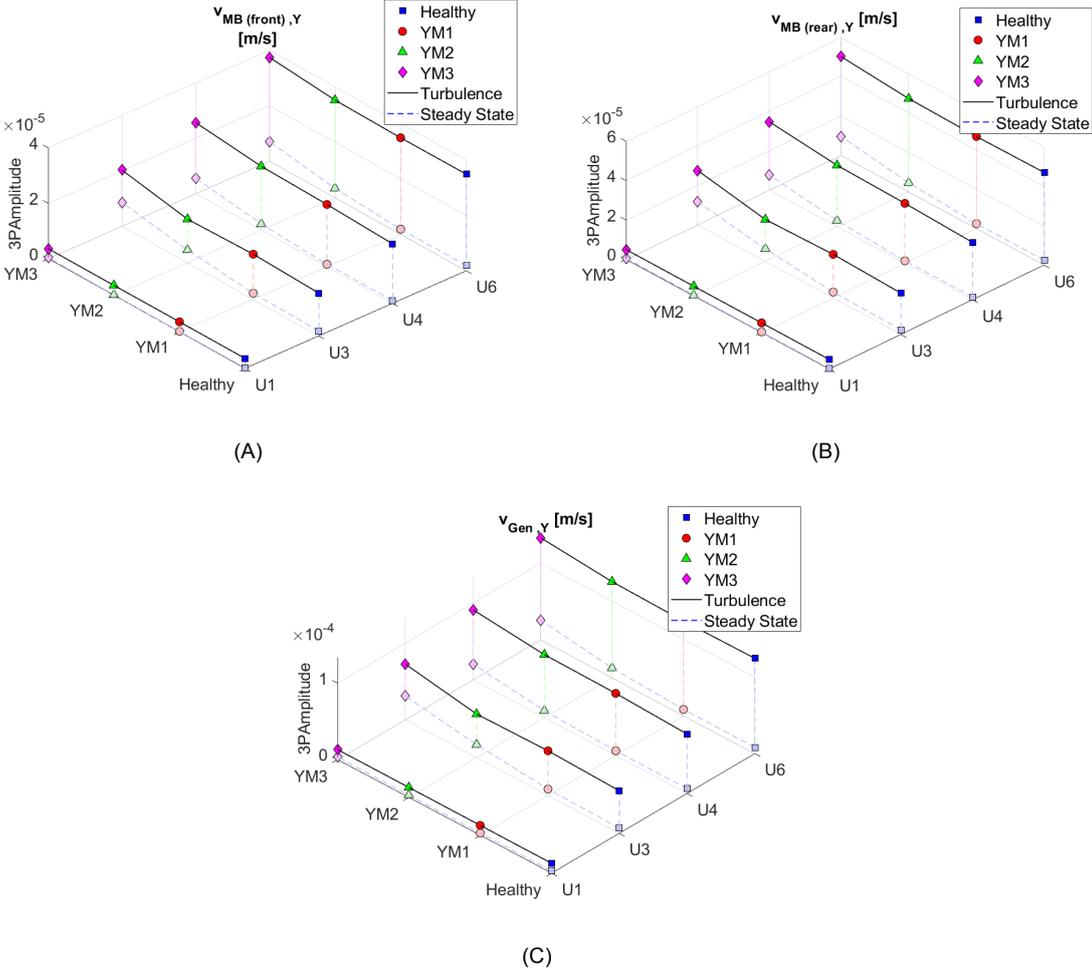


Figure E.3: YM - 3P: Averaged peak amplitudes under varying health states and wind speed conditions (A) $v_{MB (front), Y}$; (B) $v_{MB (rear), Y}$; (C) $v_{Gen, Y}$

E.2. Relative change detection histograms comparison

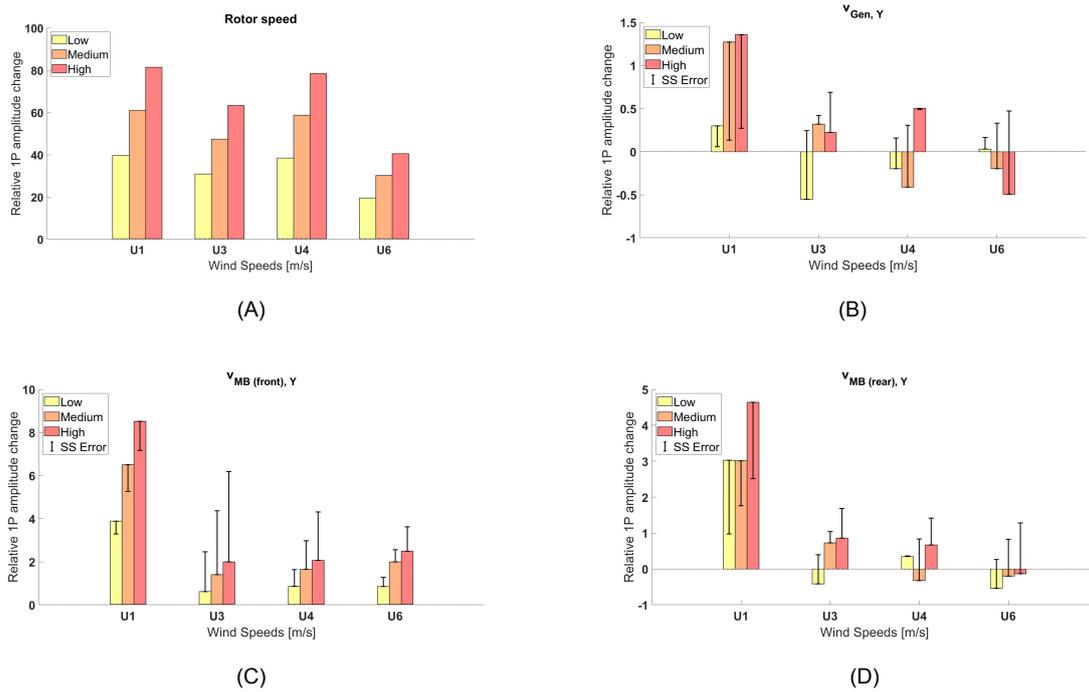


Figure E.4: MI - 1P: Relative change in averaged frequencies amplitudes for MI1/MI2/MI3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $V_{Gen, Y}$; (C) $V_{MB (front), Y}$; (D) $V_{MB (rear), Y}$

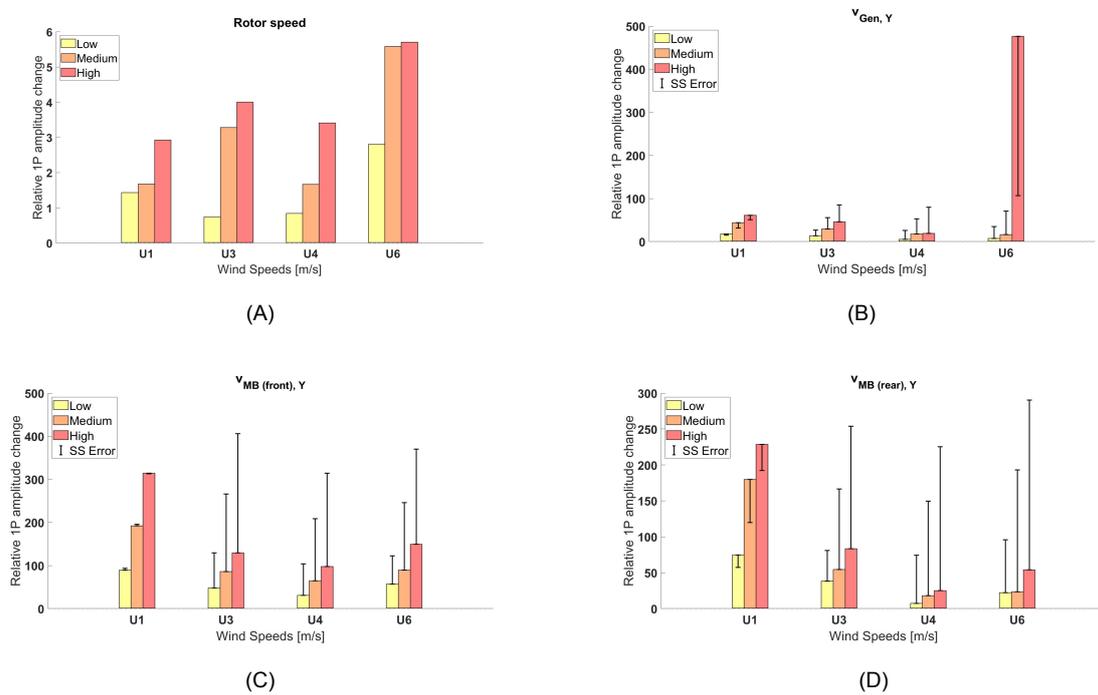


Figure E.5: PM - 1P: Relative change in averaged frequencies amplitudes for PM1/PM2/PM3 compared and normalized with respect to healthy conditions. (A) ω_{rotor} ; (B) $V_{Gen, Y}$; (C) $V_{MB (front), Y}$; (D) $V_{MB (rear), Y}$

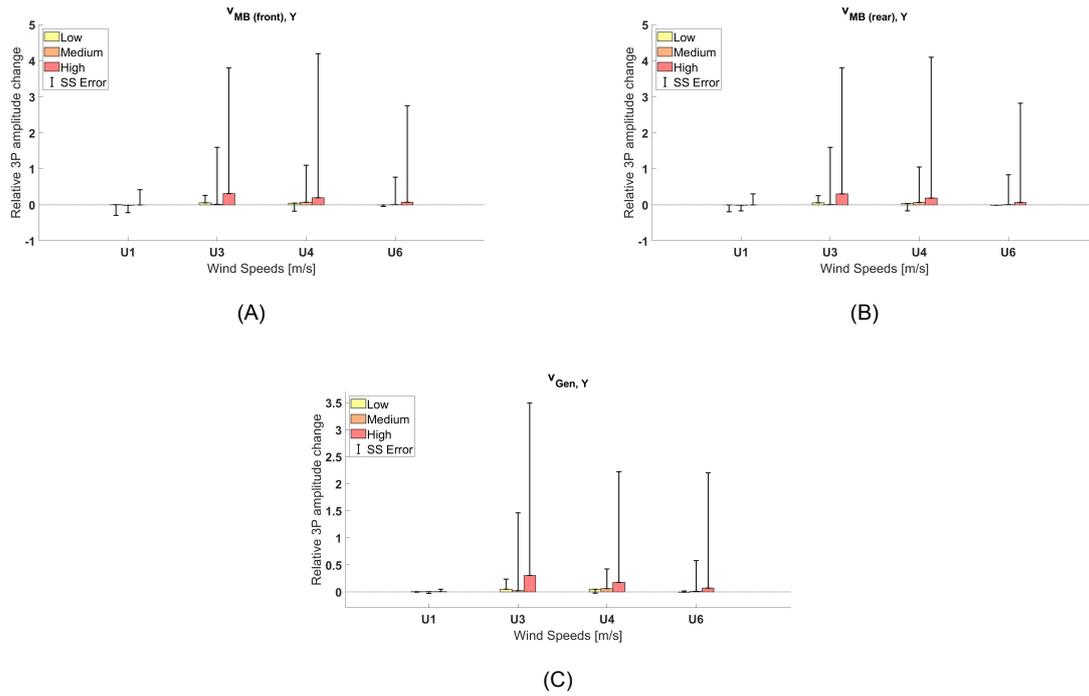


Figure E.6: YM - 3P: Relative change in averaged frequencies amplitudes for YM1/YM2/YM3 compared and normalized with respect to healthy conditions. (A) $v_{MB} (front), Y$; (B) $v_{MB} (rear), Y$; (C) v_{Gen}, Y

E.3. Turbulence correlation analysis

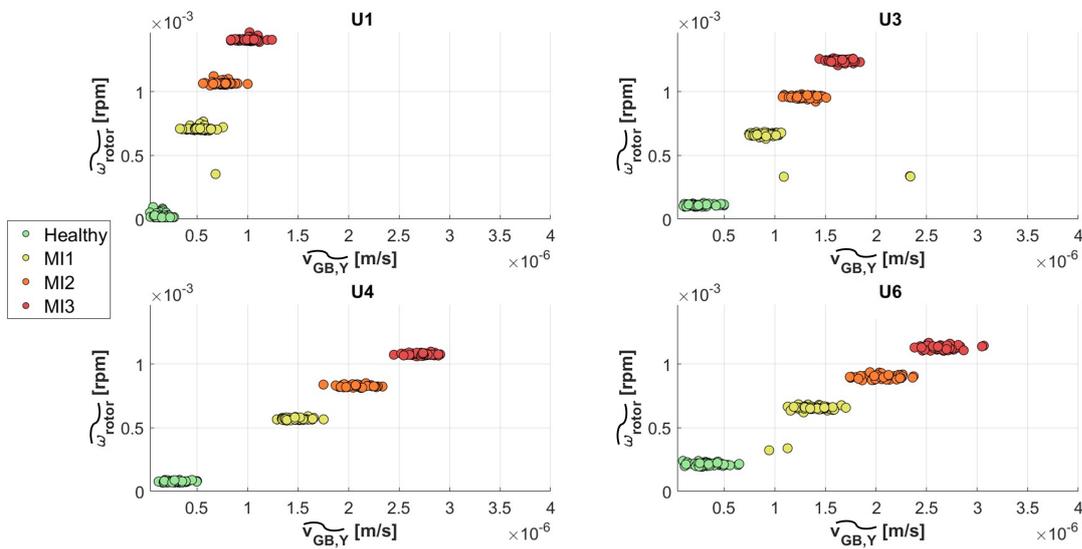


Figure E.7: Mass imbalance - 1P amplitudes scatter plots of $v_{GB,Y}, Y$, and ω_{rotor} signals. Each subplot resembles a single wind speed for different health conditions: (a) U1; (b) U3; (c) U4; (d) U6

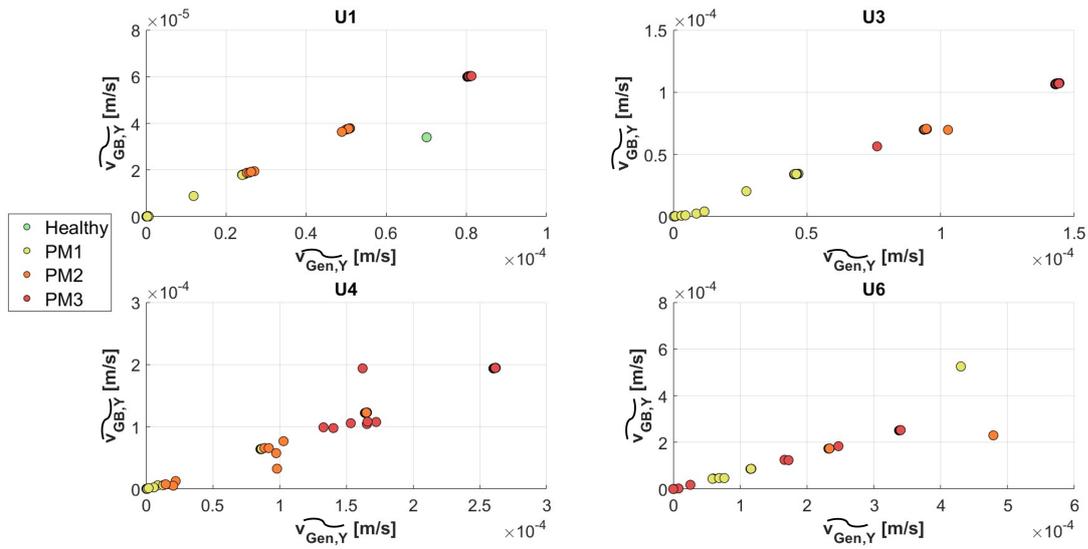


Figure E.8: Pitch misalignment- 1P amplitudes scatter plots of $v_{GB,Y}$, and $v_{Gen,Y}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U1; (b) U3; (c) U4; (d) U6

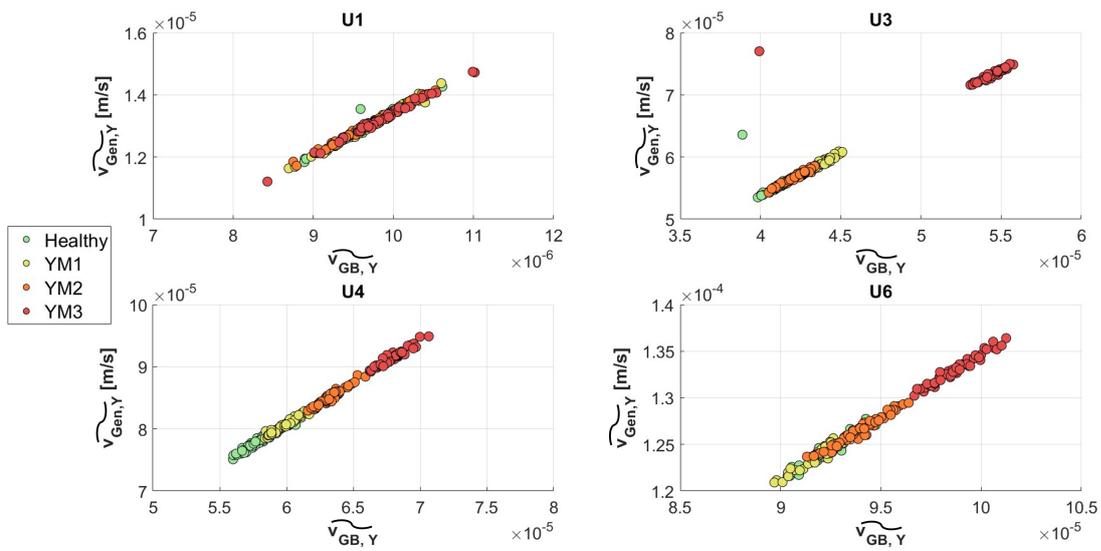


Figure E.9: Yaw misalignment- 1P amplitudes scatter plots of $v_{GB,Y}$, and $v_{Gen,Y}$ signals. Each subplot resembles a single wind speed for different health conditions: (a) U1; (b) U3; (c) U4; (d) U6