THE UNIVERSITY *of* EDINBURGH

Data-driven aerodynamic instabilities detection in centrifugal compressors

Mateusz Stajuda

School of Engineering

University of Edinburgh

A thesis submitted for the degree of

Doctor of Philosophy

2023

Abstract

Centrifugal compressors are machines of utmost importance in numerous industrial and high-tech applications. They are known to be prone to the appearance of aerodynamic instabilities at low mass flow rates, when operating close to peak performance. Instabilities are a number of flow structures that negatively impact the compressor. Their effects range from efficiency loss for inlet recirculation, through increased level of vibrations and risk of fatigue damage for rotating stall up to an abrupt machine destruction for surge.

Quick and accurate instabilities detection is a challenge. Detection of surge is often a top priority as it has the biggest consequences for the machine operation, however detecting other instabilities is also important for overall performance and long-time operability. A promising approach to detection is based on data-driven techniques, using high frequency signals sampled from the compressor to capture the dynamics of the system. Such approach could warn about the approaching onset of instability, providing ample of time for reaction. However, the signal is often composed of a number of overlapping sources and a considerable amount of noise, which makes it a challenge to extract the meaningful indication of instability. A valuable insight into the system state could be obtained if the sources and the noise were separated.

The aim of this thesis is to build an instabilities-detection methodology leveraging data-driven signal decomposition techniques. The goal is to use a pressure signal collected inside of the compressor and obtain a real-time indication of the compressor stability. Two distinct decomposition methods, Empirical mode decomposition (EMD) and singular spectrum analysis (SSA) are investigated for this purpose. The goal of each of the method is to provide components sensitive to the presence of individual instabilities to build instabilities-sensitive features. The features are combined in the feature space, dimensionality of which can be adjusted depending on the system under analysis and expected unstable conditions. Using the decomposition techniques it is possible to increase the dimensionality of a signal, enabling differentiation of different types of instabilities present in the signal that would otherwise provide an overlapping signature in the original signal.

The proposed methodology is validated with the data from a low-pressure industrial compressor, equipped with five high-frequency pressure transducers located along the flow path. The compressor was operated through a wide spectrum of conditions. In the post-processing, the data was divided into different

general conditions, being stable, locally unstable and globally unstable.

The results highlight the potential of defining robust features using both EMD and SSA for detecting general conditions, even with a relatively short input signal. The features are physically interpretable, and it is possible to provide meaningful thresholds for the detection of instabilities based solely on stable conditions. This is an important advantage, as operating the compressor in an unstable range brings risk of its damage. The overall accuracy of both methods is over 90%, with the majority of misclassifications coming from the region where the conditions transition from locally unstable to globally unstable.

For certain machines, the extension of the operating range at the expense of safety might be beneficial. The globally unstable conditions reported in the case study can be furtherly divided into transient and deep surge. It is shown that decoupling those two instabilities for a robust indication with either EMD or SSA is not fully possible, which may come from the physical character of each instability. The features values for unstable conditions have to be known to differentiate transient and surge, hence the benefit of relying solely on stable data is lost. Obtaining features sensitive to each instability requires a longer input signal and extended processing, which negatively affects the responsiveness of the detection system. To avoid such issue, it is possible to use a general condition feature. It also requires prior mapping, but a robust indication can be obtained with a short input signal.

The values of features obtained from the process show certain level of variability and tend to overlap due to noise present in the data. With a prior mapping needed for the detection of exact instabilities, a probabilistic approach to classification can be leveraged. Apart from classification, such approach provides an information about the probability of a given class, which can be used to define no-classification zones in the feature space, where the probability of each of the classes is low. It is shown that the application of probabilistic model provides comparable classification rate, but it can offer increased flexibility and limit the number of sensors to be used for detection.

The approach demonstrated in this thesis can enable better understanding of the compressor operating conditions in the proximity of the surge line. Consequently, it could be useful for ensuring that the machine can safely reach its peak performance, possibly extending its operating range for different conditions.

Acknowledgements

Throughout the period of my PhD studies, I have received a great deal of support and assistance, for which I am very grateful.

First and foremost, I would like to express my gratitude towards my supervisors, David García Cava and Grzegorz Liśkiewicz. My PhD journey would not have started without you, nor would it have been completed. I greatly appreciate all the support, encouragement and feedback I received. Having very different styles of work, you formed a perfect blend of supervision team, exerting a great impact on my studies and on me as a researcher. I am extremely grateful that I was able to have you as my supervisors.

I would like to extend my acknowledgments to my fellow PhD students, Callum, Artur, Dashty and all the others who I have met on this path. You made my PhD journey much more enjoyable, both academically and socially. I am happy to have been able to get to know you despite the strange times we lived through.

Last but not least, I would like to thank my family, especially my wife Karolina, for all the support and help throughout the uneasy period of my PhD. You were the source of joy, happiness and motivation in the difficult moments. I am truly grateful for all you have done and continue to do for me.

Declaration of Originality

I declare that I composed the the work presented in this thesis entitled "Data-driven aerodynamic instabilities detection in centrifugal compressors". Except in cases where indicated in the text, the presented work was completed in its entirety by me. Furthermore, I declare that none of the included work has been submitted for any other degree or professional qualification. Sections of the thesis that have already been published have been clearly noted in the text.

_____

Mateusz Michał Stajuda

Lay summary

Centrifugal compressors are widespread machines present in various industries as well as in everyday life. They play a crucial role in many chemical processing plants, can be found in heat pumps, combustion engines or home appliances.

A common feature of centrifugal compressors is that they can develop aerodynamic instabilities when operating close to the peak of performance and pressure ratio. Instabilities are unwanted flow structures that are detrimental to the compressor's performance and can threaten its structural integrity. Consequently, the compressors are controlled to prevent the onset of instabilities.

A common control approach is to set an operating limit that should not be crossed. Such a limit is placed at some distance from the actual onset of instabilities and does not rely on detecting them, but its location is defined beforehand by the manufacturer. This approach, although safe, unnecessarily limits the operating range of the compressor.

To allow the compressor to operate as close to the limit as possible, the presence of instabilities could be detected in real-time. By continuously probing the stability level of the compressor, the operating limit can be set as close to the actual limit as possible. The crucial part of this approach is a detection method that is sufficiently accurate and quick.

This thesis demonstrates that real-time detection can be achieved with signal decomposition techniques. By taking a pressure signal and separating it into components, the information about the presence of instabilities can be extracted. The components of the pressure signal are transformed into features that are used for making an indication of the presence of instabilities.

The decision about the presence of instabilities can be made based on a feature threshold value, however, this approach can be sensitive to singular events and noise. To further improve the detection, a probabilistic model can be implemented to take advantage of the uncertainty quantification. Consequently, a more informed decision about the compressor's stability can be made.

With the developments presented in this thesis, it should be possible to safely extend the operating range of the compressor using instabilities-sensitive features coupled with a probabilistic approach to the classification of conditions. This can be leveraged to propose a tailored control system capable of bringing important benefits to machine operation.

Publications list

Journal contributions

- **Stajuda, M.**, Cava, D. G., & Liśkiewicz, G. (2022). Aerodynamic instabilities detection via empirical mode decomposition in centrifugal compressors. Measurement, 199, 111496.
- **Stajuda, M.**, García Cava, D., & Liśkiewicz, G. (2022). Comparison of empirical mode decomposition and singular spectrum analysis for quick and robust detection of aerodynamic instabilities in centrifugal compressors. Sensors, 22(5), 2063.
- Liśkiewicz, G., Kabalyk, K., Jaeschke, A., Grapow, F., Kulak, M., **Stajuda, M.**, & Kryłłowicz, W. (2020). Unstable flow structures present at different rotational velocities of the centrifugal compressor. Energies, 13(16), 4146.

Peer-reviewed conference contributions

- **Stajuda, M.**, D. Garcıa Cava, & G. Liskiewicz. (2022) "Statistically enhanced classification of centrifugal compressor operating condition." 30th International Conference on Noise and Vibration Engineering.
- **Stajuda, M.**, Cava, D. G., & Liśkiewicz, G. (2021). Evaluation of EMD and SSA sensitivity for efficient detection of aerodynamic instabilities in centrifugal compressors. In 2021 Signal Processing Symposium (SPSympo) (pp. 258-263). IEEE.
- **Stajuda, M.**, Garcia Cava, D., & Liśkiewicz, G. (2020). Cyclostationary Approach for Instabilities Detection and Condition Monitoring of Centrifugal Compressor. In Turbo Expo: Power for Land, Sea, and Air (Vol. 84102, p. V02ET41A032). American Society of Mechanical Engineers.
- **Stajuda, M.**, Liskiewicz, G., & Garcia, D. (2019). Flow instabilities detection in centrifugal blower using empirical mode decomposition. Proceedings of the Global Power & Propulsion Society.

List of Tables

1.1 Data-driven methods for detection of aerodynamic instabilities 52

3.1 Operating parameters and dimensions of the test rig . 92

3.2 Working regimes identified in the signal obtained with quasi-dynamic protocol 100

3.3 Working regimes identified in the signal obtained with the dynamic protocol 101

3.4 Number of data points available for different conditions in thousands 102

7.1 Selected kernels and their hyperparameters 202

List of Figures

1.1 Schematic representation of a centrifugal compressor with marked sections; 0: compressor system inlet; 1: impeller leading edge; 2: impeller trailing edge; 3: end of the diffuser; 4: end of the volute; 5: compressor system outlet 3

1.2 Relation of size and rotational speed for different types of centrifugal compressors [1] 4

1.3 Variations of impeller designs a) shrouded and unshrouded impellers; b) impeller with 2D and 3D blades 5

1.4 Variations of impeller designs dependent on the flow coefficient 5

1.5 Exemplary compressor map, based on [2] 6

1.6 Typical inlet recirculation zone in the inducer and impeller; based on [3] 11

1.7 Stalled airfoil [4] 15

1.8 Impeller rotating stall evolution; numbers represent the temporal evolution of the flow in the same space between the blades [5] 16

1.9 Schematic depiction of Greitzer compressor model as an analogue to a mechanical system; adapted from [6] 21

1.10 Visualisation of frequency overlap in detection of surge and rotating stall 25

1.11 Visualisation of a blade pass character change for unstable conditions; depending on the instability, the change can be different both in terms of the character and the frequency of the occurrence 29

1.12 Jet and wake structure due to separation of the flow on a suction side [7] 30

1.13 Compressor map with instabilities detected for different rotational speed; based on [8] 32

1.14 Feature values for different number of classes to be differentiated with a single feature a) two conditions; b) three conditions; differentiating three conditions with a single feature requires establishing two thresholds, marked with red line 58

1.15 Distribution of the features for two different conditions a) dispersion of the points with x showing the mean; b) box plot related to the distribution; the mean values for the clusters differ, but their distributions overlap in the feature space, investigating single points from each class it is possible to find such that their values differ, but it does not ensure accurate classification 59

1.16 Conceptual representation of boundary types for 2-dimensional feature space and binary classification a) linear boundary b) non-linear boundary 61

1.17 Depiction of a rejection zone concept for two classes classification; the rejection zone is present where the difference in class c_1 probability $p(c_1)$ and class c_2 probability $p(c_2)$ is lower than a certain threshold value t 63

2.1 Overview of the general methodology steps; steps marked with colors are furtherly explained in detail 73

2.2 Details of the feature extraction step 74

2.3 Details of the decision-making step 75

2.4 EMD process visualisation based on an exemplary signal a) The data along with upper and lower envelopes defined by the local maxima and minima, respectively, and the mean value of theenvelopes b) proto-IMF h_1 79

2.5 EMD process visualisation based on an exemplary signal a) IMF 1 obtained after 8 iterations b) original signal $x(t)$ and the first residual r_1 after subtracting IMF 1 from the original signal 80

3.1 Compressor cross-section with marking of its most important parts [9]; A - inlet pipe, B - Witoszynski nozzle, C - impeller, D - diffuser, E - volute, F - outlet pipe 91

3.2 Elements of the test rig [9] a) experimental rig impeller and volute; b) pressure taping; c) throttling valve 93

3.3 Performance measurements of the compressor [9] a) performance map (*phi* - mass flow coefficient, ψ - pressure raise coefficient); b) mass flow rate relation to throttling 93

3.4 Compressor section with gauges location [9] 95

3.5 Pressure data in relation to TOA level for different sensors collected in quasi-dynamic protocol [9] (a) Inlet p_{s-in}; (b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out} . 96

3.6 Pressure data in relation to TOA level for different sensors collected in a dynamic protocol [9] (a) Inlet p_{s-in}; (b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out} . . 97

4.1 RMS values for different IMFs at different operating conditions and sensor locations in the experimental rig demonstrating differences in IMFs response to unstable structures in the compressor; a) Inlet p_{s-in}; b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out} 105

4.2 Mean value of RMS and 90% confidence interval for selected IMFs from p_{s-imp1} sensor before impeller a) IMF 6; b) IMF 11 106

4.3 Changes in the RMS distribution of IMFs for a varying number of sifting iterations in EMD procedure for p_{s_imp1} sensor; a) surface plot for 16 sifting iterations b) surface plot for 32 sifting iterations; c) contour plot for 16 iterations; d) contour plot for 32 sifting iterations 109

4.4 Changes in mean and confidence interval of IMFs RMS for p_{s-imp1} sensor with varying number of sifting iterations; a) IMFs capturing the presence of local instability; b) IMFs capturing the presence of global instability 110

4.5 Overview of RMS distribution of IMFs for all operating conditions using varying input signal length N_s for p_{s-imp1} sensor before impeller a) N_s = 5,000; b) N_s = 20,000; c) N_s = 50,000; d) N_s = 100,000 . . . 111

4.6 Changes in mean and confidence interval of IMFs RMS for p_{s_imp1} sensor before impeller with varying input signal length for selected IMFs a) IMF 6; b) IMF 10 . 112

4.7 Time of processing a signal of different length through EMD for different number of sifting iterations; the dashed black line is an online processing limit above which the method no longer can provide real-time indication 114

4.8 RMS of IMFs for all sensors and all operating conditions; a) IMF 4; b) IMF 5; b) IMF 6; c) IMF 7 116

4.9 Contours of averaged spectral content of IMF 6 for a) sensor p_{s-imp1} before impeller; b) sensor p_{s-imp2} at impeller centre 118

4.10 Averaged spectral content of IMF 6 for a) sensor p_{s-imp1} before impeller; b) sensor p_{s-imp2} at impeller centre for different operating conditions 119

4.11 Mean and confidence interval of RMS for IMF 6 at before impeller sensor for different signal lengths; a) $N_s = 5{,}000$; b) $N_s = 10{,}000$; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation 120

4.12 Accuracy, in percent, for the features based on different IMFs and using different input signal length . . . 121

4.13 RMS of IMFs for all sensors and all operating conditions; a) IMF 9; b) IMF 10; c) IMF 11; d) IMF 12 122

4.14 Mean and confidence interval for selected IMF obtained for data from selected sensors a) IMF 9, p_{s-in}; b) IMF 10, p_{s-in}; c) IMF 9, p_{s-out}; d) IMF 10 , p_{s-out}; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation 124

4.15 Accuracy of detection, given in percent, for selected IMFs and input signal lengths a) p_{s-in}; b) p_{s-out} . . . 126

4.16 Averaged spectral content of IMF 9 for different operating conditions and sensors a) p_{s-in}; b) p_{s-out} . . . 127

4.17 Spectral content of selected high IMFs for three consecutive input signals obtained for $N_s = 10{,}000$ at TOA = 10% 128

4.18 Sum of selected IMFs RMS for signal length $N_S = 10{,}000$ samples; a) IMFs from 8 to 12; b) IMFs from 8 to 15 129

4.19 Accuracy of globally unstable conditions detection with summation of IMFs obtained for signal of input length $N_s = 10{,}000$ for inlet sensor p_{s-in} 130

4.20 Feature space representation of compressor data using μ_L and μ_G features to capture general operating conditions of a compressor; thresholds T_{μ_L} and T_{μ_G} defined based on stable operating conditions 132

4.21 RMS of IMFs for all sensors and all operating conditions; a) IMF 7; b) IMF 8 134

4.22 Mean and confidence interval of RMS for IMF 8 at inlet sensor for different signal lengths; a) N_s = 10,000; b) N_s = 50,000; c) N_s = 100,000; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation 135

4.23 Averaged spectral content of IMF 8 for p_{s-in} sensor a) contour for all operating conditions; b) line plots for selected TOAs 136

4.24 Mean and confidence interval of RMS for IMF 11 at inlet sensor for different signal lengths; a) N_s = 10,000; b) N_s = 50,000; c) N_s = 100,000 137

4.25 Contour of mean frequency spectrum of IMF 11 from selected sensors a) p_{s-out} sensor; b) p_{s-in} sensor . . . 138

4.26 Feature space representation of compressor data using μ_L and μ_G features to capture exact instabilities present in the compressor system; scatter points with black edge represent peaks of conditions, transparent points represent remaining data 140

5.1 Relative RMS of RCs at different operating conditions and sensor locations in the experimental rig demonstrating differences in RCs response to unstable structures in the compressor; a) Inlet p_{s-in}; b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out} 145

5.2 RMS of RCs for all sensors and all operating conditions demonstrating their response to different unstable flow structures; a) RC 1; b) RC 2; c) RC 3; d) RC 4 147

5.3 Mean and confidence interval of the RMS of selected RCs from p_{s-imp1} sensor before impeller a) RC 1; b) RC 2 148

5.4 Relative RMS of selected RCs from p_{s-imp1} sensor before impeller for different lengths of the input signal; a) N_s = 5,000; b) N_s = 100,000 150

5.5 Mean and dispersion of RMS for selected RCs obtained using different input signal length N_s for p_{s-imp1} sensor before impeller a) RC 1; b) RC 2 150

5.6 Relative RMS of RCs for p_{s-imp1} sensor before impeller for different window lengths; a) L = 100; b) L = 200; c) L = 500; d) L = 1000 152

5.7 Mean and dispersion of RMS of selected RCs for p_{s-imp1} sensor before impeller for different window lengths L a) RC 1; b) RC 2; solid line represents the mean value, dashed line demonstrates the confidence interval of 90 % 153

5.8 Time of SSA decomposition depending on the input signal length N_s and window length L for obtaining different number of RCs a) 15 RCs; b) 2 RCs; the dashed black line is an online processing limit above which the method no longer can provide real-time indication 154

5.9 RMS of selected RCs for all operating conditions obtained for sensors a) p_{s-imp1}; b) p_{s-imp2} 156

5.10 Mean and confidence interval of RC 2 obtained for sensor p_{s-imp1} for different input signal length a) N_s = 1,000; b) N_s = 5,000; c) N_s = 10,000; the red horizontal line represents a threshold defined as $Q(0.995)$ of data from compressor stable operation 158

5.11 Contour of averaged spectral content of selected components of RC 2 obtained from different sensors a) p_{s-imp1}; b) p_{s-imp2} 159

5.12 Accuracy of local instability detection with selected RCs obtained for different window lengths 160

5.13 RMS of RC 1 for all sensors and conditions a) full range; b) limited range to highlight the changes in locally unstable region 161

5.14 Mean and confidence interval of RMS of RC 1 obtained for window length $L = 50$ from selected sensors a) p_{s-in} sensor at the inlet; b) p_{s-out} sensor at the outlet; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation 162

5.15 Contour of the mean frequency spectrum of RC 1 for all operating conditions for selected sensors and window lengths a) p_{s-in}, $L = 50$; b) p_{s-out}, $L = 50$; . . . 162

5.16 Accuracy of indication based on threshold derived from stable conditions for a) p_{s-in}; b) p_{s-out} 163

5.17 Feature space representation of compressor data using μ_L and μ_G features with thresholds T_{μ_L} and T_{μ_G} defined based on stable operating conditions 165

5.18 RMS of RC 2 for all sensors and all operating conditions for different window lengths; a) $L = 200$; b) $L = 500$; 167

5.19 RMS of RC 2 for all sensors and all operating conditions for different window lengths; a) $L = 200$; b) $L = 500$; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation 168

5.20 Contour of mean frequency spectrum of RC 2 for all operating conditions for p_{s-in} sensor and different window lengths a) $L = 200$; b) $L = 500$; 169

5.21 Mean and confidence interval od RMS of RC 1 obtained for selected sensors with different window length a) p_{s-in}, $L = 200$; b) p_{s-in}, $L = 1000$; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation 170

5.22 Contour of mean frequency spectrum of RC 1 for all operating conditions for p_{s-in} sensor and different window lengths a) $L = 200$; b) $L = 1000$ 171

5.23 Feature space representation of compressor data using μ_L and μ_G features to capture exact instabilities present in the compressor system; scatter points with black edge represent peaks of conditions, transparent points represent all data 172

6.1 Features sensitive to locally unstable conditions obtained with different decomposition methods a) EMD; b) SSA; the red horizontal line represents a threshold defined as $Q(0.995)$ of data from compressor stable operation 178

6.2 Contour of an averaged spectral content of μ_L feature obtained with different decomposition methods a) EMD; b) SSA 178

6.3 Averaged spectral content of μ_L feature for selected operating conditions, obtained with different decomposition methods a) EMD; b) SSA 179

6.4 Features sensitive to globally unstable conditions obtained with different decomposition methods a) EMD; b) SSA; the red horizontal line represents a threshold defined as $Q(0.995)$ of data from compressor stable operation 180

6.5 Contour of an averaged spectral content of μ_G feature obtained with different decomposition methods a) EMD; b) SSA 180

6.6 Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 5% obtained with different decomposition methods a) EMD; b) SSA . . 181

6.7 Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 10% obtained with different decomposition methods a) EMD; b) SSA . . 182

6.8 Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 20% obtained with different decomposition methods a) EMD; b) SSA . . 183

6.9 Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 30% obtained with different decomposition methods a) EMD; b) SSA . . 183

6.10 Representation of changes in the components for stable operation at TOA 30 % for 100 input subsequent signals; a) μ_G feature; b) spectral content of μ_G . . . 184

6.11 Feature space representation of compressor data using μ_L and μ_G features obtained from SSA decomposition; scatter points with black edge represent the dynamic data 186

6.12 Feature space representation of compressor data using μ_L and μ_G features obtained from EMD decomposition; scatter points with black edge represent the dynamic data 187

6.13 Confusion matrix between the conditions using features based on different decomposition methods a) SSA and b) EMD 188

6.14 Classification of data for different TOAs based on feature space obtained using SSA-based features scattered in Figure 6.11 189

6.15 Classification of data for different TOAs based on feature space obtained using EMD-based features scattered in Figure 6.12 190

6.16 Class detection based on the feature space representation and thresholds for EMD and SSA 191

6.17 Timing of the methods for different input length, decomposition parameters and number of components; dashed black line is an online processing limit above which the method no longer can provide real-time indication 194

6.18 Feature space representation using only components obtained from sensor p_{s-imp1} before impeller a) SSA-based features b) EMD-based features 197

7.1 Classification of one-dimensional data with GPC using different kernels a) DP, kernel; b) RBF kernel c) DP^2 kernel d) RQ kernel 205

7.2 Comparison of probability prediction with GPC model for different kernels a) DP kernel; b) DP^2 kernel . . 207

7.3 GPC results for the training data comprised of four classes of conditions. The colourmaps represent the probability for each of the classes. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points for different classes defined in the legend 208

7.4 Changes in probability predicted by the model for the training points presented in Figure 7.3; vertical lines signify the boundaries of classes while colours demonstrate the regions defined as stable (green), inlet recirculation (yellow), transient (red) and deep surge (grey); the class probability is given for classes: c_1 - stable, c_2 - inlet recirculation, c_3 - transient and c_4 - deep surge 209

7.5 Demonstration of the rejection zones placement for a GPC model trained with 20 points for each TOA and different rejection threshold level t_{reject}, as defined by Equation (7.12) a) 0.1; b) 0.2; c) 0.3; d) 0.4; grey regions on the plot demonstrate the rejection zones, while white lines show the boundaries between classes 212

7.6 GPC model trained with 20 data points for each TOA; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold threshold $t_{reject} = 0.3$ 214

7.7 GPC results for the training data comprised of four classes of conditions. The colourmaps represent the probability for each of the classes. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points 216

7.8 GPC model trained with data limited to two points per TOA; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold $t_{reject} = 0.3$ 217

7.9 GPC results for the training data comprised of four classes of conditions. The colourmaps represent the probability for each of the classes. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points for different classes defined in the legend 218

7.10 GPC model trained with data limited to peaks of instabilities; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold $t_{reject} = 0.3$ 219

7.11 GPC results for the model trained on data from the sensor before impeller. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points for different classes defined in the legend 221

7.12 GPC model trained with data from a single sensor before impeller; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold $t_{reject} = 0.3$ 222

7.13 Confusion matrix for investigation of classification accuracy for the models based on all data for two sensors a) *Model 1*; b) *Model 1R* incorporating the rejection zone 224

7.14 Confusion matrix for investigation of classification accuracy of the models based on limited data for two sensors a) *Model 2*; b) *Model 2R* incorporating the rejection zone 225

7.15 Confusion matrix for investigation of classification accuracy for the models based on all data from a single sensors a) *Model 3*; b) *Model 3R* incorporating the rejection zone 225

7.16 Comparison of the threshold-based model boundaries (solid black lines) with GPC-defined boundaries obtained with *Model 1R* (white solid lines); the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; the shaded region demonstrates a rejection zone for rejection threshold $t_{reject} = 0.3$ 226

7.17 Confusion matrix between the general conditions using features based on SSA decomposition 227

Contents

1 Introduction **1**

1.1 Centrifugal compressors generals 2

1.2 Compressor performance and off-design operation 6

1.3 Aerodynamic instabilities 9

1.3.1 Inlet recirculation 10

1.3.2 Rotating stall 14

1.3.3 Surge 19

1.4 Detection requirements, precursors and order of instabilities . . 25

1.4.1 Required pace of detection 26

1.4.2 Inlet recirculation as a precursor of surge 27

1.4.3 Rotating stall as a precursor of surge 28

1.4.4 Precursors to rotating stall 28

1.4.5 Order of instabilities in centrifugal compressors 31

1.5 Data-driven instabilities detection 32

1.5.1 Time-domain methods 33

1.5.2 Frequency-domain methods 34

1.5.3 Time-frequency methods 36

1.5.4 Correlation methods 37

1.5.5 Chaos-based methods and entropy 42

1.5.6 Data-driven decomposition methods 44

1.5.7 Selection of a signal type for monitoring 54

1.6 Methodologies for instabilities detection and classification 56

1.6.1 Features extraction and feature space representation . . . 56
1.6.2 Decision-making . . . 58
1.7 Challenges in data-driven instabilities detection . . . 65
1.8 Aim, objectives and contributions to knowledge . . . 68
1.9 Thesis outline . . . 70

2 Methodology for data-driven instabilities detection 72
2.1 Methodology overview . . . 73
2.2 Input signal . . . 76
2.3 Signal decomposition . . . 77
2.3.1 Empirical mode decomposition (EMD) . . . 77
2.3.2 Singular spectrum analysis (SSA) . . . 82
2.4 Decomposition components processing . . . 85
2.4.1 Obtaining features from components . . . 85
2.4.2 Selection of the components . . . 86
2.5 Feature space representation . . . 86
2.6 Decision making . . . 88
2.6.1 Threshold-based classification . . . 88
2.6.2 Probabilistic classification . . . 89

3 Case study definition 91
3.1 Object of research . . . 91
3.2 Pressure data acqusition . . . 94
3.3 Instabilities and data sets . . . 98
3.3.1 Observed instabilities and their physical character . . . 98
3.3.2 Available data sets . . . 100
3.4 Goals of the detection system . . . 102

4 EMD-based detection of instabilities 103

4.1 Compressor data decomposition 104

4.2 Influence of the decomposition parameters 107

4.2.1 Number of sifting iterations 107

4.2.2 Input signal length 110

4.2.3 Timing of the decomposition 113

4.3 General compressor stability identification 115

4.3.1 Locally unstable conditions 115

4.3.2 Globally unstable conditions 121

4.3.3 Feature space representation 131

4.4 Globally unstable conditions differentiation 133

4.4.1 Transient and deep surge features 133

4.4.2 Feature space representation 139

4.5 Summary, discussion and conclusions 140

5 SSA-based detection of instabilities 143

5.1 Compressor data decomposition 144

5.2 Influence of the decomposition parameters 149

5.2.1 Input signal length 149

5.2.2 Window length 150

5.2.3 Timing of the decomposition 154

5.3 General compressor stability identification 155

5.3.1 Locally unstable conditions 156

5.3.2 Globally unstable conditions 160

5.3.3 Feature space representation 164

5.4 Exact instabilities detection for increased accuracy 166

5.4.1 Transient and deep surge features 166

5.4.2 Feature space representation 171

5.5 Summary, discussion and conclusions 173

6 Comparison of EMD and SSA for instabilities detection 176

6.1 Comparison of instabilities-sensitive features 177

6.1.1 Local instability feature μ_L 177

6.1.2 Global instability feature μ_G 179

6.2 Threshold-based classification with μ_L and μ_G 184

6.2.1 Validation with dynamic data 185

6.2.2 Accuracy of classification 187

6.3 Practical consideration for an instabilities detection system . . . 191

6.3.1 Selection of the components 191

6.3.2 Advantages and disadvantages of the methods 193

6.3.3 Extension to accommodate other instabilities 195

6.3.4 Limiting the number of sensors for detection 196

6.3.5 Challenges and space for improvements 198

7 Probabilistic classification of instabilities 199

7.1 Gaussian process classification 200

7.2 Probabilistic classification for compressor instabilities detection 207

7.3 Rejection zones 211

7.4 Classification with limited training data availability 214

7.4.1 Limited training data representing all conditions 216

7.4.2 Limited training data representing peaks of instabilities . 218

7.5 Classification with single sensor data 220

7.6 Accuracy of classification 223

7.7 Summary, discussion and conclusions 228

8 Closing remarks 233

8.1 Conclusions 234

8.1.1 EMD-based instabilities detection 237

8.1.2 SSA-based instabilities detection 239

8.1.3 Comparison of EMD and SSA for instabilities detection . 241

8.1.4 Probabilistic classification enhancement 245
8.2 Limitations of the study 247
8.3 Further work 250

Nomenclature

List of symbols

α	Weight of kernel components in RQ kernel
$\boldsymbol{\mu}$	Feature vector
$\boldsymbol{\theta}$	Vector of hyperparameters
$\boldsymbol{c}$	Class labels vector
$\boldsymbol{D_{c_i}}$	Validation data set matrix
$\boldsymbol{D_{t,c_i}}$	Validation data set matrix
$\boldsymbol{D}$	Full data set matrix
$\boldsymbol{F}$	Features matrix
$\boldsymbol{K}$	Covariance matrix
$\boldsymbol{u_i}$	Eigenvector of an i^{th} component in SSA method
$\boldsymbol{v_i}$	Factor vector of an i^{th} component in SSA method
$\boldsymbol{X_i}$	i^{th} component of a trajectory matrix in SSA method
$\boldsymbol{X}$	Trajectory matrix in SSA method
$\dot{m}$	Mass flow
ϵ	Noise component in Gaussian process regression

λ	Eigenvalue of an i^{th} component in SSA method
$\mathcal{GP}$	Gaussian process
$\mathcal{N}$	Normal distribution
μ	Feature value
μ_*	Feature value for unknown operating conditions
ϕ	Mass flow coefficient
ψ	Pressure rise coefficient
σ	Logistic function
σ_0	Kernel hyperparameter
c	Class label
$cov(\cdot)$	Covariance of $(\cdot)$
$d(\cdot,\cdot)$	Eucliean distance
$f(\cdot)$	Function of $(\cdot)$
f_H	Helmholtz frequency
f_s	Sampling frequency
h_i	i^{th} proto-IMF in EMD method
i_i	i^{th} IMF in EMD method
L	Window length in SSA method
l	RBF lengthscale
L_M	Log marginal likelihood

$m(\cdot)$ Mean value of $(\cdot)$

N Number of features

N_s Input time series length

$p(\cdot)$ Probability value of $(\cdot)$

PR Pressure ratio

$Q(p)$ p-quantile

Q Number of IMFs created during decomposition

r_i Residual after an i^{th} step in EMD method / i^{th} element of RC

SD_i Stoppage criterion value at i^{th} step

SN Number of sifting iterations in EMD method

T_{μ_i} Threshold value for an i^{th} feature

$x(t)$ Input time series

x_i i^{th} input time series element

$y_k(t)$ Time series based on k^{th} component of the decomposition

List of abbreviations

AIRS Abrupt impeller rotating stall

CEEMD Complementary ensemble empirical mode decomposition

CFD Computational fluid dynamics

CWT Continuous wavelet transform

DMD Dynamic mode decomposition

DP Dot product

DRS Diffuser rotating stall

EEMD Ensemble empirical mode decomposition

EMD Empirical mode decomposition

FT Fourier transform

GP Gaussian process

GPC Gaussian process classification

GPR Gaussian process regression

HHT Hilbert-Huang transform

IMF Intrinsic mode function

ITD Intrinsic timescale decomposition

k-NN k nearest neighbours

LMD Local mean decomposition

PCA Principal components analysis

PIRS Progressive impeller rotating stall

RBF Radial basis function

RC Reconstructed component

RDF Rate of derivative fluctuation

RQ Rational quadratic

SSA Singular spectrum analysis

STFT Short-time Fourier transform

SVD Singular value decomposition

SVM Support vector machine

TOA Throttle opening area

VDRS Vaneless diffuser rotating stall

VMD Variational mode decomposition

WVD Wigner-ville distribution

1 Introduction

This chapter presents a general introduction to centrifugal compressors including the operation, monitoring and detection of instabilities. Firstly, the most common aerodynamic instabilities are described, focusing on their types and impact on the compressor. Subsequently, methods of instabilities detection are demonstrated and discussed, based on different approaches they might represent. Next, the methodologies that could be applied to instabilities detection are outlined and their most important aspects are highlighted. Based on the literature review, gaps in knowledge and shortcomings of the presented approaches are identified. The chapter ends with a roadmap of the thesis, giving an overview of each chapter.

1.1 Centrifugal compressors generals

Compressors are widespread machines of multiple applications and various designs. The basic role of a compressor is to increase the pressure of a fluid that is delivered to it, which might be used in different applications including the chemical industry, petroleum industry, turbo engines and many others [10]. Compressors can be divided into positive displacement, such as reciprocating, scroll, or screw machines, and dynamic, including centrifugal and axial compressors [11].

The process of increasing pressure in dynamic compressors has two main steps. First, the fluid is accelerated with a rotating impeller, which increases its kinetic energy. Secondly, the fluid is decelerated in a diffuser and its kinetic energy is changed into pressure [1]. Neither of these processes is loss-free and the level of losses depends on a number of factors. Axial and centrifugal compressors differ in geometry, principles of operation and flow patterns in the machine. Hence, despite performing a similar general task, research on centrifugal and axial compressors is often separated [12]. This especially applies to research on aerodynamic instabilities, where researchers investigate one [13, 14] or the other [15, 16] type of machine. The focus of this thesis are aerodynamic instabilities in centrifugal compressors, hence only this type of machines will be considered.

Centrifugal compressors take advantage of centrifugal effects for flow acceleration, which is reflected in their design [7]. A schematic view of a centrifugal compressor is presented in Figure 1.1. The flow enters through the inducer - an axial part of the impeller passage. Some designs may have a partial inducer (semi-inducer impeller) or no inducer at all (radial impeller) [17]. In some compressor designs, the flow might be directed through guide vanes present before an impeller, in between sections 0 and 1. Guide vanes can introduce additional curl to the flow, affecting the aerodynamics of the impeller and its performance [18]. The impeller extends between sections 1 and 2. Its shape and design may differ

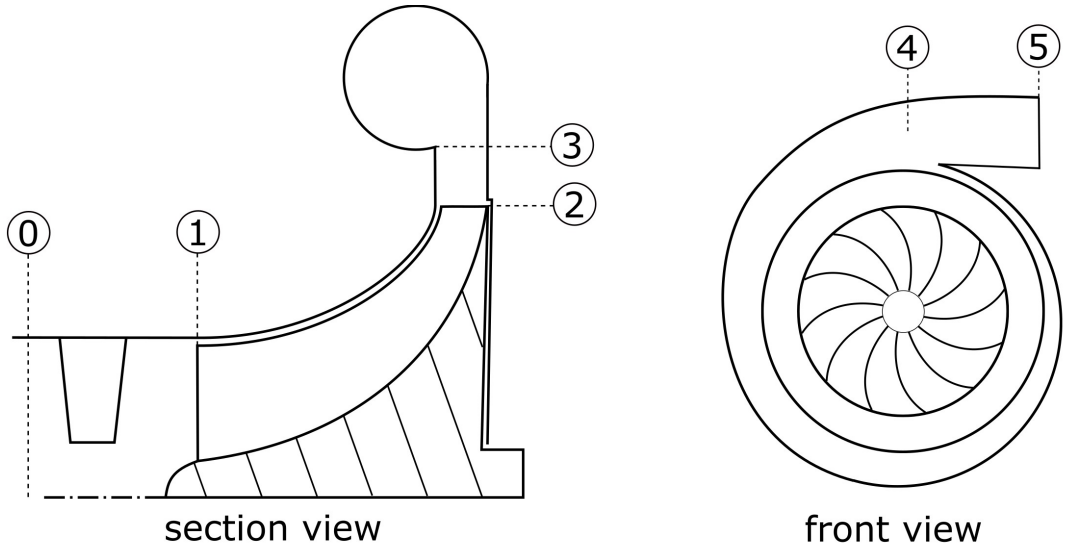


Figure 1.1: Schematic representation of a centrifugal compressor with marked sections; 0: compressor system inlet; 1: impeller leading edge; 2: impeller trailing edge; 3: end of the diffuser; 4: end of the volute; 5: compressor system outlet

importantly, depending on the application. Past the impeller, the flow enters a diffuser, located between sections 2 and 3. In the diffuser, the kinetic energy of the flow is transformed into a pressure increase. Diffusers can be vaneless, where there are no guides to the flow or vaned, where aerodynamically-shaped vanes are present [7]. After the diffuser, the flow enters a volute, where its velocity vector changes from radial to tangential in order to transfer the medium into the pipeline. The shape of a volute can differ between compressors, and it was shown to have an impact on the operation of the machine [19]. Past the diffuser, an outflow cone might be present, located between the end of the volute (section 4) and outflow (section 5).

The design of a centrifugal compressor may differ significantly, depending on its application [7]. Differences in size and geometry may be very important, resulting in very different flow patterns and behaviour close to off-design conditions [10]. With changes in size, compressors operate at different rotational speeds. Different families of compressors, as defined by Casey et al. [1] are

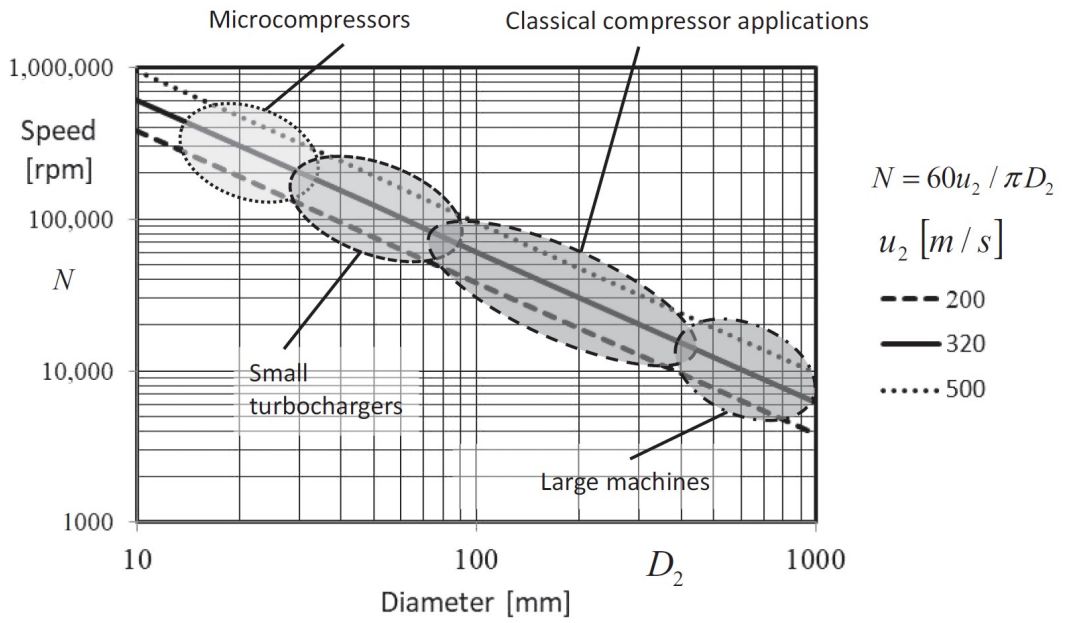


Figure 1.2: Relation of size and rotational speed for different types of centrifugal compressors [1]

represented in Figure 1.2. It can be seen that both the rotational speed and the impeller diameter can vary greatly between machines.

The design of a compressor is influenced by the function it has in the system, expected performance, the medium it works with, space constraints and many other aspects [17]. There exist variations in each of the compressor elements, which influence the performance and stability of the machine [11, 7]. Firstly, there are several variations to the impeller design. Depending on the application, the impeller can be shrouded or unshrouded (Figure 1.3a), which importantly changes the impeller aerodynamics and secondary flows present in the machine [7]. The impeller blades can be shaped in a 2-dimensional way, having a constant shape along the height of the flow channel or in a 3-dimensional way, varying across the height (Figure 1.3b). The form of the blade may also differ, depending on the expected flow coefficient for the machine (Figure 1.4). Additionally, splitter blades may be used for some high-performance impellers [7]. The variations in compressor design can also come from the type of the

diffuser, being vaned or vaneless. Some compressor designs may additionally include inlet guide vanes to alter the flow swirl at the inlet [17]. There is also a number of other shape parameters, related to the diffuser or volute, that affect machine performance and operating range [7].

The variability in design and the number of possible combinations make generalization of centrifugal compressor behaviour challenging [11]. Due to the multitude of designs combined with complex flow physics inside the machine, the behaviour of centrifugal compressors has been far less understood than that of axial compressors, especially when considering off-design operation [2].

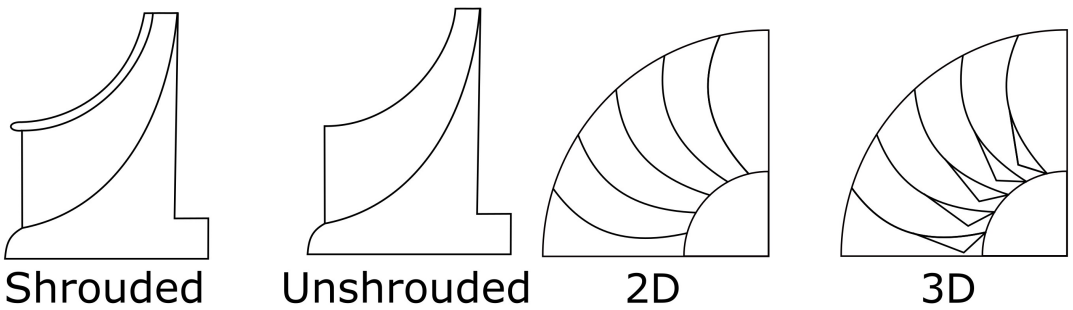


Figure 1.3: Variations of impeller designs a) shrouded and unshrouded impellers; b) impeller with 2D and 3D blades

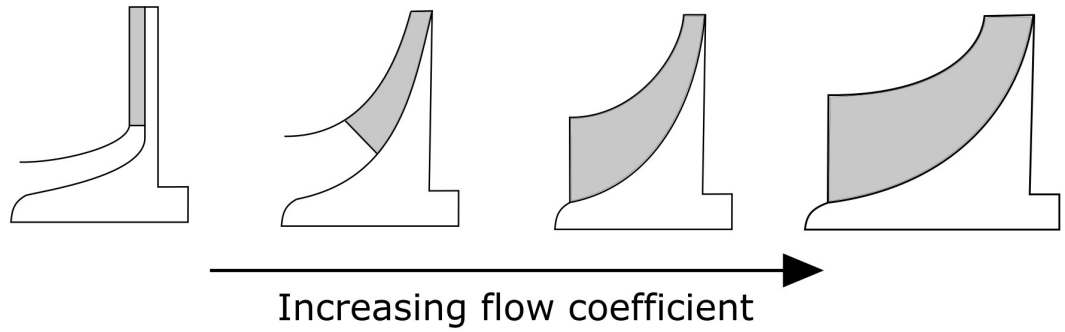


Figure 1.4: Variations of impeller designs dependent on the flow coefficient

1.2 Compressor performance and off-design operation

Compressor performance can be quantified on a global level using a compressor map that relates the mass flow ($\dot{m}$) of the machine with pressure ratio (PR) [10]. An exemplary compressor map is shown in Figure 1.5 It is possible to define a functional relationship in the form $PR = f(\dot{m})$ based on the fit of a function to empirical data, often in the form of third-order polynomial [7]. A different function can be created for each rotational speed. The compressor operates connected to a network of certain characteristics, being a function of a mass flow through that network. The operating point of a compression system locates at the cross-section of the compressor characteristics and network characteristics. If some flow or resistance perturbations occur, which is common in normal compressor operation, the operating point can oscillate around some location on the compressor map [7].

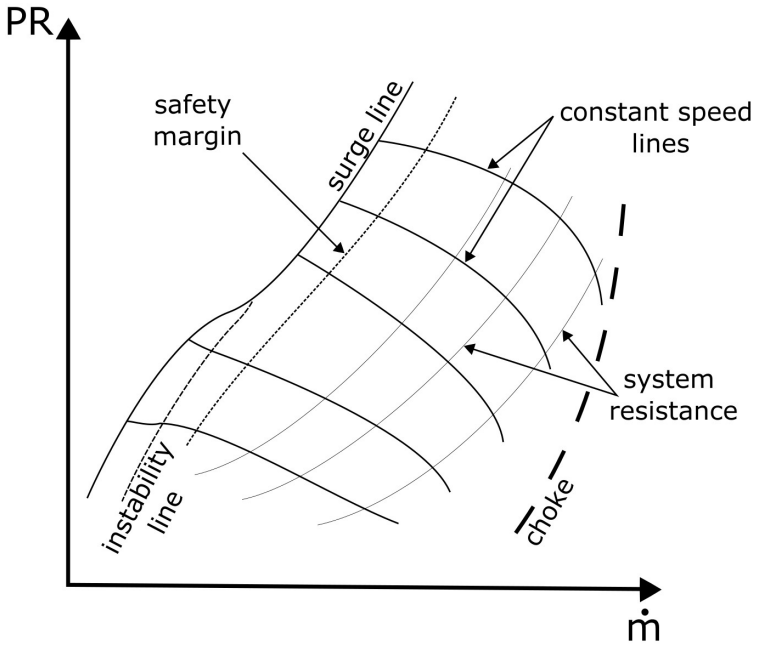


Figure 1.5: Exemplary compressor map, based on [2]

To compare different machines and fluids, the mass flow and pressure rise

can be represented in a non-dimensional form using the mass flow coefficient ϕ (Equation (1.1)) and the pressure rise coefficient ψ (Equation (1.2)) [10].

$$\phi = \frac{4\dot{m}}{\rho u_2 \pi d_2^2} \tag{1.1}$$

$$\psi = \frac{PR}{1/2\rho u_2^2} \tag{1.2}$$

Where:

$\dot{m}$ is a mass flow through the system,

ρ is a fluid density,

u_2 is a blade tip velocity,

d_2 is an impeller diameter at the outlet,

PR is a pressure ratio.

The operating range of a compressor can be defined using a compressor map [10]. This range is limited by the choke at high mass flow rates and surge or instability line at low mass flow rates [2], as shown in Figure 1.5.

The choke appears when the flow speed inside the compressor exceeds the speed of sound, so the Mach number is greater than 1. This leads to high losses [7] and is often far from the optimum design point, as the pressure rise is low. The choke line calculation is a relatively easy and well-defined task. The location of the line depends on the area of the narrowest passage location, mass flow and fluid parameters. Operation close to choke conditions is rare in normal practice, as it is typically far from the design point [2].

On the other end of the compressor map, the operating range is limited by the appearance of surge or other flow instabilities, such as inlet recirculation or rotating stall. Instabilities are a number of different aerodynamic structures, with different influence on the machine. They are extensively described and discussed in Section 1.3. All instabilities are detrimental to compressor efficiency and can lead to the destruction of a machine due to fatigue or overload of

the compressor components [2]. High pressure ratio and high efficiency of the machine are often attained close to the unstable region, therefore, operation close to the surge line is more common.

A compressor should operate in the stable region for the purpose of safety and efficiency. However, during compressor operation, the operating point can shift with a change in rotational speed, variation of the network resistance or due to the presence of disturbances [7]. Such a shift may happen abruptly, pushing the compressor from stable operation to unstable region. It is also possible that because of wear or deterioration of performance, the compressor characteristics changes. It can make the set operating point get closer and closer to the unstable region, causing a threat to machine operation [20].

Defining the exact location of the unstable region is a difficult task. It strongly depends on the flow structure inside the machine, which is to a certain extent unique for every compressor design, thus making it difficult to accurately predict in an analytical way [7]. The surge line in industrial practice is obtained experimentally by a compressor manufacturer [21], but can also be estimated theoretically, for example using Greitzer model [22]. Theoretical modelling has its limitations, as a number of parameters have to be estimated, including compressor and network characteristics [7]. The experimental procedure, although more precise, is still burdened with uncertainty. The main sources are manufacturing imperfections, as well as compressor transport and installation, affecting machine clearances. Additional errors can also be introduced with compressor curve fitting and numerical corrections of the curve due to changes in the parameters of the working medium [1, 21, 23].

Owing to an underlying uncertainty of the exact surge line location, a certain buffer from the surge line is introduced into the compressor control system to avoid surge onset [2]. This buffer may be called a safety margin [1]. The safety margin can be set at around 10 to 15% of the compressor operating range [2, 24, 25]. Such a definition ensures the safety of the machine, however, it

importantly limits the compressor operating range and may prevent the machine from operating at the peak of its performance [26]. Both performance and operating range are important for centrifugal compressors and improvements in either of these aspects can bring considerable benefits [1].

Centrifugal compressors can experience other aerodynamic instabilities than surge. The two major ones described in the literature are inlet recirculation and rotating stall [11, 1, 2]. Their inception occurs close to the compressor pressure curve peak, but the exact location depends on a number of factors [12, 7, 2, 27]. Inlet recirculation and rotating stall have a detrimental effect on compressor efficiency and their presence can cause damage to the machine [12, 2]. It makes the detection of those instabilities an important aspect of an instabilities detection system.

The compressor control system based on a compressor map can help to roughly define the regions of possible instabilities presence but does not offer the possibility of their detection in terms of defining their presence in the particular, monitored system. To ensure safety and long-term operability of the compressor without compromising the operating range or efficiency, a control method based on an online detection of instabilities can be proposed. Such a method should provide instantaneous information about the machine stability in a continuous manner. It should also be capable of detecting multiple instabilities, to be considered for use in a comprehensive compressor controller.

Accurate and comprehensive detection of aerodynamic instabilities is a challenging task. To provide a general approach, capable of quick detection and distinguishing of instabilities, the character of instabilities and their signature have to be understood.

1.3 Aerodynamic instabilities

Aerodynamic instabilities are different flow structures that can appear in the compression system when the machine works in proximity or past its peak com-

pression level, where the mass flow of a compressor is low [7]. The term instabilities in the context of centrifugal compressors encompasses a number of flow structures, generally detrimental to the compressor. The three major ones appearing in the literature are inlet recirculation, rotating stall, and surge. They can be differentiated by the location they appear at, their characteristic signature and their influence on the compressor.

1.3.1 Inlet recirculation

Inlet recirculation is an instability with the lowest overall effect on compressor performance and structural integrity, compared to rotating stall or surge [28]. In the literature, it is also referred to as rotating instability [29], inducer stall [30] or recirculating bubble [3]. The name whoosh noise is often used in reference to the audible effects that can be produced by inlet recirculation [31], frequently observed for turbocharger compressors. The naming of this phenomenon has been a subject of discussion in the scientific community [32], however in this thesis, it will be consequently referred to as inlet recirculation after the most comprehensive study of this phenomenon in centrifugal compressors [28].

Physics of inlet recirculation

In centrifugal compressors, inlet recirculation has a form of an axisymmetric swirling zone appearing at the leading edge of the impeller close to the shroud of the compressor [33]. It can extend upstream into the inducer and downstream into the impeller [3], in some cases reaching even the diffuser [34]. The extent of inlet recirculation may vary depending on the operating conditions as well as the parameters of the machine. The recirculation rotates in the plane of the flow and has a radial velocity component concurrent with the impeller rotation direction, which arises from the interaction with the impeller [1]. A schematic view of inlet recirculation is presented in Figure 1.6.

Inlet recirculation was first documented for centrifugal pumps [35] and subsequently for axial fans and blowers [36]. It was not extensively researched for centrifugal compressors, with only minor mentions of such phenomenon from several researchers [30, 37, 38]. Over the years the interest in inlet recirculation was increasing. It was partly related to advances in turbocharger compressors, as the effects of inlet recirculation are particularly noticeable for this type of machines [33]. Over time, the effects of inlet recirculation were included in the compressors design procedures [33, 39]. Inlet recirculation can be present in a centrifugal compressor at a given operating point, providing little symptoms of its existence when considering the global performance of the machine [2]. However, it can cause a broadband excitation of the compressor elements, leading to fatigue damage in a long term [7].

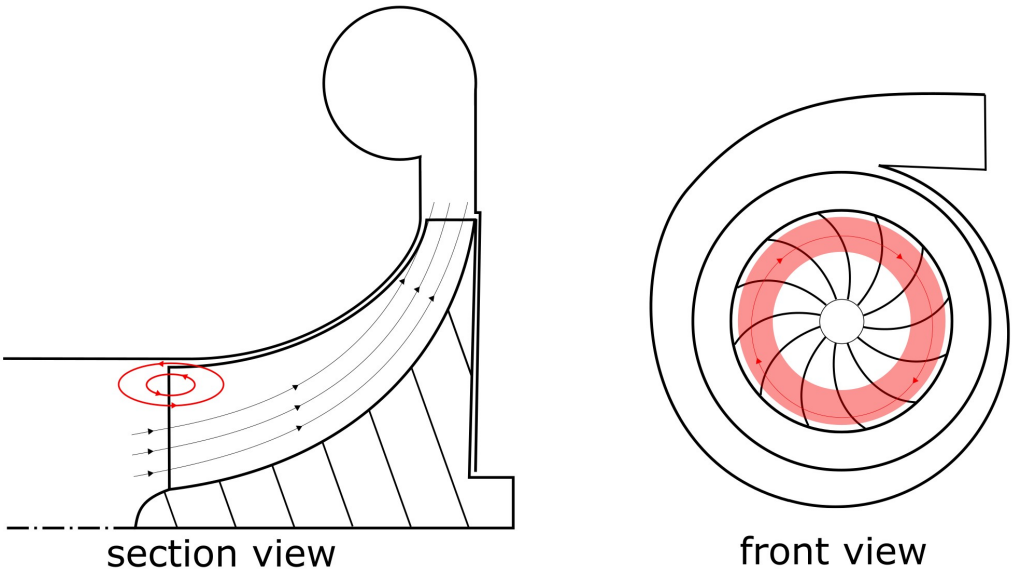


Figure 1.6: Typical inlet recirculation zone in the inducer and impeller; based on [3]

The mechanism of inlet recirculation onset was investigated numerically, which provided great insight into the flow structure inside the machine [15, 16, 40]. It was suggested that inlet recirculation is triggered in the impeller, near the leading edge, when the tip vortex from a previous blade hits the pressure

side of another blade [15, 16]. This happens as with decreasing mass flow, the tip vortex is no longer pushed downstream along the rotation axis of the impeller, but rather starts to move perpendicular to that axis [40]. Assuming that the tip vortex is the only factor responsible for initiating recirculation, this instability would be eliminated in shrouded impellers since the tip gap, which is necessary for tip vortex creation, does not exist in such impellers. The observations of inlet recirculation for shrouded rotors were reported by McKee et al. [41], where it was shown that such a structure was only present for inducer or semi-inducer impellers, not for purely radial impellers.

Ribi [2] provided an alternative theory on the onset of inlet recirculation, arguing that recirculation can be caused by a leading edge separation on the suction side. It was demonstrated that inlet recirculation is initiated within the impeller passage, triggered by a separation on the suction side and enhanced by the tip leakage flow [33, 34]. Following such an explanation, the recirculation is expected to appear for the shrouded compressor impeller as well, but to a lower extent as the tip leakage enhancement would not be present.

Inlet recirculation can extend through the whole impeller, up to the trailing edge of the compressor blade [33, 34], causing a significant flow blockage. Having reached the diffuser, recirculation might alter the diffuser flow patterns and affect its stability. The extent and influence of those effects on compressor stability are specific to compressor geometry [39].

Inlet recirculation effects

The presence of recirculation alters the flow path in the impeller, as it provides a blockage to the flow on the shroud side, as shown in Figure 1.6. Due to its rotation in agreement with the impeller rotation direction, it introduces a pre-swirl on the incoming flow and pre-heats it, as higher energy fluid is reversed to the impeller inlet and mixed with the incoming flow [42]. Decreasing the available throat area results in increasing the flow velocity in the axial direction,

while the presence of the pre-swirl increases the meridional velocity component magnitude [28].

The presence of a flow blockage near the shroud may have a stabilizing effect on the global flow as due to the lower area, the axial velocity increases, influencing the incidence angle and offsetting the appearance of more harmful instabilities [1]. The numerical study of Lin et al. [3] suggested that inlet recirculation effect on compressor efficiency can be both positive and negative, depending on the stage of recirculation development dictated by the mass flow rate. In the first stage, the presence of recirculation decreases the active flow area, resulting in an increased flow velocity for the same flow rate. The recirculating zone also pre-swirls the flow in the direction of compressor rotation. Both of these mechanisms decrease the incidence angle, resulting in lower incidence loss. With a further decrease in the flow rate, those effects are countered by important pre-heating of the flow, leading to a drop in efficiency due to the thermodynamics of the compression process [3]. With an even greater mass flow decrease, a compressor can experience other instabilities, such as a rotating stall or surge.

Inlet recirculation was shown to have a broadband signature [43, 9], which can lead to a broadband excitation of the compressor elements [44]. It might, in the long term, lead to fatigue damage of the blades or shorten the life span of other compressor components [2, 44]. Due to the presence of inlet recirculation in its developed form, increased energy losses can be observed. In some cases, inlet recirculation can account for up to 35% of the total compressor losses [28].

Inlet recirculation detection

Detection of inlet recirculation can be based on different principles. The broadband excitation caused by swirling flow can be sensed using a pressure sensor positioned in the vicinity of the impeller leading edge [45]. It is also possible to measure changes in the tangential velocity component before the impeller,

which increases when inlet recirculation is present [41]. The reversal of a high-energy flow from the compressor could be captured through temperature measurements before the impeller, it was shown that inlet recirculation results in a higher standard deviation of the temperature signal [46]. In small centrifugal compressors used in turbochargers, the presence of recirculation can be responsible for the so-called whoosh noise [31, 46], therefore the acoustic signature of inlet recirculation could also be used for the detection.

With the limited impact of inlet recirculation on the compressor, its rapid detection is not of paramount importance. However, accounting for its effect in a compressor system might be helpful for fatigue damage safety. Energy savings could also be made if inlet recirculation can be avoided. Inlet recirculation can also serve as a surge precursor in some cases [43, 8, 47], therefore, its rapid detection during the transition from stable to unstable conditions can be leveraged for surge onset prediction.

1.3.2 Rotating stall

Rotating stall is a local instability that has a more severe impact on the compressor than inlet recirculation, both in terms of safety and machine performance [7]. Stall was researched to a higher extent for axial compressors than for centrifugal compressors [48]. It was due to its much more severe impact on axial machines, where stall results in an instantaneous loss of pressure ratio and creation of high loads on the slender blades of an axial compressor [17]. Rotating stall in centrifugal compressors is also dangerous, thus its detection was of concern for a number of researchers [2, 49].

Physics of rotating stall

The term stall describes a detachment of the flow, often caused by an adverse pressure gradient [2], which can be very well visualized for airfoils (Figure 1.7). When exceeding a specific angle of attack, the flow detachment takes place and

results in an important drop in airfoil lift [12]. In compressors, the term stall can refer to the separation on the blades of the impeller or vaned diffuser or a separation in the vaneless diffuser. In each case, stall is caused by an adverse pressure gradient in the boundary layer [12]. The presence of stall often results in an important change in the flow field inside the compressor and drop in compressor performance [7].

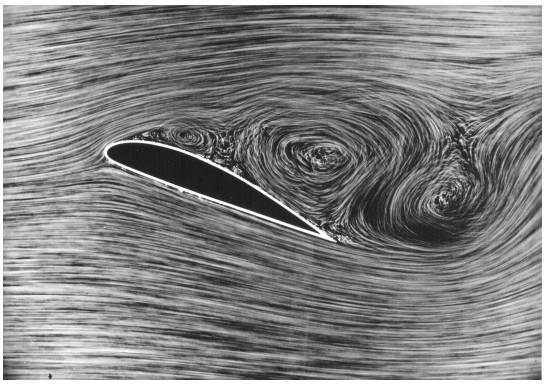


Figure 1.7: Stalled airfoil [4]

Rotating stall in centrifugal compressors can have different forms. In general, it is manifested by the presence of a number of low-energy cells that rotate around the annulus of the compressor [7]. It is also possible that a single cell is present in the compressor. The number of cells can vary across machines [2] or even within the same machine for different stages of development [50]. Within stall cells, the flow is stagnant or reversed, constituting a blockage to the normal flow inside the machine [7]. These zones locally change the loading of the compressor blades and increase the level of vibration of the machine due to the force imbalance [44].

The rotation of stall is caused by the interaction of cells with the core flow [2]. When a blade passage is stalled, the flow cannot go normally through a passage and is redirected to adjacent channels. For one of those channels, due to an additional component of a velocity vector, the incidence angle is increased, while for the other it gets decreased [5]. In the channel with increased incidence,

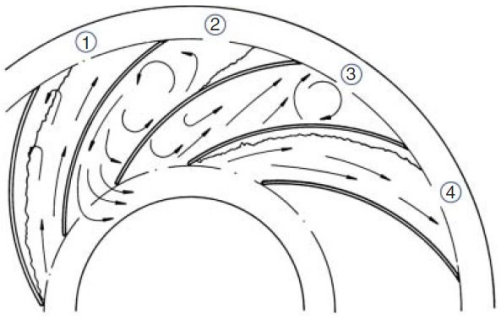


Figure 1.8: Impeller rotating stall evolution; numbers represent the temporal evolution of the flow in the same space between the blades [5]

the separation on the suction side is more likely to form, causing this passage to stall [2]. In the previously stalled channel, when no flow through the channel takes place, the pressure distribution changes and the normal flow through it is re-established. This process is visualised in Figure 1.8.

There exist different types of rotating stall, depending on the location and the onset mechanism [7]. The onset of stall is highly dependent on the compressor type, geometry and operating conditions [2]. One of the most commonly applied classifications [7] differentiates subsequent types of stall:

- Progressive impeller rotating stall (PIRS)
- Abrupt impeller rotating stall (AIRS)
- Vaned diffuser rotating stall (DRS)
- Vaneless diffuser rotating stall (VDRS)

PIRS occurs when the angle of attack of an impeller blade becomes too high, inducing local flow separations inside the impeller [7]. These separations can develop into stall cells, creating a blockage in certain passages. A different number of cells and their rotational speeds were observed in the literature, with cell numbers ranging from 1 to 5 and rotational speed between 40% and 100% of impeller rotational frequency [5, 51]. The name progressive comes from the steadily increasing amplitude of phenomena with increasing throttling [7].

AIRS arises from the interaction of the impeller with a diffuser [52]. The stall cells emerge at the outlet of the impeller, extending into diffuser. The number of cells is within the same range as for AIRS, but they rotate more slowly, between 20 to 40% of the rotational speed [2]. The appearance of AIRS is dependent on the location of the operating point on compressor curve as well as the impeller outlet flow parameters [50]. The compressor geometry has a significant impact on both parameters. Specifically, the lean angle of the blades greatly influences the sensitivity, particularly for backward lean blades which may result in VDRS instead of AIRS. However, if certain conditions introduced by flow perturbations are present, then AIRS can occur [7].

DRS has a very similar mechanism to PIRS, but takes place in the diffuser. Due to too high incidence, local separations occur which tend to grow into stall cells [53]. Owing to the interactions of the flow with the impeller, the inflow to the diffuser is more complex than the inflow to the impeller. It results in a more complicated character of diffuser stall [7]. The velocity of the cells can range from 5% up to 85% of the impeller rotational speed and their number can vary between 1 and 4 [2]. The most important role in stall onset is played by spanwise non-uniformity of the flow, which directly affects the incidence on diffuser vanes [7].

VDRS is the instability of a similar character as previous stalls, but appearing in the vaneless diffuser [52]. The stall onset is due to the combination of non-damped spanwise non-uniformity of the flow and the boundary layer separation [7]. During VDRS, the net mass flow remains the same as for stable operating conditions. The stall cells restrict the outflow from the diffuser, consequently, the flow in between them is accelerated to account for the decreased flow in stalled regions [50]. The number of cells is on average 1 to 3, and they rotate slowly, within the range between 5% to 30% of the impeller rotational speed [2].

The effects of rotating stall

The presence of rotating stall is detrimental to a compressor in a number of ways. Primarily, it results in a decrease in compressor performance due to flow blockage and increased losses introduced by the stall cells [7]. The presence of rotating stall also alters the pressure distribution on the elements of the compressor, mainly on the impeller blades [44]. This leads to cyclic loading and causes an increased level of vibrations of impeller blades and shaft. In the long term, it may result in fatigue damage of the impeller [54]. The significance of stall-related forces grows with increasing density of the process fluid [7]. Stall can also be responsible for triggering surge [7].

Rotating stall detection

Stall detection does not have to be instantaneous, as most centrifugal compressors can often tolerate the presence of stall for some period [2]. Rotating stall detection is important for compressor safety and efficiency. A method allowing for its quick detection can be useful where the presence of stall can be considered a surge precursor.

The detection of stall is often based on pressure measurements inside the machine [52, 55, 56]. The placement of the sensors may differ, depending on the expected stall type and machine design [2]. Commonly, at least two sensors placed at the same radial position are used to measure the number of cells as well as their velocity [57]. It is also possible to use a single sensor to capture stall, however, then it is not possible to precisely measure the number of cells and their velocity [58].

Pressure measurements can also be used to detect modal waves, which are responsible for progressive stall inception [59]. Modal waves are long, transverse pressure waves that travel around the annulus of the compressor [7]. They are often of small amplitude, compared to the pressure variation inside the compressor, hence their detection is not straightforward [60].

Rotating stall can also be detected through measuring the vibrations of compressor elements, induced by this instability [61, 62]. However, in such configurations, the detection of stall may be hindered by different sources of vibrations present in the system [63]. Detecting stall is not straightforward and often requires important signal processing efforts to extract the signature of this instability [64, 65]. It applies especially to an early stage of stall, where its signature in pressure or vibration signal is not clear [66].

1.3.3 Surge

Surge, contrary to inlet recirculation and rotating stall, is a global phenomenon that affects the entire compression system [12]. It is a highly dangerous instability that can lead to quick damage or even destruction of the system components, rendering the compressor inoperable [2]. Surge can have a different impact depending on the medium used by the compressor and compressor design, but it should generally be avoided due to the high load exerted on compressor elements [7]. When surge onset is detected, the common practice is to depressurise the plenum by using a bypass valve or, in extreme cases, to stop the compressor [7]. It was shown that surge suppression may be possible with different active methods, such as controlling the compressor speed [67] or changing the plenum volume [68, 69]. However, these approaches have not been widely implemented in industrial practice, partly due to the size requirements of the suppression systems and the increased complexity they bring to the control system [7].

Physics of surge

Surge is manifested by periodic pressurizing and depressurizing of a plenum, which is the volume between the compressor and a throttle or network resistance [12]. This process results in important pressure pulsations that affect the entire compressing system. The character, intensity and effect of surge on the compressor differ, depending on the compressor design, operating parameters

and the network it is connected to [12]. Surge can be divided into different types, depending on the flow conditions. The division, according to De Jager [70] is presented below.

- **Mild surge -** the flow demonstrates small pressure fluctuations of a period close to the Helmholtz frequency f_H. The flow reversal does not occur, which means that the total net flow is positive. It doesn't need to imply that there are no local reversed flow zones.

- **Classic surge -** the pressure oscillations are larger and at a lower frequency than during mild surge. The higher frequency oscillations can also be present in the form of harmonics of the base frequency as the signal becomes non-linear. Still, no full flow reversal is present in the system.

- **Modified surge -** it is a classic surge condition with a superimposed rotating stall, resulting in fluctuations in both axial and circumferential directions.

- **Deep surge -** this is the ultimate situation, where the full flow reversal occurs. It is characterized by the strongest level of pressure fluctuations.

Ribi [2] provides a slightly different taxonomy, where the term mild surge covers both mild and classic surge of De Jager and modified surge from De Jager is referred to as classic surge. In literature practice often only mild and deep surge are distinguished [2].

Several attempts were made to predict the onset of surge and model compressor behaviour under these conditions. A major contribution was the work of Greitzer [22], who proposed a mathematical model of surge. This model, first derived for axial compressors, was extended to centrifugal compressors and proven experimentally [71, 72]. It was shown that a system can be modelled as a dynamical system composed of a number of simple components. The schematic description of a compressor system is presented in Figure 1.9.

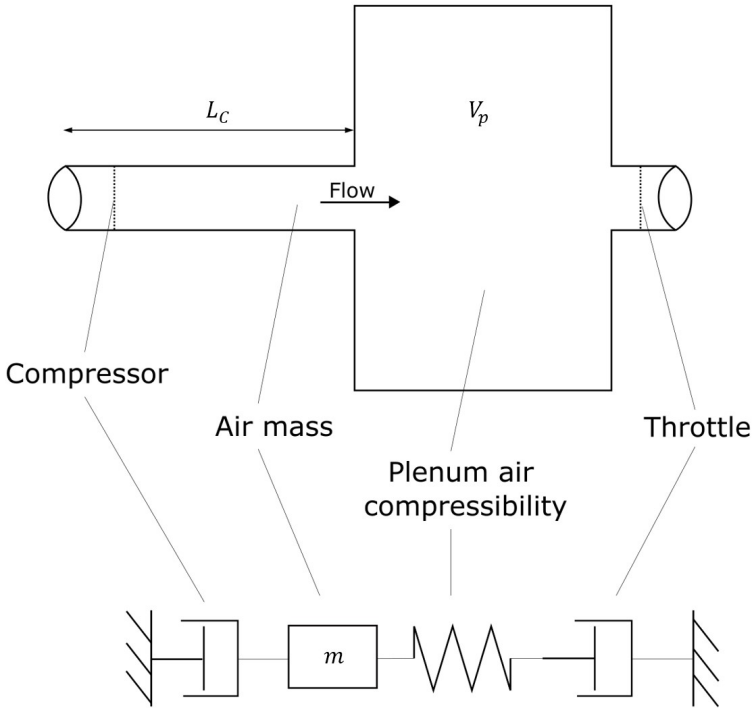


Figure 1.9: Schematic depiction of Greitzer compressor model as an analogue to a mechanical system; adapted from [6]

Greitzer model can be represented with two main equations (Equation (1.3) and (1.4)) connecting the mass flow and pressure rise with the compressor and throttle characteristics [6].

$$\frac{d\phi}{dt} = B\left(\psi_c - \psi\right) \tag{1.3}$$

$$\frac{d\psi}{dt} = \frac{1}{B}\left(\phi - \psi_T^{-1}\right) \tag{1.4}$$

Where:

ϕ and ψ are non-dimensional mass flow and pressure respectively,

ψ_c and ψ_T are the compressor and throttle characteristics.

An important element of the model is B parameter (Equation (1.5)). Its value can be understood as a measure of pressure to inertia force ratio [2].

The bigger the B, the stronger the surge oscillations, thus the effect on the compression system [22]. A bigger B parameter makes a system more prone to the appearance of unstable structures, while small B has a stabilising effect on the compressor [71]. In some studies, it was shown that the value of B parameter may define the type of instability present in the system, being either rotating stall or surge [12].

$$B = \frac{u_2}{2a}\sqrt{\frac{V_p}{A_c L_c}} \tag{1.5}$$

Where:

a is a speed of sound,

V_p is a volume of the plenum,

A_c is the throat area of the compressor,

L_c is a compressor and duct length.

Greitzer model can be used for prediction of the onset of instability using the shape of the compressor curve and the resistance curve as input [22]. There are two types of system stability: static and dynamic [6]. For static stability, the compressor performance curve must have a higher slope than the resistance curve, as only then the perturbations of pressure caused by compressor operation are damped by the system [6]. The condition for static stability can be written as in Equation (1.6).

$$\frac{d\psi_c}{d\phi_c} \leq \frac{d\psi_T}{d\phi_T} \tag{1.6}$$

The other type of stability is dynamic stability, which can be derived with the same approach as static stability. However, in this case, the dynamic behaviour of the compressor is considered, hence the difference in the inertia of the compressor and resistance must be factored in. Consequently, when a disturbance in compressor operation occurs, the compressor follows its characteristics

while, due to inertia, the resistance does not. The behaviour of resistance can be modelled using B parameter. The condition for dynamic stability is shown in Equation (1.7). For a stable compressor operation, both static and dynamic stability conditions have to be met [6].

$$\frac{d\psi_c}{d\phi_c} \leq \frac{1}{B^2}\left(\frac{d\psi_T}{d\phi_T}\right)^{-1} \tag{1.7}$$

The stability of a compressor in Greitzer model is defined based on the shape of the compressor performance curve and throttle characteristics, therefore the predicted location of instability is dependent on the shape of those curves. Various modelling approaches were proposed for obtaining compressor performance curves. These models are often 1-dimensional and rely on a number of approximations. Some models, such as a meanline model divide the compressor into parts and focus on its separate components to define the global performance [17]. The models often require the selection of several parameters, which are based on correlations and previous experiments [17]. The primary benefit of utilising simplified models lies in their ability to rapidly generate predictions for initial compressor design. However, they are not accurate enough to be suitable for compressor control when the aim is to maximally approach the surge line without the risk of crossing it [73].

The compressor performance curve can also be obtained with computational fluid dynamics (CFD). CFD can provide accurate predictions of compressor behaviour in off-design conditions [16, 9]. It was demonstrated for a turbocharger compressor that the results provided by CFD are more accurate than those produced with simpler models [74]. However, compared to the experimental results, the CFD also produced some discrepancies, affecting the presence of a predicted surge line [74]. Such differences can be caused by manufacturing imperfections or misalignment introduced during machine transport and installation [21, 23]. It is also possible that the compressor experiences wear over time, which would

furtherly alter the shape of the compressor map and, consequently, the location of the predicted surge line [1, 75].

The effects of surge

Surge can have a highly destructive effect on the compressor [12]. The presence of flow oscillations induced by surge leads to high axial loading of the bearings [57], which may cause their failure and consequently quick destruction of the machine. Surge also significantly alters the performance of the compressor, including its pressure ratio and the mass flow [2]. The flow oscillations cause strong vibrations of the compressor and connected elements [44]. What is more, the surge-induced vibrations can transfer onto the compressor network, leading to failures of the piping or other elements of the system [7].

Surge detection

Surge is characterized by strong pressure fluctuations, present everywhere in the compression system. Depending on the type of surge, a different signature can be observed [70]. The frequency of surge can be well approximated by Helmholtz frequency (Equation (1.8)), dependent on the physical dimensions of a compressor and its network. Knowing the expected frequency, it is possible to monitor the signal at a specific frequency range to capture the rise in amplitude related to surge [9, 63].

$$f_H = \frac{a}{2\pi}\sqrt{\frac{A_c}{L_c V_p}} \tag{1.8}$$

Helmholtz frequency for a number of compression systems reported in the literature ranges from below 1 Hz to over 40 Hz [76, 77], which is generally lower than rotating stall frequency [2]. However, an overlap between surge and stall frequencies might be present in some cases, as indicated visually in Figure 1.10. Surge-induced oscillations can be distinguished from rotating stall by comparing the phase at more than one sensor. Surge oscillations are in phase

for all the circumferential locations over the impeller and strongly present in locations remote from the impeller due to a global character of surge, while stall oscillations are more local and not in phase over the impeller annulus [7]. However, using a measurement system with a limited number of sensors, distinguishing those two instabilities might be challenging.

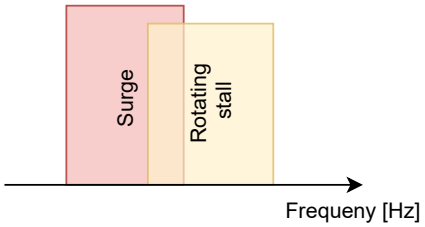


Figure 1.10: Visualisation of frequency overlap in detection of surge and rotating stall

A fully developed surge is easily detected because it produces an abrupt change in a number of parameters, including high-amplitude pressure oscillations and vibrations of the structure [63]. The change in the pressure signal can be for example observed in Figure 3.6 in Chapter 3. The challenge in surge detection lies in the prediction of its onset and ensuring system reaction prior to surge development.

1.4 Detection requirements, precursors and order of instabilities

The strategy for instabilities detection and ensuring the safety of the compressor can be either to quickly detect the symptoms of instability [62, 78] or rely on some instability precursors [2, 66]. The precursor should appear prior to the targeted instability, but as close to its onset as possible. Too early the appearance of a precursor can unnecessarily restrict the operating range of a machine. With a far less important effect of inlet recirculation and rotating stall on the compressor integrity compared to surge, they were investigated as

surge precursors for some machines [2, 47, 57]. There also exist other features that were shown to precede rotating stall in certain compressors [18, 66]. Such features can be used as additional steps in the control system, allowing a more precise indication of the compressor stability.

1.4.1 Required pace of detection

The impact of instabilities on the compressor varies, affecting the necessary pace of detection. Inlet recirculation is the mildest of instabilities. It can remain present in centrifugal compressors, in its initial stage, even without a significant effect on the performance [28, 3]. Therefore, the pace of detection for inlet recirculation is not paramount due to the low risk of compressor damage. However, considering inlet recirculation a precursor to some other instability, its quick detection may be very useful.

The effect of rotating stall is more significant than that of inlet recirculation. Its influence depends on the compressor parameters and is higher for machines operating at a high pressure ratio or using a medium of higher density [7]. The stall in centrifugal compressors can often be tolerated by the compressor in short periods [58]. Within a longer time frame, stall can lead to fatigue damage of the machine due to excitation of the blades [44, 54]. The damage may occur abruptly, with fatal consequences [79]. Hence, the detection of rotating stall in centrifugal compressors should be fast, but it does not have to be instantaneous. Again, considering stall as a precursor to surge, its quick detection becomes important.

Surge has the most grievous effects on the compressor. In extreme cases, it can lead to the destruction of a machine within a very short period of time [18]. For the majority of the compression systems, surge must be avoided due to extreme loads exerted on the compressor bearings [80]. The impact of surge, depends on the type of machine and the medium it operates with, however, in some cases only a few surge cycles can be sufficient to damage the machine [81].

Hence, the detection of surge should be as quick as possible, which promotes the focus on surge precursors research [2, 47].

1.4.2 Inlet recirculation as a precursor of surge

Inlet recirculation can appear before surge in some compressor configurations [43, 8], thus its detection can serve as a surge warning. A study of inlet recirculation potential for indicating the onset of surge was undertaken by McKee et al. [41], where they proposed a dedicated device to measure the velocity of the recirculating flow before impeller as a surge onset warning. With this method, they were able to increase the compressor operating range by approaching the surge line closer, while ensuring the safety of compressor operation. They claimed that the presence of recirculating flow was the only consistent and well-detectable surge indicator. However, the recirculation was not observable for all the tested impellers, as they noted a lack of this phenomenon prior to surge in shrouded, radial impellers.

More recent studies performed on turbocharger compressors [16, 40] and industrial compressor [60] noted that inlet recirculation appears, but it is not the ultimate surge precursor as it disappears at some distance before surge. In those studies, the last instability prior to surge was rotating stall or similar structures. Numerical and experimental studies by Poujol et al. [34, 82] performed on an aeronautical compressor with inlet guide vanes demonstrated that rotating instabilities, resembling stall, disappeared at the onset of inlet recirculation. Therefore, inlet recirculation was not a direct precursor of surge in the investigated machines. Hence, the presence of inlet recirculation can be used for detecting surge proximity, but it cannot always be considered the ultimate surge precursor.

1.4.3 Rotating stall as a precursor of surge

In a number of studies it was noted that the onset of surge was preceded by rotating stall [2]. One of the very first attempts at detecting surge precursor was in 1946 by Bullock et al. [83]. Before the compressor surged, they noticed the presence of small, irregular pressure pulsations of low magnitude. It is possible that what they observed was a signature of a rotating stall. A thorough analysis of surge precursors for different compressors was presented by Ribi [2]. The review was based on studies between 1955 and 1991, where surge precursors were noted by the authors of those studies. In most cases, a combination of unshrouded rotors and vaneless diffusers was used. The compressors operated within a wide range of Mach numbers (from 0.13 up to 1). The precursors were mostly rotating stall or mild surge, however a different behaviour of different compressors on a path to deep surge was noted. For a number of cases, the presence of rotating stall was followed by mild surge, but it was not always a rule [2]. Overall, no generalized observations regarding centrifugal compressor surge precursors were made due to differences in reported studies. It might be considered evidence that the differences in machine design and operating parameters transfer into different behaviour in off-design conditions.

1.4.4 Precursors to rotating stall

Rotating stall can be a precursor of surge, but due to its effect on the compressor performance, detection of stall predecessors was also of interest. Stall onset was shown to be detectable with a number of methods. One of them relies on the well-described stall onset mechanism, being modal wave stall [84]. By measuring the characteristics of pressure waves, the stall onset proximity can be defined [84]. It was shown that waves are present even for stable operation of the compressor, but their parameters change when approaching stall [60]. Reaching stall through amplification of the modal waves gives rise to progressive stall. Setting an appropriate threshold allows for early detection of this type of stall

[85], however, it does not aid quick detection of spike-type stall inception, which is more unpredictable and does not have to provide a clear indication of its proximity prior to inception [7]. Liu et al. [85] did show that certain precursors could be found through chaotic behaviour of the pressure signal, but they appear only a few revolutions before the onset, leaving little time for reaction.

To address that issue, another method was proposed. The pressure signal inside the compressor, especially close to the impeller, is strongly affected by the blade passing signature. The blade pass character and its alterations while approaching the unstable region can be employed to detect stall proximity. This is demonstrated in Figure 1.11. With each blade pass, a certain pressure signature is expected. This signature tends to be stable for the design operating conditions and changes when instabilities appear [86]. Deviation from expected pressure trace can be used to assess the onset of stall. It was shown by Dinghra et al. that approaching rotating stall, the disturbances to a standard pressure trace appear more often [87]. This observation can be used for the proximity of onset prediction, however, this approach bears a certain level of stochasticity to it.

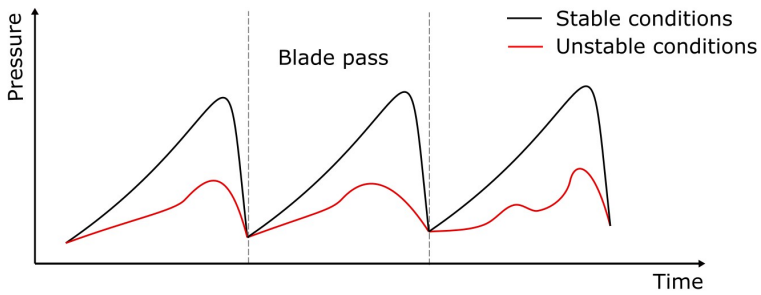


Figure 1.11: Visualisation of a blade pass character change for unstable conditions; depending on the instability, the change can be different both in terms of the character and the frequency of the occurrence

A study by Young et al. [86] confirmed that for an axial compressor, the blade pass pressure trace can be used for stall onset prediction, but the observations differ depending on the tip clearance and the level of eccentricity.

The viability of this approach for centrifugal compressors was proven by Zhang et al. [88] where the character of pressure trace deviation from the ensemble average was quantified in a statistical manner. Computing the irregularity and skewness of the deviations was used as a pre-stall indication. In their study, the presence of inlet recirculation was also noted. It was shown that it precedes the appearance of stall but could also remain present alongside stall in the unstable range.

Another approach is related to the detection of a jet-wake structure that can appear at low flow rate conditions, as shown in Figure 1.12. It gets amplified with decreasing mass flow rate, leading to a point where the flow detaches from the blade and the impeller stalls. At the interface of high and low velocity flow, a Kelvin-Helmholtz instability can be formed [66]. This is a flow phenomenon happening when two zones of different lateral velocities meet, resulting in the appearance of a vortical structure at their interface. Kelvin-Helmholtz instability was shown to appear before the onset of rotating stall in a centrifugal compressor. However, it is difficult to detect in a rotating frame of reference due to interference with rotational speed, but also holds the potential for being an early detection feature [66].

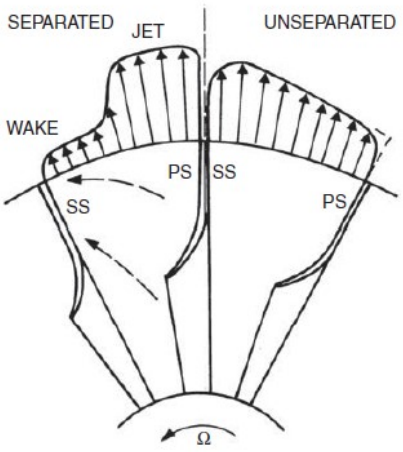


Figure 1.12: Jet and wake structure due to separation of the flow on a suction side [7]

1.4.5 Order of instabilities in centrifugal compressors

The understanding of physics of instabilities has developed over time, with a number of studies in the field [7, 2]. Generalising the behaviour of compressors at low mass flow rates has proven to be challenging due to the significant variability in machine design and the resulting variations in off-design operation.

Recent studies on turbocharger compressors provide more information on the development of instabilities and their precedence [43, 8]. It was shown that the order of instabilities for a given compressor can vary depending on the rotational speed of the impeller [8] (Figure 1.13). For the lowest rotational speeds range (60 to 80% of rated speed), instabilities appeared in order: inlet recirculation, rotating stall, mild surge, surge. At 90% the mild surge was present first before inlet recirculation, which directly preceded surge. For this speed, rotating stall was not observed. At 95% surge was directly preceded by mild surge, while for 97 and 100% there was no apparent precursor present [8]. Similar observations were made by Sun et al. [43]. Considering three rotational speeds of 50, 80 and 100 % of the rated speed, they observed a different order of instabilities for different rotational speeds. For each rotational speed, the ultimate surge precursor was different. It was a mild surge for the lowest speed, inlet recirculation for moderate and rotating stall for the highest.

There are no general rules regarding the instabilities order when decreasing the mass flow. Furthermore, some instabilities may not be observed for some speed lines. Consequently, a comprehensive instabilities detection system should be capable of capturing and distinguishing different instabilities using their signature. Then, the state of the compressor at each instant of its operation can be understood and appropriate reaction can be defined, which would depend on the type of machine and the expected risk associated with the presence of instabilities.

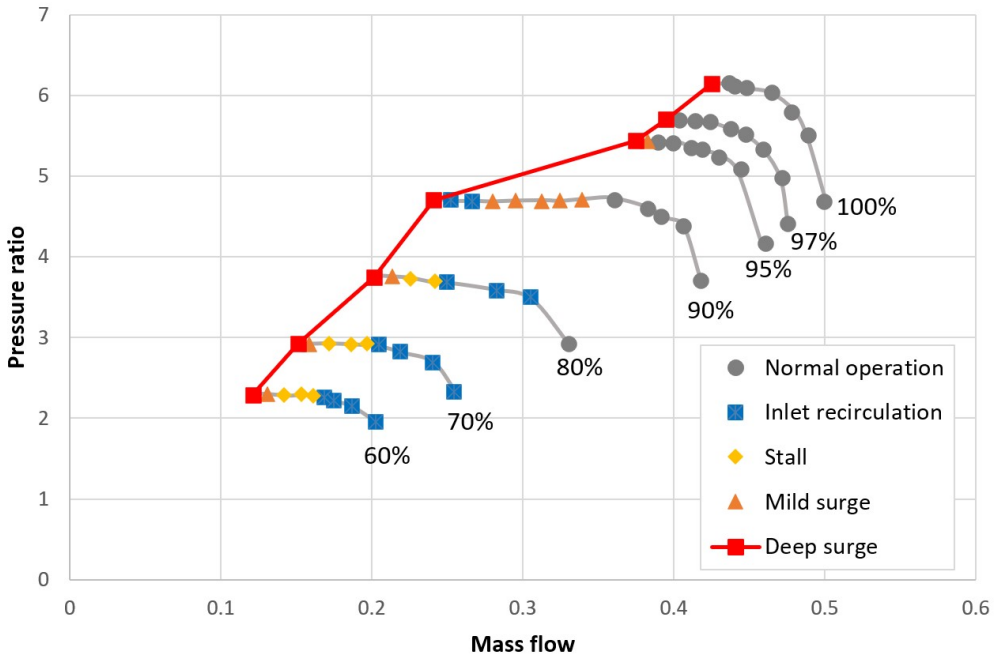


Figure 1.13: Compressor map with instabilities detected for different rotational speed; based on [8]

1.5 Data-driven instabilities detection

To provide a comprehensive detection of different compressor operating conditions, it may be possible to leverage data-driven techniques. The data obtained from the compressor is rarely comprehensible in a raw state. Therefore, it has to be processed to obtain the indication of instability. Different physical signals can be used for detection, such as pressure, vibration or temperature. They can provide different insights into the compressor operation, capturing different aspects of the instabilities. In each case, the raw data has to be appropriately processed to gain insights into the system stability.

Considering the field of instabilities detection in centrifugal compressors, methods of increasing complexity were developed over the years. Through the analysis of the literature, six groups of approaches were defined, based on the main processing idea. Those are as follows:

- Time-domain methods,

- Frequency-domain methods,
- Time-frequency methods,
- Correlation methods,
- Chaos-based methods and entropy,
- Data-driven decomposition methods.

The groups listed above do not form disjoint sets, as the methods overlap to some extent. For example, correlation methods can be considered for both the time domain and frequency domain [18, 65], some chaos indicators are defined based on the time domain statistics of a signal [89], while decomposition methods often provide a number of time series that can be processed with any of the above approaches [90, 91].

The aim of the division proposed in this thesis is to structure the literature by highlighting the main focus of the contributions. It should provide a clearer picture of the field of data-driven instabilities detection in centrifugal compressors. On the basis of the literature review, it is possible to identify the gaps in knowledge and define research directions that address these gaps. Each of the categories is described, providing its main idea in the context of the physical signature of instabilities and listing the most important contributions. The contributions are summarised in Table 1.1, with a short description of the processing idea, the category they were assigned to and the types of instabilities detected.

1.5.1 Time-domain methods

Time-domain methods focus on the signal characteristics and their changes in the time domain. Commonly, these methods exploit statistical parameters such as mean, standard deviation, kurtosis or crest factor, which are expected to change with the onset of unstable conditions. Pecinka et al. [92] used standard

deviation of pressure signal as a measure of a surge onset proximity and have shown that it can be applied to a centrifugal compressor operating in a jet engine. The standard deviation of pressure and temperature were demonstrated to be good indicators of mild and deep surge [46, 93]. Liskiewicz [9] proposed a metric called rate of derivative fluctuations (RDF), based on a pressure signal and its derivatives. It computes the changes in derivatives of the pressure signal over time, detecting when the changes are becoming more abrupt. RDF, related to some extent to standard deviation and kurtosis, was used as a surge proximity indicator [94].

All of the time-domain methods rely on the detection of an altered character of pressure pulsation happening before the surge onset. These methods have been applied to detect surge, but not for inlet recirculation or rotating stall. The biggest disadvantage of the time-domain approach is that in the majority of implementations, it does not offer the possibility to differentiate instabilities. The onset of each unstable structure would cause a change in signal statistics compared to the stable regime. However, the character of those changes may not be differentiable with time-domain statistics. Therefore, time-domain methods can be used for instabilities detection but they do not offer the possibility of constructing a comprehensive instabilities detection system.

1.5.2 Frequency-domain methods

Frequency methods rely on the frequency spectrum and its changes between different conditions to detect instabilities. In this way, a better understanding of the signal composition is obtained, as the spectral character of the observed fluctuations can be understood. It allows not only to detect the presence of unstable conditions but also to define its type.

The spectrum of a signal is often complicated and does not provide a direct indicator of conditions. Hence, it has to be interpreted to provide an indication. To avoid manual interpretation, some methods rely on the summation of

some parts of the spectrum, characteristic of instabilities. Bianchini et al. [58] proposed a criterion for vanless diffuser rotating stall detection based on a sum of the spectrum energy in the subsynchronous range. The presence of rotating stall would be visible in an increased amplitude for stall frequency, which is mostly below the rotational speed of the machine [7]. To account for differences between compressors and offer universality of detection, the sum of the spectrum is normalized by the blade passing frequency energy. This approach was validated by Romani et al. [95], demonstrating its applicability for different machines.

A similar approach was proposed by Courtiade et al. [96], where the energy of frequencies associated with instabilities, normalized by the mean of a static pressure, was used for surge detection. A range of frequencies was used rather than a specific value, as the signature of instability can vary in frequency. The power spectrum of structural response and noise generated by the compressor was thoroughly investigated by Reggio et al. [63]. They showed that significant diagnostic information can be obtained from the vibration signals collected from compressor housing at both, subsynchronous frequencies as well as frequencies much above the rotational speed. The latter might need more sophisticated processing techniques to be extracted and can be more dependent on the compressor housing design that affects the vibration response of the structure.

Spectral analysis of the pressure signal was employed by Liskiewicz et al. [97]. In that study, the aim was to identify the information that could be obtained with the use of a pressure spectrum rather than propose an indicator of instabilities. For each operating point, defined by the level of throttling, a power spectrum of a signal was computed. Combining the spectra, a spectral map was created where the characteristic instabilities frequencies could be identified. This approach may be useful for the identification of the presence and character of instabilities in research settings, but it is not directly transferable

into a detection system because of the need for interpretation and long signal acquisition time.

As much as spectral analysis can help to understand the phenomena taking place in the compressor system, it has its limitations. One of the most commonly used approaches for spectral analysis is Fourier transform (FT). Its underlying assumption is that the power spectrum is constant in time. It is often not the case for signals from centrifugal compressors, especially when approaching unstable conditions. Averaging the power spectrum for long input signals can blur information that could be extracted when investigating the variation of the power spectrum [57, 65].

1.5.3 Time-frequency methods

Time-frequency methods deal with the shortcomings of the frequency methods as they allow tracing the changes in the power spectrum in time. Time-frequency approach also helps to identify changes in signal structure that might be used for quantification of intermittent compressor states and detection of instabilities onset proximity [65].

Time-frequency representation can be obtained with a number of methods, that differ in performance [98]. One of the popular methods is the short-time Fourier transform (STFT), which extends the use of FT to capture time-variability. It was shown to be applicable for predicting surge onset in compressors [65, 99]. Although commonly used, the method suffers from time and frequency resolution limitations which arise from Gabor limit [100].

A better frequency resolution can be obtained with continuous wavelet transform (CWT). It was introduced in the application to compressors instabilities detection by Liao et al. [101], and researched for other machines by Horodko [57] or Brown et al. [102]. Surge and stall were targeted in those studies, and it was shown that CWT can be useful for instabilities detection and that it can provide quicker information of instabilities onset than Fourier transform [102].

The interest in the application of wavelet analysis has been revived recently with studies from Wang et al. [64], Liu et al. [85], Silvestri et al. [103] or Zhou et al. [104]. In those studies, the applicability of wavelets-based time-frequency analysis was studied, demonstrating again that it is possible to differentiate stable and unstable conditions with CWT. However, in most cases, no specific indicators of instabilities were proposed. Despite the high potential of CWT, several concerns about its application were raised [105]. They regarded mostly the fixed base of the transform, which can decrease the accuracy of the indication if the chosen mother wavelet does not conform well with the signal characteristics [106]. Some concerns regarding the pace of processing were also voiced in the context of using CWT for online detection, indicating it is a relatively slow method [105].

A different method to obtain a time-frequency distribution is through Wigner-Ville distribution (WVD) [107]. It offers very good resolution in both time and frequency, but is highly computationally intensive in both time and memory needed, therefore it was not widely considered for real-time instabilities detection [65]. However, the high time and frequency resolution of WVD make it a very good tool for research and visualisation purposes [65, 108].

Time-frequency methods can provide a very good insight into the overall structure of the signal, capturing the temporal changes in its spectrum. However, they tend to be either computationally expensive, like WVD or limited in their resolution, like STFT [65]. In addition, those methods provide a complex output that has to be interpreted. Translating it into an indication is not straightforward, hence the time-frequency domain methods often serve as a diagnostic and research tool [65].

1.5.4 Correlation methods

Correlation methods are another set of approaches differentiated in this thesis. They often rely on the time-domain or frequency-domain methods, but leverage

those approaches for a specific aim. Correlation methods focus on the identification of coherent structures in the signal, which are either amplified or damped with the onset of instability [65].

One of the approaches is based on the correlation of blade pass signature [86]. The pressure signature obtained from a single sensor located close to the impeller is compared for subsequent blade passes. It has been shown for an axial compressor that when operating at stable conditions, the correlation between subsequent blade passes is high. However, when the compressor approaches unstable conditions, the correlation of certain passes drops [86]. This drop can be associated with a local flow detachment, which at the early stage of instability development is quickly suppressed and does not lead to the onset of rotating stall. Dhinghra et al. [87] have shown that a singular low-correlation event may happen far from the actual instability onset. Such an event is to a certain extent random, therefore its isolated occurrence is not a valid indicator of change in conditions. However, those events are becoming more often when approaching unstable conditions, hence they can be considered precursors of instability onset. To quantify the frequency of occurrence, the distance between subsequent drops in correlation coefficients can be used [87].

There exist a number of factors, such as eccentricity or tip gap that can influence the metrics based on pressure correlation [86]. The blade pass signature is stronger for axial compressors than for centrifugal ones. To obtain a clear indication of the blade pass of a centrifugal compressor, a number of conditions has to be met. It has to be a high-pressure unit and the acquisition system must be able to provide high sampling frequency for sufficient resolution on high-speed machines. It is still possible that for highly off-design conditions, the blade pass signature becomes undistinguishable [109].

An extended study of using the blade pass correlation for centrifugal compressors was presented by Zhang et al. [88]. Similarly to [87], they measured

the deviation from a pressure signature for a number of blade passes. Furthermore, they accurately measured the magnitude of the deviation and statistically analysed the data, using the irregularity and skewness of the deviations as indicators of stability. Differences could be seen between stable, inlet recirculation, stall and surge conditions with this approach. However, considering the less clear waveform of a centrifugal compressor pressure, the indication was based on statical parameters of the obtained differences for a number of blade passes, rather than separate measures for each blade pass [88]. This implies that a larger number of blade passes must be observed for centrifugal compressors than for axial ones when using a method based on blade pass correlation.

Certain aspects of correlation and repeatable behaviour of a signal can be explored by cyclostationary analysis. This framework offers a way for detailed investigation of cyclic components in the signal and detecting modulations of different nature [110]. The basic idea of cyclostationarity is that a signal has periodic components, which can be extracted from it, for example through ensemble averaging. Then, different aspects of the remaining part can be used for construction of indicators [110]. It has been widely applied in many branches of condition monitoring, but its application to pressure signals was limited. Jurdic et al. [108] proposed a cyclostationary spectral analysis to study a wake character from a low-speed fan. The objective of the study was to examine the relationship between turbulence characteristics and the generation of noise. Interestingly, a similar approach could also be applied to distinguish between stable and unstable conditions. Stajuda et al. [109] aimed to exploit cyclostationary for the detection of instabilities in a centrifugal compressor. Although the study showed some potential, it was not fully conclusive due to marginal, but detectable variability of the rotational frequency of the compressor.

It might be possible to detect instabilities through the vibration signature of the compressor elements. Botero et al. [61] applied cyclostationary analysis to detect rotating stall in a pump-turbine, using cyclic spectral coherence of

vibration signal obtained from the diffuser vanes. The detection of stall was successfully demonstrated for the instability that was developed, with precise measurements taken from a strategically positioned vane at the anticipated stall area. If the vibration signal from the housing of the machine was to be used, the identification of the instability might have been more difficult due to a transfer function dependent on the housing mass, material and geometry [63]. Vibration and acoustic signature were investigated with cyclostationarity by Munari et al. [111]. It has been demonstrated that stall and surge can be identified by analysing the vibration response of the compressor. However, it should be noted that the findings may vary depending on the specific flow system to which the compressor is attached.

The challenge in applying cyclostationary analysis to instabilities detection in centrifugal compressors might arise from the physical characteristics of the signal. There are two possible sources of periodicity in the compressor signal. The first one is the impeller rotation. Since the rotational speed may fluctuate or change in time, the signal is then periodic in the domain of the shaft rotation angle. The other source are instabilities. The periods of instabilities are not directly related to the rotational speed of the compressor. For example, surge frequency is dependent on the whole compressor system [22], therefore it should be considered in the time domain. Thus, cyclostationary approach, well suited for vibration signals where there is often a single source of periodicity, may not be advantageous for instabilities detection.

Another correlation-based approach described in the literature is based on cross-correlation between two sensors. This configuration is commonly used in the research of rotating stall [55, 56, 57]. The sensors are often located at the same radius but shifted in an angular position. This way, it is possible to detect structures travelling around the annulus and define their velocity by investigation of the signal shift between the sensors [56]. The shift between the sensors has to be carefully chosen to avoid uncertainty related to the number

of cells and their speed [50]. With a pair of sensors, it is also possible to detect rotating pressure waves that would otherwise be indistinguishable due to noise in the signal [60]. Such pressure waves may be considered a precursor of rotating stall [7]. Two practical correlation-based approaches were demonstrated for stall detection, with Horodko [60] using the correlation of CWT coefficients, while Bianchini et al. [56] took advantage of signal cross-correlation.

The approach based on cross-correlation is suited mostly for the detection of rotating stall and its precursors. It requires two sensors to operate, but it only allows one to identify a single type of instability. The detection of other instabilities requires different types of signal processing, therefore the cross-correlation method would not provide a comprehensive instabilities detection capability.

The correlation analysis is a powerful approach, targeting the physical aspects of instabilities. However, it has a number of practical limitations. To effectively use the correlation of the blade pass signature, it is important to identify a specific location where a coherent cyclic signal can be obtained. For a centrifugal compressor, it may be over the blade or in the diffuser, as close to the impeller as possible to avoid the influence of mixing. The measurement over the blade not be possible if a shrouded rotor is considered. If a machine has a relatively low pressure ratio, the blade pass signature might quickly lose its cyclic character with decreasing mass flow [109], increasing the uncertainty of the indication. What is more, the variation of rotational speed can cause disturbances in time-resolved measurements. To avoid that issue, the measurements should be related to or at least resampled to the shaft angle [110]. A disadvantage of the time-domain correlation approach is the need for a distinctive blade-pass signature and sufficient sampling frequency to provide a certain blade-pass resolution. This is mostly possible for axial compressors, where a strong and undisturbed marking of a blade passing is visible in the pressure

signal and the rotational speed is often lower than for centrifugal compressors [10].

1.5.5 Chaos-based methods and entropy

Another type of approach relies on quantification of the chaotic behaviour of the signal. It might be considered complementary to other approaches such as cyclostationary approach or time-frequency methods, which focus on detecting repeatability and patterns in the data [110]. When approaching unstable operation, the mass flow decreases and the pressure signal in the compressor becomes more chaotic and some erratic pulsations can be observed [83]. The pulsations can have a specific frequency, as expected for rotating stall and surge or be more random, such as for inlet recirculation or some transient states [43].

The pressure signal from inside the compressor is inherently chaotic at some scales due to the turbulent character of the flow [112]. Both random and periodic structures coexist in the compressor, but their relation can change for different operating conditions [97]. There exist a number of methods that can be used for the assessment of the chaotic behaviour of the signal.

One of the approaches exploits the concept of entropy. In information theory and signal processing, entropy can be understood as a measure of the rate of new information generation in the signal [113]. If a signal is fully deterministic, the entropy is low. If it is chaotic, the entropy is high. There exist a number of entropy measures, which differ in computational procedure and the exact interpretation [114].

One of the measures applied to compressor conditions classification was based on sample entropy. Zhang et al. [115] proposed the application of improved sample entropy for the detection of rotating stall in an axial compressor. It was shown that due to the increasing complexity of the signal before the onset of stall, an increase in entropy takes place. That increase precedes the onset of stall by several revolutions, thus such feature can be used for early detection.

An extensive study of approximate entropy for instabilities detection in centrifugal compressors was performed by Lou et al. [116]. It was shown that an increase in entropy can be used to detect the onset of rotating stall. Similar results were obtained for two machines of different characteristics, proving to some extent the universality of the approach.

Another study employing a different measure of entropy was done by Hong et al. [117], where permutation entropy was used to detect rotating stall in an axial compressor. Similarly to the studies investigating sample and approximate entropy [115, 116], a change in entropy was noted for the stall inception region, making differentiation of this state from stable operation possible.

Although entropy can provide a good understanding of machinery conditions, it is very sensitive to the choice of parameters. Approximate entropy requires four different parameters to be set. Their incorrect or sub-optimal choice can have important consequences on the detection performance [116].

Another approach coming from the field of chaos is based on Hurst exponent. Hurst exponent can be used for quantification of correlation and memory of a time series [77]. With this approach, the level of pressure fluctuations and their coherence at different scales can be quantified. The algorithm also requires empirical input regarding the choice of the timescale for computation, but it was shown by Kerres [89] in a comprehensive study that Hurst exponent can serve as a surge indicator in centrifugal compressors. Hurst exponent, was also investigated by Liu et al. [118] for surge detection. It was demonstrated that it can capture the transition to surge, however its advantage over the variance of the pressure signal was not clearly evidenced. Kerres [89] noted that for the Hurst exponent to be indicative of surge conditions, a sample containing a number of surge cycles has to be used. Hence, such a system may provide too low responsiveness to be considered a valid anti-surge solution [18].

Overall, the methods based on chaos and entropy are a potential solution for condition monitoring in centrifugal compressors. However, their interpretability

remains limited as with changes of parameters, it is possible to obtain entirely different results [116]. The need for optimisation of parameters, their high number and interdependencies make the application of those methods challenging and negatively affect the interpretability. In addition, the prevalence of chaos-based methods over other approaches is not proven as none of the presented studies tried to compare entropy-based detection to more standard ones, while some hints regarding the comparison show that time-domain statistics could give similar results [118]. The computationtion of entropy is also a resource-intensive process, rendering this approach challenging in implementation for real-time compressor monitoring [113]. Most of the applications of entropy found in the literature focus on a single instability, either rotating stall [116] or surge [117], which may suggest that the detection of multiple instabilities with the same set of parameters is challenging.

1.5.6 Data-driven decomposition methods

Another set of methods used for instabilities detection are data-driven decomposition techniques. They process a signal and divide it into a number of components based on different principles. Those methods are termed data-driven as the decomposition base is dependent on the data and not fixed as in the case of the Fourier transform or wavelet transform. Adaptability to the data could provide better decomposition results, compared to the fixed-base decomposition [119].

The aim of applying the decomposition in signal processing can be very different, ranging from removing the noise or discovering the trend [90, 120], isolating the signature of interest through a single or a number of components [91] or separating different sources in the signal [121]. There exist a large number of methods that differ in principles of the decomposition [119]. Most of them can deal with signals of different origins and physical character. The signals from mechanical sources are often complex and composed of a number

of overlapping sources. Hence, the decomposition methods are a commonly used approach in the field of condition monitoring or fault detection to separate different components of the signal [122].

There are several data-driven decomposition methods used in the field of condition monitoring. The most popular ones can be divided into two families, based on the processing idea they exploit [119]. The first family takes advantage of the signal waveform shape in the time domain. An important method of this kind is empirical mode decomposition (EMD) proposed by Huang et al. in 1998 [90]. A number of methods followed, exploiting a similar concept, such as intrinsic time-scale decomposition (ITD) [123], variational mode decomposition (VMD) [124] or local mean decomposition (LMD) [125].

The other family of approaches used in the field takes advantage of the matrix decomposition. The methods include principal component analysis (PCA) [126], singular value decomposition (SVD) [127], dynamic mode decomposition (DMD) [128] or singular spectrum analysis (SSA) [91]. PCA, SVD and DMD are often applied as dimensionality reduction techniques in a number of domains [127, 128] but can also be applied for component extraction in signal analysis for fault detection and condition monitoring [129, 130]. SSA is more commonly used for signal decomposition in condition and structural health monitoring [131], as it is suited for the decomposition of time series [91].

Despite important advances and the maturity of a number of decomposition methods used for condition monitoring or structural health monitoring, the field of instabilities detection using decomposition methods remains relatively unexplored. With a vast number of studies performed in condition monitoring and fault detection in mechanical systems as indicated in review papers [132, 133], decomposition methods are a comprehensive set of tools that have the potential to be used for robust and sensitive condition monitoring. The problem of aerodynamic instabilities detection in centrifugal compressors is, in principle, similar to condition monitoring [119]. However, in detail, it differs from that of

mechanical faults or structural damages. Firstly, the structural damage to be detected is often permanent and can only aggravate over time, which is not the case for aerodynamic instabilities. Secondly, the intermittency of instabilities requires the detection to be quick, therefore the processing time as well as the input signal length required for detection have to be limited. Consequently, the developments in the field of condition motoring cannot be directly transferred into instabilities detection without prior investigation for compressor monitoring.

With the vast field of decomposition methods, it is difficult to provide a comprehensive comparison of all approaches. In a number of comparative studies, the authors acknowledge the need of restricting the number of methods used in comparison for the sake of clarity as well as due to the expectation of rendering similar results with closely related methods [122, 134]. The application of decomposition methods to compressor instabilities detection was limited, hence there are plenty of opportunities for the exploration of different approaches. Having two families of methods based on different processing approaches, it is chosen to investigate how the representative of each family performs for instabilities detection. This could serve as a solid base for a comprehensive analysis of the methods within the same family. This thesis focuses on two methods: EMD, the most mature approach among empirical decompositions, and SSA, a method well suited for time series analysis.

EMD [90] was initially developed for geoscience purposes, but it has been widely used in many other applications, including gears [135], bearings [136] as well as flow problems [137]. EMD was widely applied in the field of condition monitoring, as evidenced in the reviews [132, 138]. EMD decomposes a signal by fitting an envelope to the extrema of its waveform and processing it in an iterative manner. The process of extracting components is called sifting and the components are termed intrinsic mode functions (IMFs) [90]. The base of decomposition is dependent on the input data, making EMD highly adaptive.

The components extracted from the original signal are expected to be interpretable and meaningful, representing simple oscillations present in the system [90]. In applications to very complex signals, the components tend to represent specific frequency bands rather than simple oscillations. Therefore, to some extent, EMD can be considered a filter, with a bandwidth dependent on the number of sifting iterations [139]. The IMFs are limited bandwidth and ordered by their central frequency, with IMF 1 holding the highest frequencies, while the last extracted IMF holds the lowest ones [90]. By altering the parameters of EMD, it is possible to influence the width of a frequency band assigned to the component and the total number of components [90].

EMD has a number of shortcomings which were identified within several studies and summarised by de Souza et al. [140]. The most important ones are end effects and mode mixing. End effects arise due to the procedure involved in EMD, caused by the problem of fitting the envelope to the data at signal ends [140]. The effect is notable when considering short signals and their time-frequency representation, where the end effects can affect a major part of the analysed sample. With longer signals, the influence of end effects on the analysis is less significant. What is more, the end effects can be to some extent addressed by extending the signal at both ends [141]. There is a number of techniques employed for signal extension [140]. In most applications, addressing the end effects with any form of signal extension should not greatly influence the outcomes of the decomposition, while reducing the impact of end effects [141].

The mode mixing phenomenon [90], is defined either as a situation where a single IMF consists of waves of widely disparate scales, or waves of a similar scale being present in different IMFs [140]. Often, the noise present in the signal is defined as a cause for mode mixing [90]. If the noise alters the distribution of extrema in the signal, the decompositions of the noisy data and noise-free data can be significantly different [140]. Mode mixing can lead to challenges in interpreting and analysing the IMFs, as the mixed modes can obscure the true

characteristics of the underlying signals. However, in certain cases, mode mixing may be acceptable depending on the specific application, should appropriate method parameters be selected [142]. It was also demonstrated that EMD in its original form can perform well despite mode mixing [143].

Over the years, improvements to EMD have been proposed, with a number of them addressing mode mixing. Ensemble empirical mode decomposition (EEMD) [144] assumes an addition of white noise to the signal before decomposition and performing a sufficient number of independent decompositions that are subsequently averaged. The presence of noise forces the algorithm to explore all possible paths in the process of creating the components. A sufficient number of trials is necessary to average out the influence of the added noise. EEMD was shown to be less prone to mode mixing, compared to EMD. However, the procedure for EEMD requires numerous repetitions of the decomposition process, with a suggested number of trials being around 100 or higher [144]. It makes the method several orders of magnitude more time-consuming than the original EMD.

Another variation to EMD is complementary ensemble empirical mode decomposition (CEEMD) [145], which addresses some of the downsides of EEMD. In CEEMD, a pair of realizations of a signal are used. For one, the white noise is added and for another, it is subtracted. This limits the number of decomposition procedures that have to be performed. As much as the method allows decreasing mode mixing effects, it is sensitive to the choice of white noise parameters [145] and still requires performing more than a single decomposition. Therefore, in a number of applications, the original EMD approach remains a viable method [138], especially in applications where the mode mixing and end effects are not very important [143].

The IMFs are time series, that often have to be furtherly processed to extract meaningful information. Several approaches to the processing of the IMFs were proposed in the literature. In the original manuscript of Huang et al. [90] EMD

was coupled with Hilbert transform. It gave rise to Hilbert-Huang transform (HHT), capable of providing a time-frequency representation of the signal [90]. This approach, although useful for the analysis of the signal, is not suited to directly provide indicators of conditions. It rather helps to obtain time-frequency representation, similar to CWT or STFT. A number of studies focused on the energy of IMFs, demonstrating that changes in energy can provide diagnostic information. In those studies, either the absolute energy level of a given component was used as an indicator [146] or it was the relation of the energy of selected IMFs [135]. In some studies, also other types of measures were used for IMFs, such as statistical parameters [147] or entropy [148]. In practical terms, any method applicable to time series could be used for the analysis of IMFs.

Despite a wide range of applications in many domains [132, 138], the use of EMD and related decomposition methods for instabilities detection has been limited. Basic studies confirming the potential of EMD for rotating stall [149] and surge [78] have been published, however, a number of issues such as pace of detection or influence of the changes in the parameters of decomposition were not addressed. More recently Liu at al. [148] demonstrated the possibility of surge detection with EEMD by investigating the changes in the entropy of the components. Yue et al. [150] have demonstrated that EMD can perform well in obtaining a time-frequency representation of the data from a pressure sensor in an axial compressor. They also demonstrated a possibility of improvement of the representation by combining EMD with LMD. Wang et al. [151] used VMD to investigate the time-frequency changes in the pressure signal from a centrifugal compressor. They have shown that data-driven decomposition of a pressure signal can provide a very good insight into compressor operation.

EMD-related methods were also shown to be useful for the characterization of the turbulence in the flow [112, 152], which might suggest that if the presence of instabilities would cause a difference in turbulence scales, it could be captured with such methods. Therefore, EMD has the ability to offer various insights into

the pressure signal, which can be utilized to detect and characterize different types of instabilities.

Signal decomposition with a fundamentally different idea than EMD is offered by SSA. SSA is a principal component analysis (PCA) extension to time series analysis [91]. SSA was initially applied mostly to atmospheric or geoscience data [91]. Over the years, its application was extended to biomedical signals [153], condition monitoring [154], structural health monitoring [155] or studies on turbomachinery [156]. SSA decomposes a signal into a finite number of independent components, ordered according to the variance content in each of them [91]. These components are called reconstructed components (RCs). SSA is well suited for processing non-linear and non-stationary signals of different origins. RCs are time series of the same length as the input signal and can be furtherly processed.

With SSA, the signal is transformed into a matrix form by collating a number of lagged vectors, produced based on the original signal. Such a matrix undergoes an eigenvalue decomposition and is subsequently reconstructed. The details of the procedure can be found in [91]. The RCs are ordered by their variance in decreasing order, therefore each subsequent component has a lower contribution to the reconstruction of the original signal. SSA extracts all rotational patterns included in the signal, also the non-linear ones [91]. RCs can be used for trend identification, detection of oscillatory components, periodicity extraction, signal smoothing, noise reduction, feature extraction and detection of structural changes in time series.

SSA is often used for filtering the signals, assuming the low variance components are due to noise and can be removed for obtaining a cleaner signal [91]. Then, a signal is reconstructed using only a selected number of components [91]. A different approach is to investigate the components independently, assuming that selected ones hold the information of interest [156, 157].

The challenge in the application of SSA lies in the selection of the intrinsic parameters of the method. The crucial parameter is window length L, which defines the embedding dimension in the decomposition. A number of studies focused on defining the window length for different types of analysis, but the issue remains mostly case-specific [158, 159]. With too long a window, the signal may get over-decomposed, which introduces the need for grouping of the components [91]. By increasing the window length, the decomposition time also increases. It is related to the size of the matrices to be used for computation, which are defined by the window length. Therefore, when applying SSA for instabilities detection, the goal is to find an appropriate window length in order to obtain meaningful components, but avoid the need for grouping of the components and limit the computation time required.

The application of SSA and related methods in the field of compressor instabilities detection was limited. One of the first studies implementing a method similar to SSA, was done by Komatsubara and Mizuki [160], where they used SVD to improve the phase portrait representation of the system for stall and surge. The SSA was applied to instabilities detection by Garcia et al. [156], where they aimed to find the components that could be used directly for the detection of surge in a centrifugal compressor. The work of Logan et al. [47] has demonstrated a similar analysis, but aimed at finding the indicators of inlet recirculation. It was shown that the signature of inlet recirculation can be obtained through investigation of the energy of specific RCs. The quality of indication and selection of RCs was highly dependent on the choice of intrinsic parameters of SSA, mainly window length. The study did show the potential of the method, however, the authors did not focus on demonstrating its feasibility for real-time condition monitoring. Kozhukhov et al. [157] have used the autocorrelation of a selected RC to identify the presence of rotating stall in a centrifugal compressor. They focused on periodicity related to rotating stall to detect the instability.

Table 1.1: Data-driven methods for detection of aerodynamic instabilities

First author, year	**Method description**	**Method type**	**Instability**
McKee, 2005 [41]	Increased lateral flow velocity before impeller	Time domain	IR
Pecinka, 2017 [92]	Standard deviation, kurtosis and crest factor	Time domain	S
Liu, 2013 [93]	Standard deviation of pressure and temperature	Time domain	S
Liskiewicz, 2019 [94]	Derivatives of pressure signal	Time domain	S
Bianchini, 2019 [58]	Energy spectrum in subsynchronous range	Frequency domain	RS
Reggio, 2018 [63]	Spectrum of vibration and acoustic signals	Frequency domain	IR, RS, S
Romani , 2021 [95]	Energy spectrum in subsynchronous range	Frequency domain	S
Courtiade, 2013 [96]	Normalized instabilities frequencies amplitude	Frequency domain	S
Lieskiewicz, 2014 [97]	Spectrum of pressure signal	Frequency domain	IR, S
Dehner, 2013 [161]	Sound pressure level spectrum	Frequency domain	S
Bousquet, 2016 [66]	Spectrum of pressure signal	Frequency domain	IR, RS
Liskiewicz 2015 [45]	Continuous wavelet transform	Time-frequency	IR, S
Kryllowicz, 2002 [57]	Continuous wavelet transform	Time-frequency	RS, S
Botero, 2014 [61]	Cyclostationarity	Time-frequency	S
Champ, 2021 [62]	Wigner-Ville distribution, wavelet transform, short-time Fourier transform	Time-frequency	RS
Wang, 2019 [64]	Continuous wavelet transform	Time-frequency	RS
Oakes 2002 [99]	Time-frequency representation of pressure signal	Time-frequency	RS, S
Liu 2019 [85]	Wavelet analysis	Time-frequency	RS
Liao 1996 [101]	Wavelet transform	Time-frequency	RS
Brown [102]	Wavelet analysis	Time-frequency	RS, S
Zhou, 2021 [104]	Wavelet transform	Time-frequency	S
Champ, 2020 [65]	Autobispectrum components sum in subsynchronous range	Correlation	RS

Horodko, 2006 [60]	Correlation of continuous wavelet transform coefficients	Correlation	RS, S
Young, 2012 [86]	Correlation of blade-pass pressure trace	Correlation	RS, S
Dinghra, 2007 [87]	Correlation of blade-pass pressure trace	Correlation	RS
Zhang, 2022 [88]	Correlation of blade-pass pressure trace	Correlation	RS, S
Silvestri, 2021 [103]	Autocorrelation of vibration signal	Correlation	RS, S
Stajuda, 2020 [109]	Cyclostationarity	Correlation	IR, S
Munari, 2018 [111]	Cyclostationarity	Correlation	IR, RS
Kerres, 2017 [77]	Hurst exponent	Chaos and entropy	RS, S
Zhang, 2022 [115]	Improved sample entropy	Chaos and entropy	RS
Lou, 2020 [116]	Approximate entropy	Chaos and entropy	RS
Hong, 2020 [117]	Permutation entropy	Chaos and entropy	RS
Liu, 2019 [118]	Fractal dynamics and entropy	Chaos and entropy	S
Xue, 2019 [162]	Bifurcation of pressure signal	Chaos and entropy	RS
Liu 2020, [148]	EEMD combined with entropy	Decomposition	S
Li, 2011 [149]	EMD and Hilbert Transform	Decomposition	RS
Yue, 2020 [150]	Joint EMD and local mean decomposition	Decomposition	RS
Wang, 2022 [151]	Variational mode decomposition	Decomposition	IR, RS, S
Garcia, 2014 [156]	SSA and phase portraits	Decomposition	S
Kozhukhov, 2020 [157]	SSA and autocorrelation function	Decomposition	RS
Komatsubara, 1997 [160]	Singular value decomposition	Decomposition	RS, S
Liu, 2022 [163]	Probability distribution of a signal reconstructed using Fast Wavelet Transform	Decomposition	S

IR - inlet recirculation; RS - rotating stall; S - surge

Both EMD and SSA demonstrate strong potential as signal decomposition methods. They were extensively applied in the field of condition monitoring of machines and structures, providing very good results [138, 164]. From the available studies [47, 78, 151, 156] it could be concluded that further investigation of the decomposition methods for the purpose of building a comprehensive compressor monitoring system is recommended.

1.5.7 Selection of a signal type for monitoring

A data-driven detection of instabilities is based on a live signal collected from a machine. There are several physical quantities that can be measured with different sensors and at different locations. From a practical point of view, the measurement should introduce as little change to the compressor operation as possible. It is also important to have it robust, easy to maintain and low cost. Those requirements tend to be contradictory in many ways.

Various physical quantities can be used for indication, such as acceleration caused by pressure [45], vibrations [65], acoustic emission [111] or temperature [161]. The selection of the signal may differ depending on the compressor system and the task of the monitoring setup. Measurements of some physical quantities may be easier to perform than others, but they tend to be more difficult to interpret.

Considering the difficulty of system adaptation and performing the measurement, measuring noise with a microphone placed nearby the machine is the easiest [65, 111]. No change to the compressor system is also introduced with a vibration sensor mounted on the housing of the compressor [65]. Pressure and temperature measurements are more difficult, as they require introducing some alterations to the system. The pressure measurements are often performed with high-frequency sensors, mounted flush to the compressor inner walls [45, 58]. Temperature measurements for compressors are performed in the flow, therefore a thermocouple has to be placed in the flow path [42, 161].

Being the easiest to perform, both acoustic and vibration measurements are often difficult to interpret, especially for a complex system such as a centrifugal compressor [65]. Vibration measurements are affected by the transfer function of the compressor elements [119], while acoustic data is prone to background noise [65]. Temperature measurements were shown to be applicable for the detection of a number of instabilities [46, 161], however, they offer a lower time resolution of measurement and have larger inertia than pressure measurements. The effect of temperature changes is dependent on the compression ratio of the machine, hence the detection for low-pressure units may be less accurate. The thermocouples have to be located in the flow field, creating an obstruction to the flow while being susceptible to damage [46]. Pressure signal measurements are the most commonly used for capturing instabilities as well as their differentiation [47, 58, 62, 77]. They offer high temporal resolution of the measurement, which can be coupled with acquisition system supporting high sampling frequency. Mounted flush to the walls, they do not obstruct the flow and are not susceptible to excessive wear. Modern sensors are also miniaturised, enabling their installation even for the smallest turbochargers.

Flow instabilities are aerodynamic structures that have a significant impact on the flow and pressure within the machinery. The variations of the flow induced by instabilities are the source of vibrations or acoustic emission, which can also be measured for the purpose of detection. Using the signal registered at the origin of instabilities should provide the quickest and most sensitive indication. Pressure measurements were shown to be capable of providing quick information about the onset of instability [62]. The downside of pressure measurement is that it is often highly noisy, as flow turbulence has different timescales, which are captured by the sensor. This problem can be addressed by applying signal processing methods that can distil and enhance the information present in the signal.

1.6 Methodologies for instabilities detection and classification

A general data-driven detection and classification process used in many branches of the industry follows a number of general steps [119]. First, the signal is acquired and conditioned. Subsequently, it is processed to extract features that can provide a better representation of the data in the context of the application. The features can then be used for inference about the observed system. The inference can be direct if the difference in the conditions is clear and a low number of features is used. In more complicated situations, the decision process can be supported with a classification algorithm.

The goal is to propose appropriate methods for each step of the methodology to allow differentiation of the operating conditions or states of the system. The crucial aspect in data-driven detection is obtaining the most informative features, as they are vital for the accuracy of detection and classification [119].

1.6.1 Features extraction and feature space representation

The process of feature extraction is not straightforward and has been a subject of research for a number of scientists [165, 166].

Feature extraction is a process where relevant information is extracted from raw data to create a meaningful representation of this data. Different signal processing techniques can be used for obtaining features and the extraction approach should be suited to the investigated problem. Using features based on experimental data, it is not possible to avoid the effect of noise. It can be attributed to measurement errors or the unsteadiness of the process. The latter is especially the case with pressure measurements in centrifugal compressors, as turbulence at different scales are affecting the measurement. The presence of noise leads to the dispersion of a feature, even if it represents the same operating

conditions. The dispersion is understood as a width of the distribution of feature values with respect to the mean value. The dispersion can differ, depending on a number of parameters such as the processing method, its parameters or signal length. This results in the need to consider the stochastic nature of the features.

The features should allow differentiating the conditions of the system, so they are to have a different value for different operating conditions. Using EMD or SSA, the obtained features can be considered interpretable [90, 91] as specific physical aspects of instabilities are targeted. Hence, using data-driven decomposition techniques, it is possible to obtain interpretable features that can be related to instabilities. The features can be based on specific components that best isolate and highlight the physical characteristic of instabilities [47], a combination of components [91] or some metrics derived from the obtained components [148]. It was shown that data-driven techniques can be used for extraction of instabilities features in centrifugal compressors [47, 148, 162].

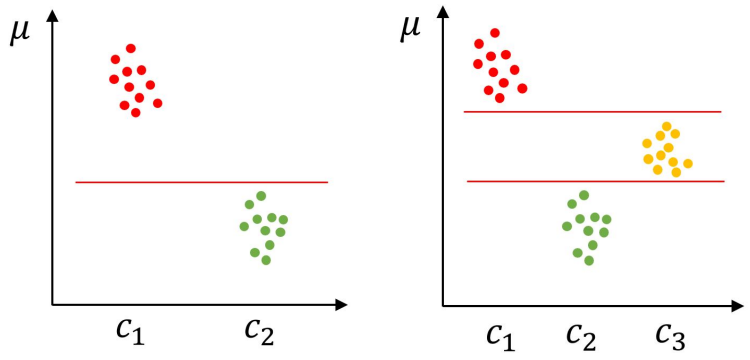
With the decomposition techniques, it is possible to increase the dimensionality of the signal by producing a number of features from a single measurement. This can allow the differentiation of instabilities with a number of features obtained using the same decomposition method and the same input signal. It can importantly increase the responsiveness and limit the complexity of a detection system. It is desirable to have a single feature for distilling a signature of each instability, as then every instability can be monitored independently. The features can also be used jointly for a compound indication of conditions.

The increased dimensionality of the data can be explored through the feature space. The feature space represents the set of all possible combinations and values of features that describe the data. Each measurement is typically represented as a data point, the location of which is defined by the vector of feature values. The feature space dimensionality is defined by the number of features used. By representing the data in the feature space, it is possible to take

advantage of all features in a joint manner and apply a number of algorithms for data classification.

1.6.2 Decision-making

There are several strategies that can be used for inferring the condition of the system based on the features. In the most basic approach, a single feature may be considered for the representation of two different conditions present in the system, as shown in Figure 1.14a. Then, the differentiation can be made based on the value of this feature. The decision threshold can be introduced to enable the classification of new points. It is possible to establish such a threshold based on a single class, consequently, the other class does not have to be sampled. In the case of compressor instabilities detection, it implies that the compressor does not have to be operated in an unstable regime. It helps to limit the risk of machine damage while obtaining the data for the classification model.



(a) (b)

Figure 1.14: Feature values for different number of classes to be differentiated with a single feature a) two conditions; b) three conditions; differentiating three conditions with a single feature requires establishing two thresholds, marked with red line

The interpretable threshold approach is the most commonly met in the literature on instabilities detection in centrifugal compressors [47, 58, 78]. The authors show that a certain feature has different values for stable and unstable conditions therefore they can be distinguished by this feature value. It is also

possible that a given feature takes some distinct values for different conditions (Figure 1.14b) and this property can be used for their differentiation [162]. The disadvantage of such an approach is that the thresholds can no longer be established based on a single class of conditions and at least two out of three classes have to be mapped.

Most of the studies focus on a single instability, however by increasing the number of features, more states can be differentiated. It was demonstrated that by examination of the pressure signal, it is possible to detect different instabilities within a compression system [8]. However, a comprehensive method for such detection without the need for expert examination was not proposed.

Most of the physical processes are burdened with noise that results in the dispersion of the features based on such signals. The extent of dispersion can depend on a number of factors, including the process dynamics, input length and noise level. For accurate classification, the features for different conditions should not overlap in the feature space. Hence, it is important to ensure that not only mean values are different, but also that dispersion is low enough with respect to the means.

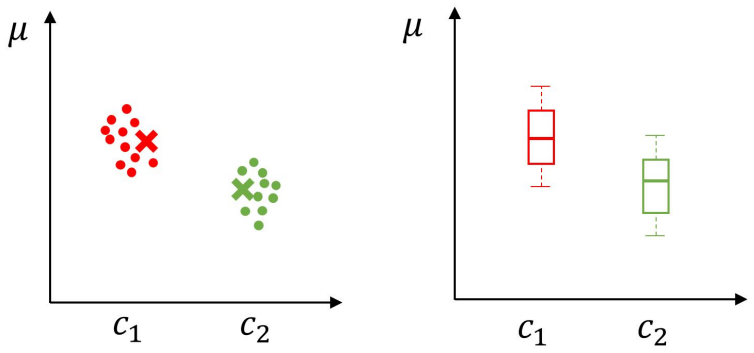


Figure 1.15: Distribution of the features for two different conditions a) dispersion of the points with x showing the mean; b) box plot related to the distribution; the mean values for the clusters differ, but their distributions overlap in the feature space, investigating single points from each class it is possible to find such that their values differ, but it does not ensure accurate classification

In order to ensure the feature's robustness, it is essential to evaluate it with a variety of input signals. Relying on a single observation can often lead to misleading conclusions. This issue is demonstrated in Figure 1.15 where the mean values of the features for two classes are different, but the dispersion makes those features overlap, preventing accurate classification. By selecting a single point from two classes, it is possible to find such that they have different values of the feature, but it does not guarantee accurate classification for all the points from two classes.

The variability of the features is rarely investigated for compressor instabilities detection, despite its great importance in practical applications. One of the studies acknowledging the stochastic nature of the pressure signal was done by Dhingra et al. [87]. It was demonstrated that the detection of a singular change in signal characteristics does not have to indicate the onset of instability, as such events may happen randomly during normal operation. Therefore, it was recommended to investigate a number of consecutive periods for the purpose of detection. Xue et al. [162] proposed an indicator based on the entropy of a pressure signal and demonstrated that such a feature can be used for accurate detection. However, the aspect of dispersion was not investigated in detail. The dispersion of the data was also considered by Kerres [89], who investigated Hurst exponent as a surge indicator. It was indicated that good accuracy can be obtained if the input signal length is of several surge cycles. The majority of research studies neglect the stochastic aspect of indicators, simply showing that the characteristic derived from a chosen signal input allows for distinguishability. It is possible that conclusions similar to that drawn by Kerres [89] would be obtained for other methods such as [47, 65, 94], rendering them unsuited for quick instabilities detection.

A clear distinction between the classes based on a single feature may not be feasible in some cases. Then, it may be possible to increase the number of features to separate the conditions. With a multi-dimensional feature space,

the boundary between conditions can have different shapes. Considering the 2-dimensional feature space for ease of visualisation, the separation may be possible either with a linear boundary or a non-linear one (Figure 1.16). The shape of a non-linear boundary can be provided in many ways.

Constructing optimal boundaries between conditions can be done with classifiers. A classifier is an algorithm used to assign labels or categories to newly observed input data based on previously known features and their patterns [167]. Classifiers can be supervised or unsupervised [168]. A supervised classification algorithm is trained on a portion of data with defined classes and aims to find the best separation between the classes, which is then used for new inputs. The research regarding classifiers and classification approaches is extensive in many fields [167]. Many engineering challenges to be solved with a data-driven approach can be expressed as classification problems [168] and the same applies to instabilities detection problem in centrifugal compressors.

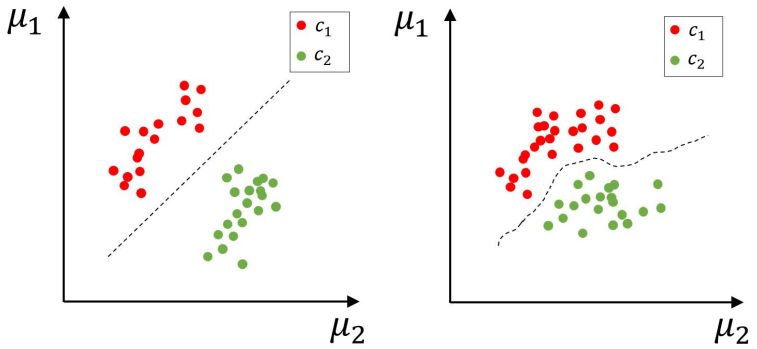


Figure 1.16: Conceptual representation of boundary types for 2-dimensional feature space and binary classification a) linear boundary b) non-linear boundary

The use of classifiers is not common for instabilities detection in centrifugal compressors. In mosts, the classification aspect is not mentioned explicitly. Researchers focus on single instabilities and demonstrate a feature that can provide a difference between conditions [47, 58]. Hence, a simple threshold is

provided, commonly based on the maximum of reference conditions to demonstrate a change occurring for unstable conditions [78, 94]. The feature used for classification often has a physical interpretation, so the threshold imposed on such feature can also be interpreted.

There exist a number of classification algorithms that were applied in the field of engineering [169]. The two commonly used types of classifiers are k-nearest neighbours (k-NN) and its variants [170], as well as support vector machine (SVM) [171]. However, recent developments evidence that classifiers with probabilistic output can provide similarly good performance while providing higher interpretability of the classification [172].

Apart from interpretable classifiers, there also exist a large group of deep learning methods that could possibly be used for instabilities detection [173]. Those approaches are often regarded as 'black box' models, because their inner processing mechanisms are unexplainable. However, such models, if trained with a sufficient amount of input data, can often provide a high-accuracy classification for complex problems [174]. Their advantage is that they can operate directly on raw data, eliminating the need for feature engineering, which is often time-consuming and requires expert knowledge [119]. However, deep learning models often require a large amount of training data to perform well. Above all, such models lack interpretability, which is very important in safety-critical applications such as centrifugal compressor instabilities detection. Hence, deep learning models are not considered a viable solution and only interpretable classification approaches are investigated.

The first, commonly used interpretable classifier is k-NN. It is based on the similarity of the location of points in the feature space [168]. K-NN classifies the new input based on the classes of a number of the nearest points in the feature space. Both the number of points and distance type can vary, depending on the application [168]. Commonly, Euclidean distance is used, however, it is possible to take advantage of Manhattan distance, Mahalanobis distance or other types

of distances and the choice of distance type can influence the classification results [175]. The similarity measure is also at the root of k-means classifier, where the distance from the distribution is used rather than from a selected number of points [176]. The classification with k-NN and similar methods provides non-linear boundaries between the classes, but it can render inaccurate results if an inappropriate type of distance is selected [175]. It also may be sensitive to the shape of clusters and cause problems if the clusters are elongated [177].

A more advanced classifier in terms of the classification algorithm is SVM. SVM aims to find the optimal hyperplane that separates the data points of different classes with the largest possible margin [168]. The hyperplane should maximize the distance between the nearest data points of each class, which are called support vectors. SVM can provide good accuracy classification in many engineering problems and has been used extensively for classification in different domains [171], however it is not interpretable [169].

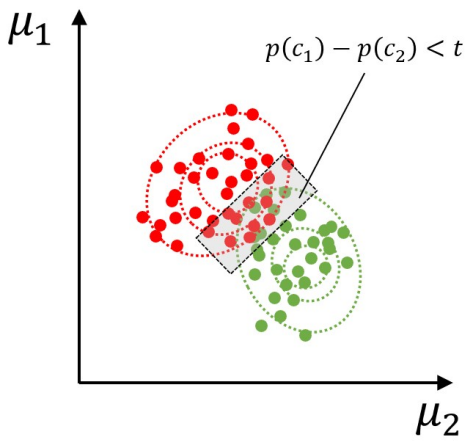


Figure 1.17: Depiction of a rejection zone concept for two classes classification; the rejection zone is present where the difference in class c_1 probability $p(c_1)$ and class c_2 probability $p(c_2)$ is lower than a certain threshold value t

Both K-NN and SVM provide the classification outcome but do not allow obtaining information about the uncertainty of the assigned class. Such output can be obtained with a probabilistic classifier and it can be useful when the

interpretability of the classification is important. What is more, using the estimates of probability, a reject option can be included in the classification [178]. It might be possible to define a region where the classification is not performed to avoid errors. It is demonstrated in Figure 1.17, where the grey rectangle represents a rejection zone. It is especially useful when the misclassification is expensive, which is the case for centrifugal compressors as the activation of an anti-surge system results in loss of performance or machine stoppage [80].

A powerful approach that offers non-linear, data-driven classification with probabilistic output is Gaussian process classification (GPC). It was not widely employed, however, it has shown to be suitable for a number of engineering problems [179, 180, 181] and that it could offer high accuracy, often being among the best-performing classification methods [172]. GPC offers a probabilistic approach, where for a new data point, the model output the probability of it belonging to each class. This information is used for classification, where the most probable class is selected as a true class, but also enables the introduction of a rejection option for the locations where the classification is not certain [182]. This region could be based, for example, on the relation of probability for different classes. One of the approaches for a classifier with a probabilistic output is to define a rejection zone based on the difference between class probabilities [178]. Then, if the difference is smaller than some threshold, the indication is not certain and should be rejected.

An important advantage of GPC is that it can provide high accuracy while using a limited number of data points [183], which is beneficial in the context of centrifugal compressor monitoring due to challenges in obtaining data for unstable conditions. GPC is a performant classification method that could be used to aid instabilities detection in centrifugal compressors.

The application of a classifier can help to increase the accuracy of a system but also brings a number of requirements for its implementation. Using a classifier, all the classes have to be represented, which may be challenging to

obtain experimentally for some machines [18]. The quality of classification is dependent on the selection of the classifier and the quality of the input features, with the latter shown to be of greater importance [119]. Hence, despite the selected classification and decision-making approach, the focus of the detection process should be on obtaining the most informative features.

1.7 Challenges in data-driven instabilities detection

Data-driven detection of instabilities is a challenge due to a highly non-linear and non-stationary character of the signals obtained from compressor monitoring. The difficulty is increased by a dynamic onset of instabilities, resulting in the need for real-time operation, especially when detecting surge. The ideal instabilities detection method should be:

- capable of providing real-time indication,
- robust and immune to false alarms,
- sensitive to all types of instabilities,
- capable of differentiating instabilities.

Those requirements can possibly be met by data-driven detection methods, however, identifying the best approach is not an obvious task. There is a number of performant methods developed over the years. However, there are no benchmark data sets in the field of instabilities detection that could be used to compare them. This is a major difference compared to condition monitoring, where such sets are available and commonly used and made available [184]. The cost of obtaining the data set for the compressor is much higher than that for bearings or gears, therefore the experimental campaigns are commonly performed with industrial partners and the data is often not shared. Using the same

data set for comparison of data-driven methods can provide valuable insights into the differences in their performance for instabilities detection.

Based on the overview of instabilities detection methods presented in Section 1.5, a number of conclusions can be drawn and shortcomings of the current studies can be identified. Many of the investigations in the field of instabilities detection focused on demonstrating that indication of instabilities can be obtained using a specific signal processing method, leveraging a change in frequency spectrum [65]. A small number of papers define features based on the proposed methods [162]. Such features could be used as an input for a decision-making unit. Defining the features is crucial because it allows other researchers to compare and evaluate them for other machines and data sets without the need for expert input in decision-making.

A clear definition of the features also opens the possibility of a structured investigation on the dispersion of the potential indicators. Little attention is paid to the stochastic aspects of the signals from a compressor and its effect on detection accuracy. The demonstration of method utility is often performed on a single instance of a signal [47, 58, 94]. As much as it can prove the potential of the method, it does not allow for the understanding of its robustness.

The lack of definition of indicators also results in no discussion on the pace of detection, which is crucial, especially in case of surge [94]. Some considerations about the speed of reaction are present in the literature [78, 185], however not in the stochastic context of the signal. The pace of detection in terms of a data-driven system can have two components. The first is related to the length of the signal that needs to be collected for a robust indication. The other is related to the processing time of a method that is applied to extract features.

The longer input signal allows for increasing the amount of information present in the signal. It may be expected that the features obtained based on a longer signal will be less dispersed. Therefore, from this perspective, the signal acquisition time should be the longest possible. However, the need for

a long input signal results in an extension of the acquisition time as well as increases the processing time. Both of those aspects negatively affect the pace of detection offered by a system. Hence, the optimum should be found, regarding the robustness and pace of detection. The optimum can differ depending on the signal processing method.

Compressors operate in a continuous manner and so should an instabilities detection system. Data-driven decomposition can operate on the oncoming batches of data, providing independent information based on each batch. The system needs to function in real-time, which requires the processing time to be equal to or shorter than the time it takes to collect another batch of data. Only then, all the incoming data is analysed on time and there is no backlog. The online operation can be considered a condition for the evaluation of an instabilities detection system.

The number of works employing data-driven decomposition methods for instabilities detection is very low compared to the field of condition monitoring. The body of literature in the latter domain has shown a great potential of decomposition methods in uncovering the dynamics of the system and defining indicators of fault or damage [164, 186]. Both domains have a similar system operation idea but differ in signal type, detected phenomena, and detection pace requirements. Methods like EMD or SSA can derive multiple features from a single measurement. This makes them potentially suitable for detecting various instabilities in a compressor system using a limited number of sensors.

A centrifugal compressor system can develop a number of instabilities in a different order, depending on the rotational speed [43, 8]. Despite a well-documented need for comprehensive instabilities detection, a general framework covering the detection of different instabilities with a single approach has not been found in the literature. Such a framework, preferentially taking advantage of a single method for the detection of all instabilities should be beneficial as it might decrease the computational time and effort compared to using a number

of methods. It could be possible to build such a framework by taking advantage of data-driven decomposition methods.

1.8 Aim, objectives and contributions to knowledge

Based on the identified challenges and gaps in the literature, the aim of the thesis is formulated as:

Development of a data-driven condition monitoring methodology for instabilities detection in centrifugal compressors based on decomposition techniques applied to pressure signals.

To fulfil the aim of the thesis, a number of objectives were defined:

(O1): Identifying and analysing methods used for instabilities detection in centrifugal compressors to understand the characteristic features leveraged for detection (Chapter 1);

(O2): Developing a methodology employing decomposition techniques to obtain instabilities-sensitive features based on an input pressure signal (Chapter 2);

(O3): Evaluating feasibility of general compressor stability detection using features obtained with data-driven decomposition methods (Chapters 4 and 5);

(O4): Performing a parametric study on crucial parameters of selected decomposition methods in the context of robustness and pace of detection (Chapters 4 and 5);

(O5): Comparing the performance of EMD and SSA in the context of extracting instabilities-sensitive features using threshold-based approach (Chapter 6);

(O6): Developing a probabilistic classification model for classification in order to increase the accuracy of classification and address the misclassifications at the transition between condition observed for threshold-based approach with rejection zone (Chapter 7).

Fulfilling the aim and objectives of the thesis led to producing a number of contributions to knowledge, listed below:

(C1): A comprehensive literature review of instabilities in centrifugal compressors and instabilities detection techniques, with emphasis on the signature of instabilities and how it is targeted by different detection methods (Chapter 1);

(C2): Developing a methodology for accurate instabilities detection from pressure signal and defining case study to be used for its validation (Chapters 2 and 3);

(C3): Demonstrating that data-driven decomposition techniques can be used effectively for comprehensive detection of instabilities in centrifugal compressors (Chapters 4 and 5);

(C4): Evaluating EMD and SSA for instabilities detection in centrifugal compressors; identifying their advantages and limitations as well as demonstrating the challenges to be addressed when using these methods for instabilities detection with pressure signal (Chapter 6);

(C5): Creating a probabilistic framework for instabilities detection based on Gaussian process classification and discussing its advantages as well as

limitations in application to instabilities detection in centrifugal compressors (Chapter 7);

(C6): Demonstrating that using data-driven decomposition techniques coupled with a probabilistic classification algorithm, it is possible to accurately detect instabilities using data from a single pressure sensor located before the impeller of the compressor (Chapter 7).

1.9 Thesis outline

The thesis is organised as follows. In Chapter 1, general information about compressors and their stability was introduced. Next, the physics of instabilities was discussed followed by a review of methods applied for instabilities detection. It was demonstrated that the field of data-driven decomposition methods application for instabilities detection remains largely unexplored. The selection of two different instabilities methods with high proven potential in other fields was performed in light of challenges that have to be faced by instabilities detection methods.

Chapter 2 opens with the outlook of the proposed methodology for instabilities detection. Next, the elements forming that methodology are discussed. First, the theory of empirical mode decomposition and singular spectrum analysis is introduced. Then, selected methods of component processing are demonstrated and discussed. Next, the probabilistic framework is defined along with its objectives. Finally, the criteria for extended comparison of the methods and validation of the framework are introduced.

Chapters 4 and 5 are devoted to EMD and SSA, including a parametric study on the most important intrinsic parameters of the methods as well as input signal length that affects the method responsiveness. The case study is performed on the machine described in Chapter 3. Firstly, the detection of global conditions is demonstrated, followed by an investigation of decoupling

two regimes of global instability. It is demonstrated that general conditions can be well detected with both methods, however the global instability conditions are not decoupled by any of the approaches. However, it is possible to define features sensitive to the severity of global instability. Then, the differentiation can be made based on the feature value.

Chapter 6 demonstrates the comparison of the features obtained with EMD and SSA. The detection accuracy is computed for both sets of features and the outcomes are investigated in detail. This chapter also presents a number of practical considerations, regarding the selection of the components, pace of detection, representativeness of data obtained with different protocols and a possibility of limiting the number of sensors used for detection.

In Chapter 7, a probabilistic framework is proposed based on Gaussian process classification (GPC). The output of GPC offers not only classification but also the certainty of belonging to a class. Hence, it is possible to define rejection regions based on the level of probability. With a proper selection of the kernel function, the rejection zone can help to reduce the number of misclassifications and increase the accuracy of a method. GPC can provide accurate classification using on a small amount of data for training. It can also be leveraged to use only one sensor for detecting the stability of the system.

Chapter 8 presents the conclusion and further work considerations. It collates all the important observations and reiterates the main findings presented in this thesis. The limitations of the developed approach are also demonstrated and a path towards improvement of the proposed instabilities detection framework is proposed.

2 Methodology for data-driven instabilities detection

This chapter introduces the methodology for instabilities detection in centrifugal compressors based on data-driven decomposition methods. First, the overall framework for obtaining instability indication based on the pressure signal is introduced. It has two main steps, being feature extraction and decision-making. For the feature extraction, empirical mode decomposition (EMD) and singular spectrum analysis (SSA) are investigated. The features obtained with those methods are introduced into the decision-making module, where two supervised classification approaches are investigated. The first assumption is that only the data from stable conditions are used to define linear thresholds between conditions. The other is a probabilistic classification approach, where all the classes have to be represented in training and the boundaries are more flexible. The comparison of the methods allows investigating and quantifying the advantages of using a probabilistic approach for data-driven instabilities detection.

2.1 Methodology overview

The methodology developed in this thesis follows the general steps of the approach commonly used in the broad domain of condition monitoring [119]. The general flow of the methodology is shown in Figure 2.1, while it substeps are described in detail in Figures 2.2 and 2.3. The input signal collected from the compressor is processed in order to extract features indicative of system conditions. Using those features, a decision about the compressor system stability can be made. The novelty of the work lies in the methods used in both feature extraction and decision-making steps.

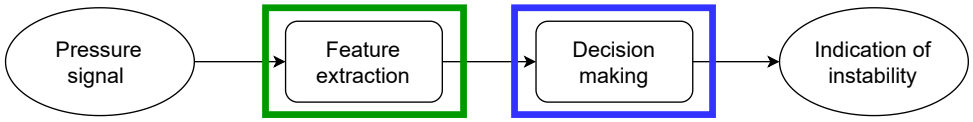


Figure 2.1: Overview of the general methodology steps; steps marked with colors are furtherly explained in detail

The feature extraction approach proposed in the methodology is based on data-driven decomposition methods and includes a number of steps. Its details are demonstrated in Figure 2.2. The very first step is signal pre-processing such as mean removal or normalisation, and it is optional. The need for pre-processing may depend on the type of decomposition method used and in some methods it might not be necessary. The next step is a signal decomposition. The pressure signal is decomposed into a number of components that are to capture and highlight different characteristics of the signal that could be indicative of instabilities. The decomposition can be any method that leads to obtaining the components that are time series. In this thesis, empirical mode decomposition (EMD) and singular spectrum analysis (SSA) are proposed. The decomposition components are processed to extract the features. In this study, the features are obtained as the RMS of the components. The RMS is related to the energy

of the signal [187] and was shown to be useful for building sensitive features in similar cases [47, 146].

The features may be based on individual components or groups of components that best highlight the presence of instability. The selection of the components to be processed into features is an important issue. In this thesis, the components that best capture the difference between stable and specific unstable conditions are used. The output from feature extraction process is a vector μ containing a number of instability-sensitive features. The dimensionality of the vector depends on the number of components selected from the decomposition.

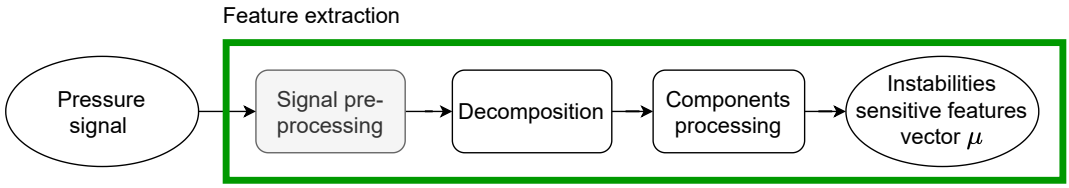


Figure 2.2: Details of the feature extraction step

The formulation of the decision-making step may depend on the input data type, available classes and the aim of the detection. The methodology, with the details of the decision-making step is shown in Figure 2.3. The decision-making is based on classification of the data for unknown operating conditions. A supervised classification is employed, which has two phases. The first one is the training phase, where the available data for the known conditions are processed through feature extraction step. When using more than a single feature, it is possible to take advantage of the feature space representation, where the values stored in the feature vector μ are considered coordinates of a point in space. The features are processed in a joint manner through a feature space representation concept, which allows greater insight into the operating conditions of the system and allows applying a number of classification techniques to the data.

The decision-making requires a priori training of the model, as indicated with a dashed line box in Figure 2.3. The pressure signal for known conditions is processed through feature extraction and the data is scattered onto the feature space. Based on the distribution of the data, boundaries for different conditions can be created. The boundaries can then be placed onto the feature space, mapping the regions of different conditions. When the training phase is done, it is possible to use the model for instabilities detection. The signal sampled from unknown operating conditions is subjected to feature extraction process to obtain features. Those features allow to position the new conditions in reference to the boundaries in the feature space to make an indication of instability. The advantage of the proposed method is that the instability indication can be obtained using only a single feature vector from unknown operating conditions.

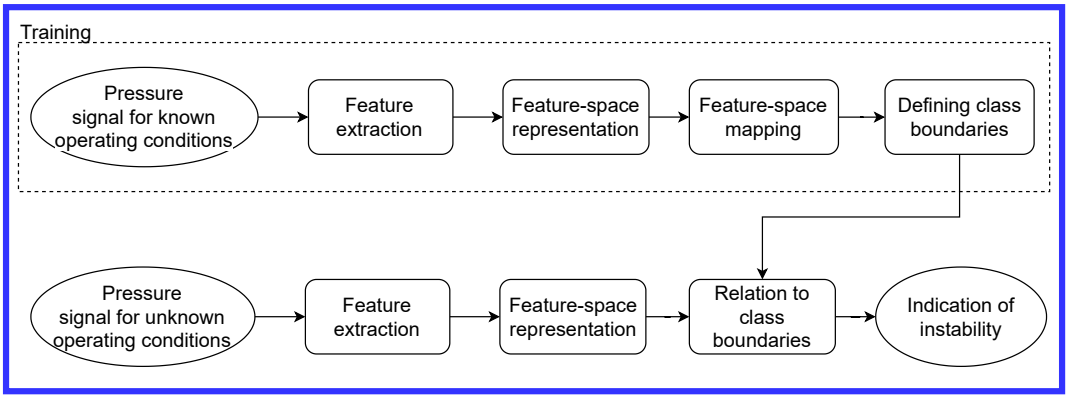


Figure 2.3: Details of the decision-making step

Two different approaches to feature space mapping and definition of class boundaries are proposed in this thesis. The first method is a threshold-based approach. It is assumed that the features obtained within the feature extraction procedure are interpretable and represent the intensity of instability. The intensity of instability is understood as the strength of the instability presence in the compressor. Hence, the values of features representing instabilities should be low in stable region and increase when the instability onsets. Therefore, by computing the extent of the feature values for the stable range, it is possible

to establish a threshold value. If such a value is exceeded, the instability is detected. The advantage is that the training phase can be based solely on stable operating conditions. However, with this approach, the number of used features should correspond to the number of instabilities to be detected. Each feature should independently react to each instability in order to ensure differentiation of the conditions.

The other approach is based on a probabilistic classifier. The interpretability of the features is not as crucial in this scenario, as the classification boundaries are determined by the data distribution in the feature space. Such approach requires representation of all conditions, not only stable ones. The output is the class boundaries but also the probability of each class, which may be used for obtaining better insights about the classification.

Two distinct decomposition methods are investigated for processing the input signal into features - empirical mode decomposition (EMD) and singular spectrum analysis (SSA). Both were demonstrated to have potential in this application [47, 78], however they have not been thoroughly investigated nor compared. The performance of the methods is dependent on a number of factors, which are discussed in Section 2.3. The features obtained through EMD and SSA are investigated for an interpretable detection based on threshold values for different instabilities. This study also focuses on a potential improvement in detection that can be offered with an advanced non-linear classification method based on Gaussian process. The outcome is a proposition of the best approach to be used for instabilities detection in centrifugal compressors.

2.2 Input signal

The first step in the methodology is obtaining an appropriate input signal. This methodology assumes the use of a high-frequency pressure signal sampled inside the compressor. The signal from several locations inside the machine is used, which is described in detail in Chapter 3.

The signal acquired during measurements $x(t) = (x_1, x_2, ..., x_{N_s})$ is a discrete signal of length N_s. The length N_s is related to the physical acquisition time through sampling frequency f_s. For a fixed f_s, the larger the N_s, the longer the acquisition time.

The length of the signal can affect the processing outcomes. The N_s should be such that the obtained features are representative for the compressor system and their dispersion is not too large. On the other hand, the N_s should not be too large in order for the system to be responsive and provide the quickest possible information about the onset of instability. The influence of the input signal length N_s on the feature extraction procedure will be demonstrated in detail in Chapters 4 and 5.

2.3 Signal decomposition

The second step of the methodology is the decomposition of the input signal $x(t)$. The aim is to extract and enhance the characteristics of aerodynamic instabilities contained in the input signal. The decomposition methods investigated in this study, EMD and SSA, were selected based on their performance in other domains. They represent two different families of approaches, hence their comparison in the context of instabilities detection is of interest. An extended discussion on the decomposition methods and selection made in this study is provided in Section 1.5.6.

2.3.1 Empirical mode decomposition (EMD)

EMD is a decomposition method based on the assumption that any signal consists of different simple modes of oscillations, called intrinsic mode functions (IMFs), which can be separated in the decomposition process [90]. The procedure of extracting IMFs is called sifting. Sifting algorithm is based on the waveform of the signal and employs a process where the components are extracted in an iterative manner, starting from IMF 1. IMFs can evidence changes in both

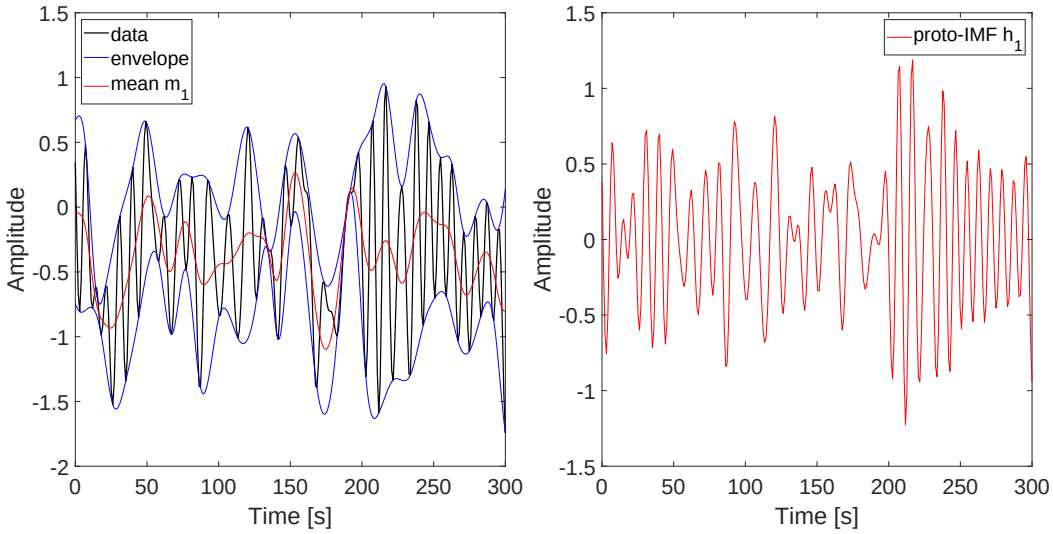
amplitude and frequency of a physical phenomenon expressed in the analysed signal [90]. For a function to be termed an IMF, it must meet two conditions [141]:

- the number of extrema and zero crossings of an IMF must be identical or differ by one;
- the mean value of an IMF envelope defined by local maxima and minima must be equal to zero.

These assumptions are, in practical implementations, replaced with stoppage criteria that govern the decomposition process [141]. It is due to the fact that a high number of sifting iterations introduces an error to the decomposition and depletes the IMFs of physical meaning [188]. A small number of iterations is beneficial in terms of decomposition time, as the more sifting iterations are needed, the longer the decomposition process. The practical effects of changing the parameters of the decomposition process are discussed in Chapter 4.

Using EMD, the input signal $x(t) = (x_1, x_2, ..., x_{N_s})$ is decomposed into a set of IMFs. Each IMF is a time series of the same length as the input signal $x(t)$. The first step in EMD algorithm is to identify local maxima and connect them, creating the upper envelope. Then, the procedure is repeated for local minima to obtain the lower envelope [90]. Then, the envelope mean can be calculated, which is demonstrated in Figure 2.4a. It is possible to use a different type of interpolation between extrema. The most commonly applied is the approach using cubic spline interpolation, which is also used in this thesis. The upper and lower envelopes should contain all the data points of the signal. The first component of the process, $h_1(t)$ is called a proto-IMF and is shown for an exemplary signal in Figure 2.4b. It is obtained from Equation (2.1), where $m_1(t)$ is the signal envelope mean, understood as an arithmetic mean of upper and lower envelopes values for each time instant t.

$$h_1(t) = x(t) - m_1(t) \tag{2.1}$$



(a) (b)

Figure 2.4: EMD process visualisation based on an exemplary signal a) The data along with upper and lower envelopes defined by the local maxima and minima, respectively, and the mean value of theenvelopes b) proto-IMF h_1

Then, it is checked if proto-IMF $h_1(t)$ meets the conditions of an IMF or a stoppage criterion is fulfilled. If neither is true, the next process step takes place. The $h_1(t)$ is treated as input data and the sifting procedure is repeated. The $h_2(t)$ is created by subtracting the mean of $h_1(t)$ envelope from $h_1(t)$ itself, as shown in Equation (2.2). Then, the check against conditions is repeated for $h_2(t)$. Subsequent steps are performed according to Equation (2.3).

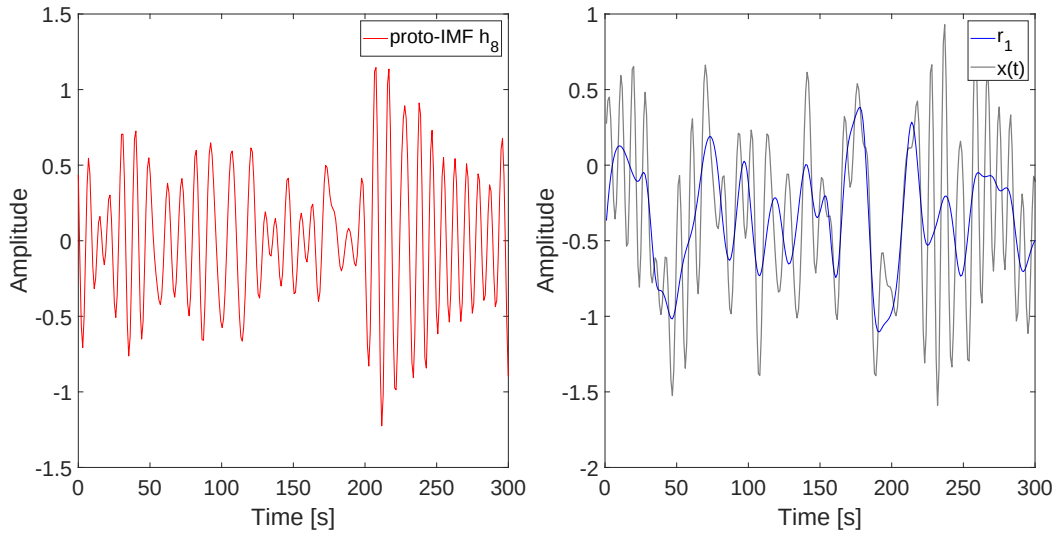
$$h_2(t) = h_1(t) - m_2(t) \tag{2.2}$$

$$h_s(t) = h_{s-1}(t) - m_s(t) \tag{2.3}$$

The process is repeated until the proto-IMF h_s meets IMF criteria or the stoppage criterion is achieved. If sifting process is completed after S iterations, $h_S(t)$ is designated $i_1(t)$ and becomes the first intrinsic mode function (IMF). For

the exemplary signal, the proto-IMF was meeting the conditions after 8 sifting iterations (Figure 2.5a). To continue the sifting process, $i_1(t)$ is extracted from the original data $x(t)$ and the output of this operation is termed intermediate residue $r_1(t)$ - Equation (2.4). The residue and the original signal are shown in Figure 2.5b.

$$r_1(t) = x(t) - i_1(t) \tag{2.4}$$



(a) (b)

Figure 2.5: EMD process visualisation based on an exemplary signal a) IMF 1 obtained after 8 iterations b) original signal $x(t)$ and the first residual r_1 after subtracting IMF 1 from the original signal

Next, $r_1(t)$ is taken as the new original data and the sifting process is repeated. The IMFs are extracted until the final residue is either a constant or monotonic function or some predefined decomposition stoppage criterion is met. The signal after decomposition is divided into Q IMFs containing oscillatory modes of the signal $x(t)$ and a residue $r_Q(t)$, holding information about the trend in the data (Equation (2.5)). If the decomposition is stopped due to achieving a global decomposition stoppage criterion, the character of the residual may be different.

$$x(t) = \sum_{i=1}^{Q} i_i(t) + r_Q(t) \tag{2.5}$$

Two stoppage criteria must be defined for EMD process, sifting stoppage and global decomposition stoppage. Globally, the decomposition can be stopped if the energy contained in the residue is lower than a certain level, defined in reference to the input data [90]. The choice of this value influences the number of IMFs obtained from the decomposition and the character of the residue, as with high energy threshold the residue can still contain oscillatory modes after the stoppage. The setting of the threshold value for the decomposition stoppage has no influence on the character of the IMFs extracted in the process, only on their total number.

The sifting stoppage criterion has influence on both, the number of IMFs and their character, mostly frequency content [141]. For noisy signals such as experimental pressure signals from compressor, each IMF covers a frequency band located around a specific central frequency [189]. Within a range of sifting iterations, EMD can work as a dyadic filter, resulting in the central frequency of each subsequent IMF being half of the previous one [141]. Increasing the number of iterations, the central frequencies of IMFs are located closer to one another, consequently, the frequency band covered by each IMF becomes narrower and the number of extracted IMFs increases [139, 141]. The number of sifting iterations SN can be controlled by a choice of a sifting stoppage criterion.

It was shown that if the number of sifting iterations is high, then the IMFs amplitude is constant and they no longer have a physical meaning [188]. Consequently, different approaches were proposed for the sifting stoppage, based on a change of energy between subsequent iterations, a relation of peak count to zero crossings or a number of sifting iterations [188]. It is recommended to minimize the number of sifting iterations in order to reduce errors caused by spline fitting to the extrema, introduced with every iteration. The optimum

number of eight sifting iterations per IMF was proposed [188]. Use of a low number of sifting iterations is beneficial for a compressor instabilities detection system responsiveness as it decreases the decomposition time.

There exist a number of stoppage criteria that can be used [190]. A stoppage criterion chosen in this study was a Cauchy type criterion, introduced originally by Huang [90] and applied in other investigations [191, 192]. When considering a discrete signal of finite length N_s sampled at the time instants $t = 1, 2, ..., N_s$, the criterion can be expressed as in Equation (2.6). The sum of squared differences between the previously calculated $h_{s-1}(t)$ and newly obtained $h_s(t)$ proto-IMFs is computed for the entire signal length and its sum is divided by the total energy of the newly obtained proto-IMF. Describing a relative change of proto-IMF energy between each sifting operation, SD_s has to be above a specified value for the process to continue. If the value is smaller than a set threshold, the sifting process is stopped.

$$SD_s = \frac{\sum_{t=1}^{N_s} |h_{s-1}(t) - h_s(t)|^2}{\sum_{t=1}^{N_s} h_{s-1}^2(t)} \tag{2.6}$$

2.3.2 Singular spectrum analysis (SSA)

SSA is a non-parametric time series analysis method, which shares the mathematical formulation with Principal Component Analysis. SSA procedure consists of two main parts, which are decomposition and reconstruction. Each of those steps is composed of certain sub-steps. In this study, SSA is applied to a time series $x(t) = (x_1, x_2, .., x_{N_s})$ of length N_s.

Decomposition

The very first step of decomposition is called embedding. A set of K lagged vectors $\boldsymbol{x_i}$ of length L is constructed, as per Equation (2.7), where $K = N_s - L + 1$. L is a window length parameter that has to be set. This parameter has

an important influence on the decomposition outcomes, governing the number of RCs and their content [91].

$$\mathbf{x_i} = (x_i, x_{i+1}, ..., x_{i+L-1})^T \tag{2.7}$$

The vectors $\boldsymbol{x_i}$ are assembled into a trajectory matrix $\mathbf{X}$ as per Equation (2.8). A trajectory matrix $\mathbf{X}$ is a Hankel matrix of dimension $[L \times K]$. The embedding step is necessary to increase the dimensionality of the input data.

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_{N_s} \end{bmatrix} \tag{2.8}$$

The next step is the singular value decomposition of the trajectory matrix $\mathbf{X}$, as per Equation (2.9).

$$\mathbf{X} = \sum_{i=1}^{L} \mathbf{X}_i \tag{2.9}$$

It is obtained with the eigenvalue decomposition of the squared matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ of dimension $[L \times L]$. This decomposition provides a set of eigenvalues in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$) and their corresponding eigenvectors ($\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_L$). The decomposition of $\mathbf{S}$ leads to obtaining L components in form of elementary matrices. The sum of those matrices makes the original trajectory matrix $\mathbf{X}$. Each individual component $\mathbf{X_i}$ is defined by an eigentriple (Equation (2.10)).

$$\mathbf{X}_i = \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^T \tag{2.10}$$

where:

$$\mathbf{v}_i = \frac{\mathbf{X}^T\mathbf{u}_i}{\sqrt{\lambda_i}} \tag{2.11}$$

Reconstruction

Having decomposed the time series into a number of matrix components, it has to be reconstructed into a time series. Each individual component matrix $\mathbf{X}_i$ contains particular information of the original trajectory matrix $\mathbf{X}$ and hence, each one contributes towards the reconstruction of $\mathbf{X}$ in a different degree. As the eigenvalues λ_i are in decreasing order, the first individual components contribute more than the last ones. At this step, it is possible to group the components by summation of selected individual component matrices. The approach to grouping may differ depending on the task at hand and structure of the time series [91]. It is also possible to reconstruct the components individually and group them after the decomposition, if desired. In this thesis, it is assumed that the components can independently hold the information that is of interest for the instabilities detection system. Grouping components prior to their analysis might lead to loosing some important information about the conditions of the compressor. Therefore, a grouping step is omitted, and the components are reconstructed independently.

In order to reconstruct each individual component in form of time series, it is necessary to apply diagonal averaging to the individual component matrices. Let $z_{m,n-m+1}$ be an element of an individual component matrix $\mathbf{X}_i$. The reconstructed components (RC), $RC_i = (r_{i_1}, r_{i_2}, ..., r_{i_{N_s}})$ corresponding to this individual component matrix $\mathbf{X}_i$ are calculated as per Equation (2.12).

$$r_{i_n} = \begin{cases} \frac{1}{n} \sum_{m=1}^{n} z_{m,n-m+1} & 1 \leq n \leq L \\ \frac{1}{L} \sum_{m=1}^{L} z_{m,n-m+1} & L \leq n < K \\ \frac{1}{N_s - n + 1} \sum_{m=n-K+1}^{L} z_{m,n-m+1} & K \leq n < N_s \end{cases} \tag{2.12}$$

The input signal $x(t) = (x_1, x_2, \ldots, x_{N_s})$ is now decomposed into L independent RCs as shown in Equation (2.13). Each of those RCs is a time series and can be processed individually.

$$x(t) = \sum_{i=1}^{L} RC_i \tag{2.13}$$

2.4 Decomposition components processing

After decomposition, the initial time series can be represented as a sum of components for SSA or sum of components and residual for EMD. It is assumed that the information of interest is contained within specific components. The aim of the feature extraction process is to process selected components and obtain features sensitive to instabilities.

2.4.1 Obtaining features from components

The features can be built either on selected components or on groups of components, obtained through their summation. Assuming $y_k(t) = (y_1, y_2, ..., y_{N_s})$ to be a selected component of the decomposition or sum of components in a form of a time series, the feature μ_k is obtained through computing the root mean squared value of y_k, according to Equation (2.14).

$$\mu_k = \sqrt{\frac{1}{N_s} \sum_{t=1}^{N_s} |y_k(t)|^2} \tag{2.14}$$

The feature represents the energy of y_k. The idea pursued in this thesis is to define features that each of them is associated to the intensity of a selected instability, which can be done through energy of time series related to instabilities. For defining features, specific components must be selected.

2.4.2 Selection of the components

The choice of the components for building features of instabilities is a challenging task. The procedure for selection may differ, depending on the available data. The approach proposed in this thesis is based on the comparison of the feature values for stable and unstable conditions. It requires the data from unstable conditions, however owing to such approach it is possible to validate the performance of a feature as a potential indicator. The goal of the selection is to obtain a single feature for each investigated instability.

A perfect feature should ensure a maximum difference between its values for different conditions. It applies to the mean values as well as data dispersion. When investigating two classes of operating conditions, it is possible for the machine to transition smoothly from one condition to another. As a result, the feature values may overlap near the interface between the two conditions. The very first aspect of the selection is comparing the mean value, as the difference in the mean is crucial for differentiation of conditions. Consequently, if the means differ, the dispersion is evaluated.

2.5 Feature space representation

The features obtained through feature extraction for each input signal are assembled in a feature vector μ_i and can be represented in the feature space. Using the signal collected from the compressor, a number of feature vectors can be constructed. Their number is dependent on the total acquisition time and the selected input signal length N_s. Assuming a total of M input signals, a feature matrix $\boldsymbol{F}$ (Equation (2.15)) can be created, where each row is a feature representation for time instant covering the input signal acquisition time.

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \\ \vdots \\ \boldsymbol{\mu_M} \end{bmatrix} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,N} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{M,1} & \mu_{M,2} & \cdots & \mu_{M,N} \end{bmatrix} \tag{2.15}$$

$$\boldsymbol{c} = (c_1, c_2, ..., c_M)^T \tag{2.16}$$

For every feature vector μ_i, a label $c_i = \{1, 2, .., m\}$ is assigned, representing the class of conditions. Using feature matrix $\boldsymbol{F}$ and a vector of labels $\boldsymbol{c}$ (Equation (2.16)), it is possible to define a data set $\boldsymbol{D} = \{\boldsymbol{F}, \boldsymbol{c}\}$ which combines all of the available data for a compressor for known operating conditions.

To perform a supervised classification of the compressor operating conditions, a training data set must be defined. Such data set, marked $\boldsymbol{D_T} = \{\boldsymbol{F_T}, \boldsymbol{c_T}\}$ contains M_T feature vectors from the general data set $\boldsymbol{D}$. It is a common practice to also define a validation set, which is used to investigate the quality of classification by the trained model. Such data set is marked $\boldsymbol{D_V} = \{\boldsymbol{F_V}, \boldsymbol{c_V}\}$ and has a size M_V. The size of $\boldsymbol{D_T}$ can differ, as well as the classes included in $\boldsymbol{D_T}$.

The choice of the number of classes is problem-dependent and subjected to input data availability. In the most basic approach, the conditions can be divided into two classes representing stable and unstable compressor operation. In a more accurate approach, a class can be assigned to conditions when a specific instability is present. The number of classes should then be compatible with the number of instabilities detected in the system. It is also possible to assign a class to the data collected at each set point of the external resistance as then the labelling is more straightforward and less prone to interpretation errors. However, in such case the external resistance levels have to be furtherly linked with the compressor instabilities present for each set point.

2.6 Decision making

The decision-making block of the methodology has two stages. Firstly, the classification algorithm has to be trained to provide the location of the boundaries between classes that are used for classification. Secondly, given a new feature vector $\boldsymbol{\mu_*}$ for unknown operating conditions c_*, the classification can be made. The concept of feature space representation is used by the proposed classifiers.

Two approaches are investigated in this thesis. The first one is based on interpretable thresholds derived for each of the feature space dimensions. It is implemented for both EMD and SSA features to demonstrate the differences between the outcomes of two decompositions. The other is a probabilistic approach based on Gaussian process classification that can help to address some of the shortcomings of the threshold approach.

2.6.1 Threshold-based classification

A threshold-based approach assumes that the features are interpretable and each feature is representative for a single instability. Then, it is possible to define a single threshold for each feature, based on stable conditions. The relation of the feature vector $\boldsymbol{\mu_*}$ components to thresholds allows classification of the conditions.

Training

In the training phase for the threshold-based classification, the data set $\boldsymbol{D_T}$ contains only feature vectors representing stable operating conditions. The threshold is for a given feature μ_i, marked T_{μ_i} is obtained as a quantile of the data distribution, defined according to Equation (2.17). The p value can be set according to the requirements within the range $[0, 1]$. Using the quantile of a distribution rather than a maximum value, it is possible to limit the influence of the outliers that could be present in the data. A separate threshold is obtained for each dimension of the feature space.

$$T_{\mu_i} = Q_{\mu_i}(p) \tag{2.17}$$

Instabilities identification

Instability detection is performed by interrogating the new observations to the defined thresholds. A signal for a new observation from an unknown operating class conditions c_* is obtained and decomposed. The components $y_k(t)$ are processed using Equation (2.14). A feature vector $\boldsymbol{\mu_*} = (\mu_{*1}, \mu_{*2}, ..., \mu_{*N})$ is obtained and its components are interrogated against the thresholds T_{μ_i}. For two-dimensional feature space, two comparisons have to be performed. The first one helps to limit the class selection problem to two classes, while the second one defines the exact class of the two.

$$\mathbf{H}_1 : \mu_{*1} \leq T_{\mu_1} \rightarrow \text{Class } c_1 \text{ or Class } c_2 \tag{2.18a}$$

$$\mathbf{H}_2 : \mu_{*2} \leq T_{\mu_2} \rightarrow \text{Class } c_1 \tag{2.18b}$$

Based on this interrogation, a class of conditions can be assigned to the new observation. Hence, the decision about stability of the compressor is made based on a single feature vector obtained for unknown conditions.

2.6.2 Probabilistic classification

The probabilistic classification approach proposed in this thesis is based on Gaussian process classification (GPC). In this section, the rationale and requirements for its implementation are introduced, while mathematical formulation and the details of the method are investigated in Chapter 7.

Training phase

In the training phase for the probabilistic classification, the data set $\boldsymbol{D_T}$ contains feature vectors representing all operating conditions to be classified. The boundaries between classes are no longer straight lines and they are driven by the data and GPC model parameters. The training of the model relies on optimisation of the model parameters that increase the fit of the model to $\boldsymbol{D_T}$. The model with optimised parameters is considered trained and used for classification.

Instabilities identification

The GPC provides a probability estimate for each of the classes at each location in the feature space. Those estimates are data-driven and can be used for defining class boundaries. The classification is based on interrogating the model for the location defined by the feature vector for unknown conditions $\boldsymbol{\mu_*}$. The output of the model is the probability for each of the classes, assuming classification into three classes, $GPC(\boldsymbol{\mu_*}) = [p(c_1), p(c_2), p(c_3)]$. The highest class probability defined the classification for $\boldsymbol{\mu_*}$. The probabilistic output also enables introducing no-classification zones. Such locations are expected to represent the regions of low model certainty, where the probability of neither class is importantly higher than the others. Rejection zone can be used to refrain from making a classification, which may increase the accuracy of the model. The details of the probabilistic approach implementation are provided in Chapter 7.

3 Case study definition

3.1 Object of research

The methodology developed in this thesis was validated on a single-stage centrifugal compressor. The compressor is a low pressure unit of industrial type. It was designed by Magiera [193] and adjusted for the research on instabilities by Liskiewicz [9]. The cross-section of the machine is presented in Figure 3.1.

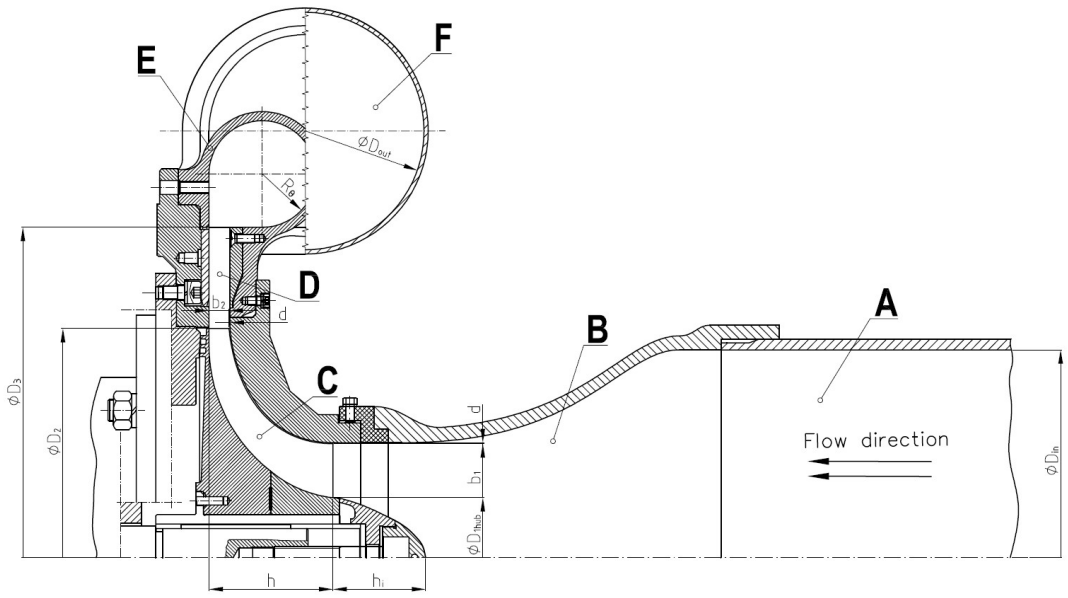


Figure 3.1: Compressor cross-section with marking of its most important parts [9]; A - inlet pipe, B - Witoszynski nozzle, C - impeller, D - diffuser, E - volute, F - outlet pipe

The compressor was designed to operate at ambient inlet conditions. The air was delivered to the compressor through the inlet pipe **A** of diameter D_{in} = 300 mm. Subsequently, it was accelerated by the Witoszynski nozzle **B** for

obtaining a more uniform inflow and directed onto the impeller **C**. The impeller was unshrouded and had 23 blades (Figure 3.2a). Downstream of the rotor, the air entered a vaneless diffuser **D** and subsequently the volute **E** of circular cross-section. The radius was gradually increasing streamwise from the volute tongue gap of 5 mm towards the outlet pipe of diameter D_{out} = 150 mm. The outlet pipe was made of two straight sections connected with the right-angle elbow. The section between the volute and the elbow was 250 mm long, while the section behind the elbow was 3750 mm long.

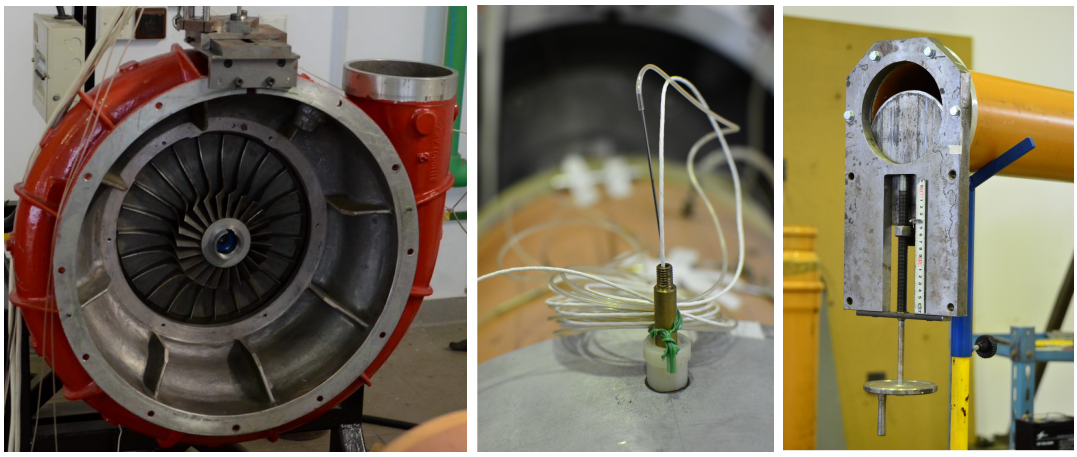
A throttling valve was mounted at the end of the outlet pipe (Figure 3.2c). The valve was operated manually and its position was controlled with aid of a measuring scale. The compressor was driven by an asynchronous AC motor (400V=15kVA) with an inverter. The design point was attained at f_{rot} = 120 Hz, $\dot{m}$ = 0:8 kg/s and PR = 1.12. However, in order to avoid a risk of the impeller damage at surge, the unit was run with a lower rotational speed of f_{rot} = 100 Hz, resulting in a nominal flow rate of $\dot{m}$ = 0.75 kg/s and pressure ratio PR = 1.08. For the selected rotational frequency, the impeller tip speed equal to u_2 = 103 m/s, well below the speed of sound a = 341 m/s. The dimensions and operating parameters of the test stand are aggregated in Table 3.1.

Table 3.1: Operating parameters and dimensions of the test rig

Parameter	Value
Pressure ratio	1.08
Rotational frequency	100 Hz
Mass flow	0.75 kg/s
Number of impeller blades	23
Inlet pipe diameter	300 mm
Rotor inlet diameter at hub	86.3 mm
Rotor inlet diameter at shroud	126 mm
Rotor inlet span	38.9 mm
Rotor outlet diameter	330 mm
Rotor outlet span	14.5 mm
Diffuser outlet diameter	476 mm

The operating conditions of the compressor were changed by altering a throttling valve position. By increasing the throttling level, the compressor mass flow

was decreased and compression ratio increased, driving the compressor towards a potentially unstable region. The throttling level was defined through throttle opening area (TOA). The TOA value is expressed in percent, obtained as a ratio of an open throat are available to the flow to the total area of the throat.



(a) (b) (c)

Figure 3.2: Elements of the test rig [9] a) experimental rig impeller and volute; b) pressure taping; c) throttling valve

Figure 3.3 demonstrates the performance curve of the compressor and mass flow dependence on TOA [9]. In this study, the data for mass flow rate were coming from an approximation of a velocity profile at the inlet to the compressor [9].

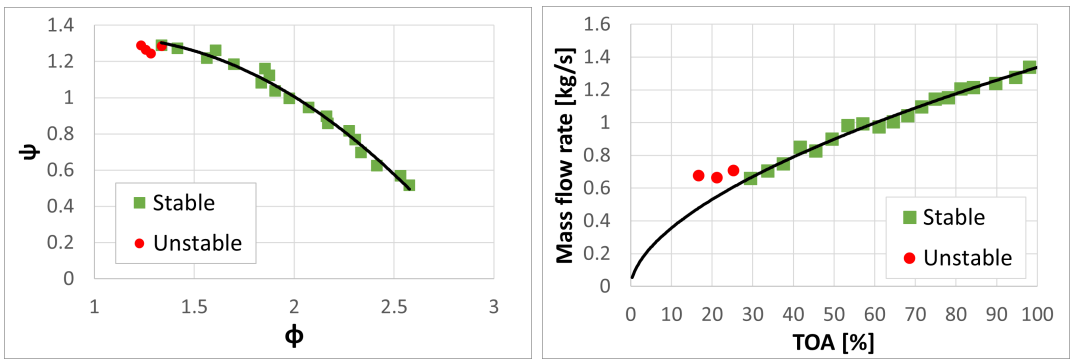


Figure 3.3: Performance measurements of the compressor [9] a) performance map (*phi* - mass flow coefficient, ψ - pressure raise coefficient); b) mass flow rate relation to throttling

With the onset of unstable conditions, the computation of the mass flow based on the velocity profile is no longer valid. This applies to the last three points, where the computed mass flow was almost constant and did not follow the trend established based on previous data points. To have a better understanding of the flow conditions, high frequency pressure data was collected.

3.2 Pressure data acqusition

Five dynamic subminiature Kulite transducers (XC0-080 SERIES) were mounted flush to the shroud walls to measure the static pressure in different locations along the flow path, as demonstrated in Figure 3.4. The sensor p_{s-in} was located at the inlet of the test stand. The sensors p_{s-imp1} was mounted before impeller, 0.2 L from the leading edge. Sensors p_{s-imp2} and p_{s-imp3} were mounted over the impeller, 0.4 L and 0.9 L from the leading edge of the impeller. The last sensor, marked p_{s-out}, was located at the outlet of the compressor, in the volute. The data was collected with sampling frequency of 100 kHz and each measurement lasted for 20 seconds, which was the limitation of the acquisition system memory buffer.

Two different collection protocols were used in this study, leading to obtaining two data sets: quasi-dynamic (QD) and dynamic (D). The aim for quasi-dynamic protocol was to obtain an accurate mapping of compressor operating conditions, collecting large amount of data while exactly controlling the level of throttling. In the procedure, a specified TOA value was set and 20 second of pressure signal was collected for this valve position. Subsequently, the TOA was changed by a small value and a new set of data was collected. This procedure was repeated for each TOA in the entire operating range of the machine, allowing to obtain the pressure signal independently for different operating conditions. The full control of throttling level could be maintained and it was possible to define the step in TOA change, which was approximately 0.5% of TOA. For each throttle position, over 2 million points were collected.

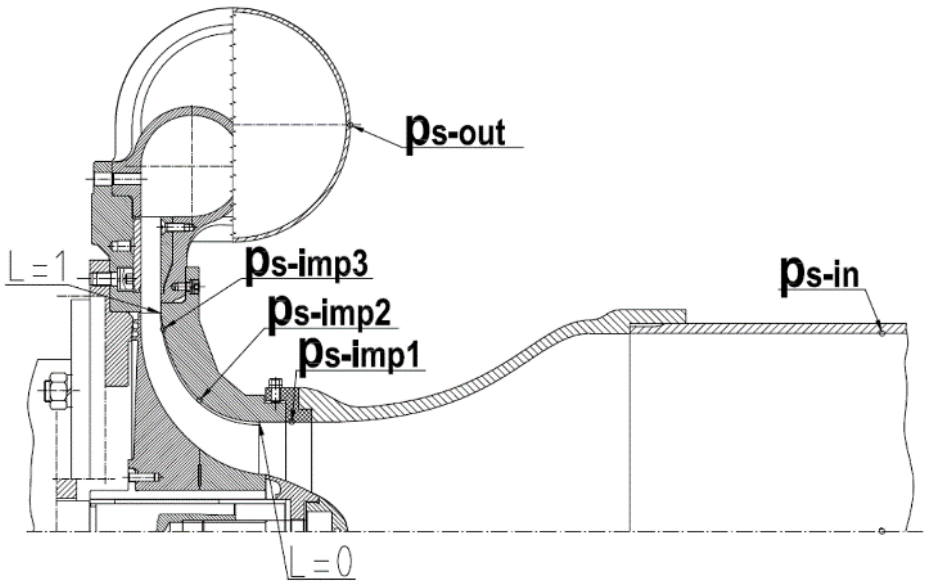


Figure 3.4: Compressor section with gauges location [9]

To provide an overview of the signal, Figure 3.5 demonstrates the amplitude of the oscillations and the mean value for compressor data collected through a quasi-dynamic protocol for all sensors in the entire operating range.

The other protocol was for used for obtaining dynamic data. The valve position was dynamically changed during the measurement, decreasing the TOA value. It aimed to emulate a quick transition from stable to unstable conditions that could occur in the compression system, due to a sudden change in machine operation or network resistance. Because of the manual operation of the valve, the readout of the exact throttle position was not available. Each measurement lasted for 20 seconds and covered the region of TOAs from stable operation to deep surge. Dynamic data is useful for validation of the data-driven model performance. Figure 3.6 demonstrates signals from inlet and outlet, collected through a dynamic protocol.

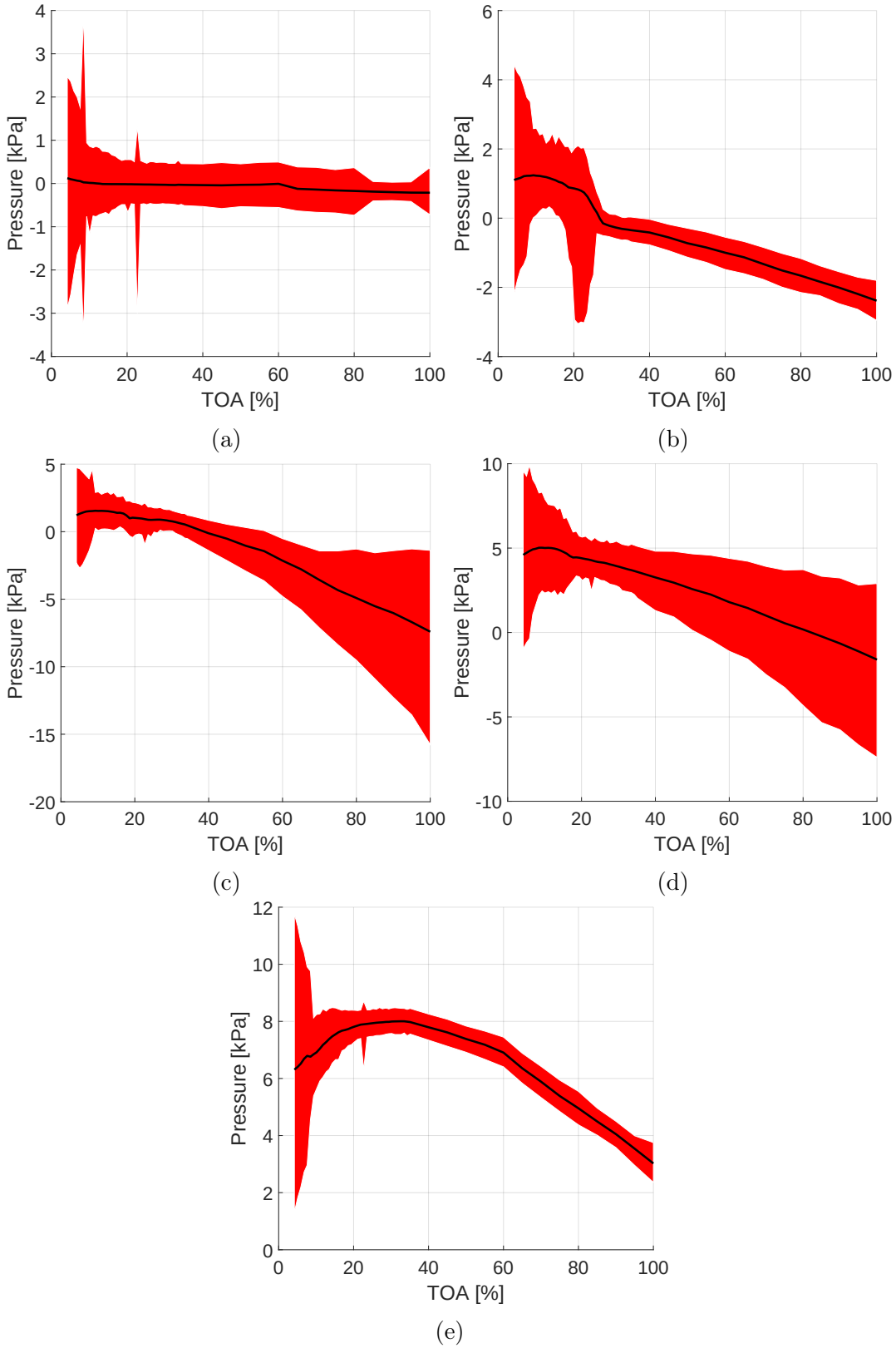


Figure 3.5: Pressure data in relation to TOA level for different sensors collected in quasi-dynamic protocol [9] (a) Inlet p_{s-in}; (b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out}

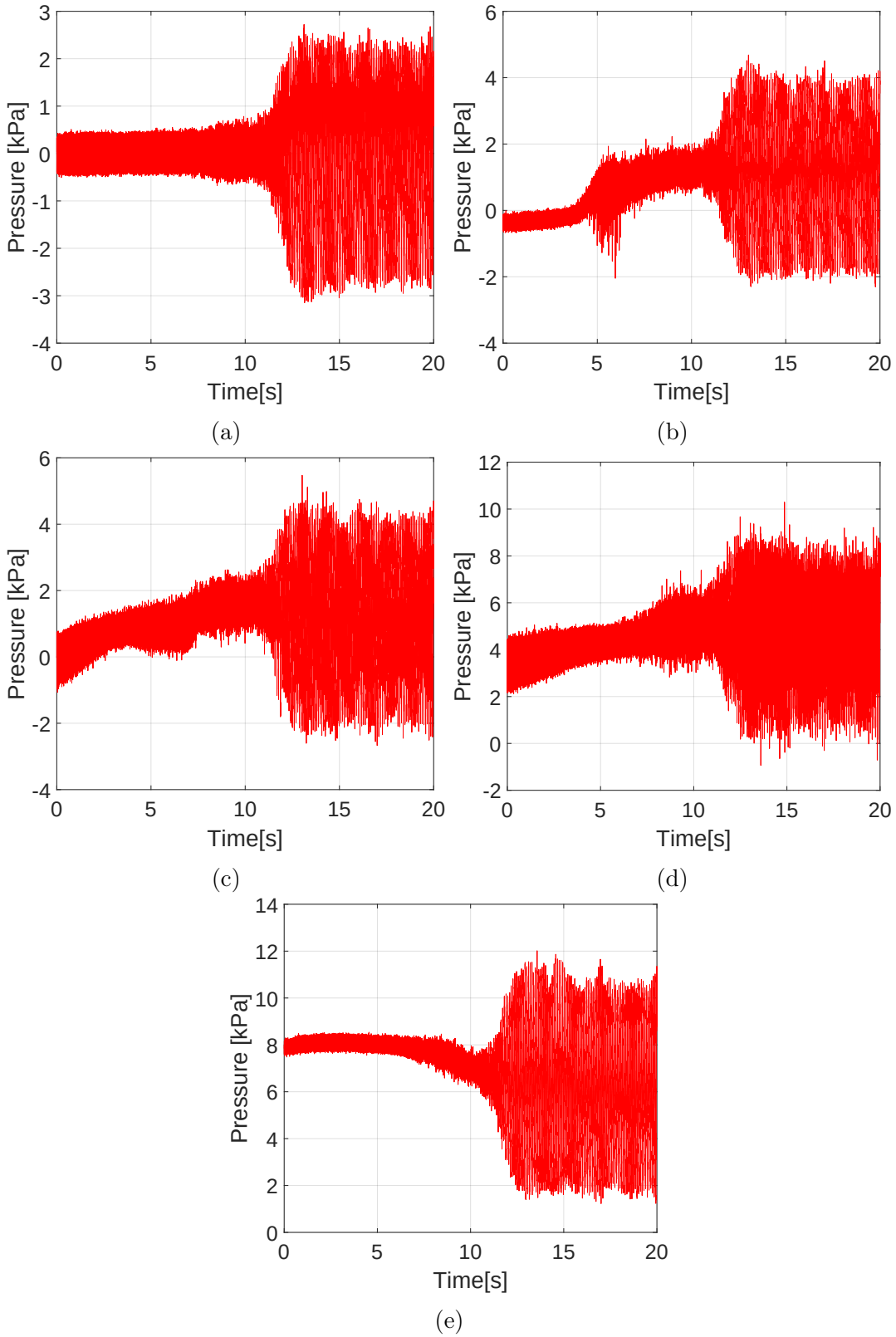


Figure 3.6: Pressure data in relation to TOA level for different sensors collected in a dynamic protocol [9] (a) Inlet p_{s-in}; (b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out}

3.3 Instabilities and data sets

The performance and analysis of the instabilities of the compressor was a subject of different research activities and its working regimes are reasonably well understood [9, 45, 97]. Several signal processing methods were used to detect and define instabilities present in the signal, including continuous wavelet transform [45] or spectral maps [97]. Thus, the data can be efficiently used for benchmarking different instabilities detection methods.

3.3.1 Observed instabilities and their physical character

Three distinct, general stability conditions were identified in the experimental data: *stable*, *locally unstable* and *globally unstable*. The *globally unstable* conditions were furtherly differentiated into transient and deep surge.

Stable operating conditions are obtained for a number of TOA, close to the compressor design point, where no unstable structures are observed. They correspond to normal compressor operation without disturbances.

Locally unstable conditions are characterized with a presence of inlet recirculation. Its signature is the strongest for the p_{s-imp1} sensor before impeller, therefore it is a preferential location for its detection. Inlet recirculation does not have a single dominating frequency, but causes a broadband excitation in the range of frequencies around 1000 Hz [45]. Its effect can extend streamwise into the impeller as the inlet recirculation progresses. The alternation in the flow characteristics at the impeller could potentially bring differences in the flow further downstream of the impeller, however it has not been shown before. The signature of inlet recirculation may be difficult to identify using the data from the sensors downstream that below impeller due to strong mixing in the impeller and the volute.

Globally unstable conditions, as defined in this study, incorporate transient and deep surge conditions. They both are termed global as their signature is vis-

ible throughout the whole compressor system. The exact physics of the transient conditions has not been understood [9]. The behaviour based on the pressure signal seems chaotic and intermittent, no coherent structures have been observed in the data. It is possible that transient conditions bear some resemblence to mild surge according to Ribi taxonomy [2], where global pressure oscillations of higher frequency than at deep surge are present in the system. However, there is no single dominating frequency, which would be expected for a mild surge [7] and oscillations cover the range of frequencies around the rotational speed of the impeller, ranging from 40 to 200 Hz. A clear signature of rotating stall was not observed in the system, although certain traces of sub-synchronous frequencies amplified for the sensor at the impeller end were noted [9]. Lacking a strong link between established instabilities and observed signal characteristics, this thesis sticks to the transient conditions name proposed by Liskiewicz. Deep surge conditions are characterized by a distinct change in signal characteristics. The pressure signal amplitude grows importantly and the majority of signal energy is present at Helmholtz frequency, which is at approximately 11 Hz for the investigated machine [9].

The compressor under analysis was adapted to withstand increased loading caused by presence of instabilities, hence the possibility of collecting data at highly unstable conditions. The effect of instabilities on the machine was not measured directly, for example by accelerators or force gauges, which could relate pressure signature to machine elements loading and vibrations [44, 103]. It is assumed that the presence of local instability is unwanted, but can be tolerated while the global instability should be avoided. However, in some applications where the extension of the operating range is of high priority and the effects of instabilities are not very strong, like in case of turbochargers [7, 194]. Then, further differentiation of globally unstable conditions could be of use, as it might offer higher flexibility of the detection system and better adjustment to the requirements of the specific application. Thus, the analysis

shown in this thesis is two-fold. First, it is demonstrated how general stability conditions can be detected. Subsequently, the possibilities of global instability differentiation are investigated.

3.3.2 Available data sets

The flow conditions and corresponding TOA values, as labeled by Liskiewicz [9] are shown in Table 3.2, where both general stability and the specific instability observed are included.

Table 3.2: Working regimes identified in the signal obtained with quasi-dynamic protocol

Valve position	General condition	Instability
$TOA < 10\%$	*globally unstable*	deep surge
$TOA \in (10\%, 17\%)$	*globally unstable*	transient phase
$TOA \in (18\%, 27\%)$	*locally unstable*	inlet recirculation
$TOA > 27\%$	*stable*	-

The analysis in this study focuses on the range from approximately 5% to 35% of TOA, covering a region from the deep surge up to the stable operation. Data for each TOA was assigned a label, although it was observed that the measurements for TOAs from the extremities of the region can be less representative for a given instability than those from the centre of the set [189]. Therefore, the most representative data sets for each condition are defined. They will be referred to as peaks of instabilities. They can be used to investigate the separability of the operating conditions using data-driven techniques. The peaks are defined based on the experimental study [9] as follows:

- Deep surge - 5 % of TOA
- Transient phase - 15 % of TOA
- Inlet recirculation - 20 % of TOA
- Stable - 30 % of TOA

The available dynamic data represent the transition from stable to unstable conditions. The range from approximately 35% to 5% was covered over 20 seconds of signal acquisition. The control over the throttle position was limited and no readback of this value was possible, therefore the exact TOA value is not known. Consequently, the dynamic data is presented with respect to the acquisition time rather than TOA (Table 3.3).

Table 3.3: Working regimes identified in the signal obtained with the dynamic protocol

Time	General condition	Instability
$t < 6s$	*stable*	-
$t \in (6s, 9s)$	*locally unstable*	inlet recirculation
$t \in (9s, 14s)$	*globally unstable*	transient phase
$t > 14s$	*globally unstable*	deep surge

Table 3.4 summarizes the number of data samples available for each condition in each protocol. A significantly larger number of data points is available for quasi-dynamic measurements, hence this data is mostly used for training and validation of the model, while dynamic data is used solely for validation. The data is divided into training and tests sets. The training set is used to define class boundaries. The test set is employed to validate if the boundaries are suitable also for the data that was not presented to the model in training. The train-test split proposed in this study is 75% for train and 25% for test. The division presented in Table 3.4 was used for the threshold-based classification with both EMD and SSA. The number of training and validation points for probabilistic approach was altered, which is given in detail in Chapter 7

The available data is divided into input signals. Those inputs are recomputed into features, following the proposed methodology. The selection of the input signal length N_s affects the number of features that can be obtained. For example setting N_s = 10,000 it is possible to obtain 200 features for each TOA.

Table 3.4: Number of data points available for different conditions in thousands

General condition	Instability	Quasi-dynamic		Dynamic	
		Train	Test	Train	Test
stable	-	13,500	4,500	-	600
locally unstable	inlet recirculation	18,000	6,000	-	300
globally unstable	transient	12,000	4,000	-	500
globally unstable	surge	12,000	4,000	-	600

3.4 Goals of the detection system

Requirements of the detection system can vary depending on the type of monitored compressor, its characteristics and application. The presence of instabilities themselves is not harmful for the compressor, but rather the effects they exert on the compressor elements. With different design and characteristics of the machines, the effects can differ [2]. Nevertheless, precise indication of the exact operating conditions of the compressor is advantageous and should be pursued by an instabilities detection system.

In this thesis, the aim is to provide a quick and robust detection of a general condition of a compressor. This means that *stable*, *locally unstable* and *globally unstable* conditions can be detected and differentiated. By fulfilling this aim, the information when to take action of suppressing the instabilities can be given to an anti-surge system. Furthertly, it is investigated if the methods used in this study could allow to differentiate the type of *globally unstable* conditions, being transient or deep surge. In some cases, it might be needed to enable temporal compressor operation in unstable range, prior to deep surge. This applies for example to turbochargers, where the operating range is very important and the effects of global instabilities on the system not as grevious as for industrial units [7]. Should a system be capable of quick and robust identification of the exact instabilities, it could enable greater flexibility in setting the objectives of the system, improving the available working range of the compressor.

4 EMD-based detection of instabilities

This chapter presents the potential of applying an EMD-based method to compressor pressure signal in order to extract and highlight the signature of unstable flow structures. First, a general analysis of the EMD decomposition is performed to understand the IMFs response to changing operating conditions. Subsequently, a parametric study of the decomposition parameters is performed to provide guidance about their selection. Further sections focus on defining instabilities-sensitive features that can be used in an EMD-based methodology for instabilities detection. At first, the differentiation of general stability conditions (*stable, locally unstable, globally unstable*) is performed with the use of the developed methodology. This approach is shown to offer safe operation to the machine and identification of unstable regimes but may be restrictive to the operating range. Hence, the possibility of further differentiation of the *globally unstable* conditions into transient and deep surge is investigated. It is difficult to extract features that are only sensitive to one of these instabilities and the signature of both instabilities is hard to fully separate with EMD due to the overlap in their spectral characteristic. Consequently, it is proposed to take advantage of the global instability sensitive feature. It holds the signature of both transient and surge and can be used for their differentiation. The content of this chapter is partly based on a manuscript co-authored by the author of this thesis [195].

4.1 Compressor data decomposition

To understand the response of IMFs to the presence of instabilities, the compressor pressure data from different operating conditions was processed through EMD and features based on all components were extracted, as defined in Section 2.3. The quasi-dynamic data was used and each input signal of length N_s was decomposed independently. For the initial analysis, the number of sifting iterations (SN) was held at eight, as suggested by Wang [188]. The length of the input signal N_s was set to 10,000 samples, which equals 0.1 seconds of a wall clock time with the sampling rate used. The selection of N_s was based on the physical characteristics of the surge phenomenon in the compressor system. The Helmholtz frequency f_H was estimated to be 11 Hz [9], resulting in a surge cycle of duration below 0.1 seconds.

Figure 4.1 presents the mean values of IMFs RMS for operating conditions ranging from stable operation to deep surge. For each position of the throttling valve, 150 input signals were used. A varying number of IMFs was obtained for different conditions, ranging from 11 to 16. IMFs from 1 to 15 were analysed as IMF 16 was rarely obtained and it was deemed of low diagnostic value. It is expected that an increase in the mean RMS value of selected IMFs can be associated with the appearance of aerodynamic instabilities in the compression system.

IMFs differ in their response to instabilities. IMF 1, holding the highest frequency range, seems insensitive to operating conditions for all sensors, as it is constant in the analysed TOA range. Low IMFs, from 2 to 6 behave differently depending on the sensor location. For p_{s-in} at the inlet and p_{s-out} at the outlet their RMS is steady over the whole analysed range. For p_{s-imp2} at the impeller centre and p_{s-imp3} at the impeller end, the values are high for high TOA, when the mass flow rate is high (see Figure 3.3) but also for low TOA, when the flow becomes unstable.

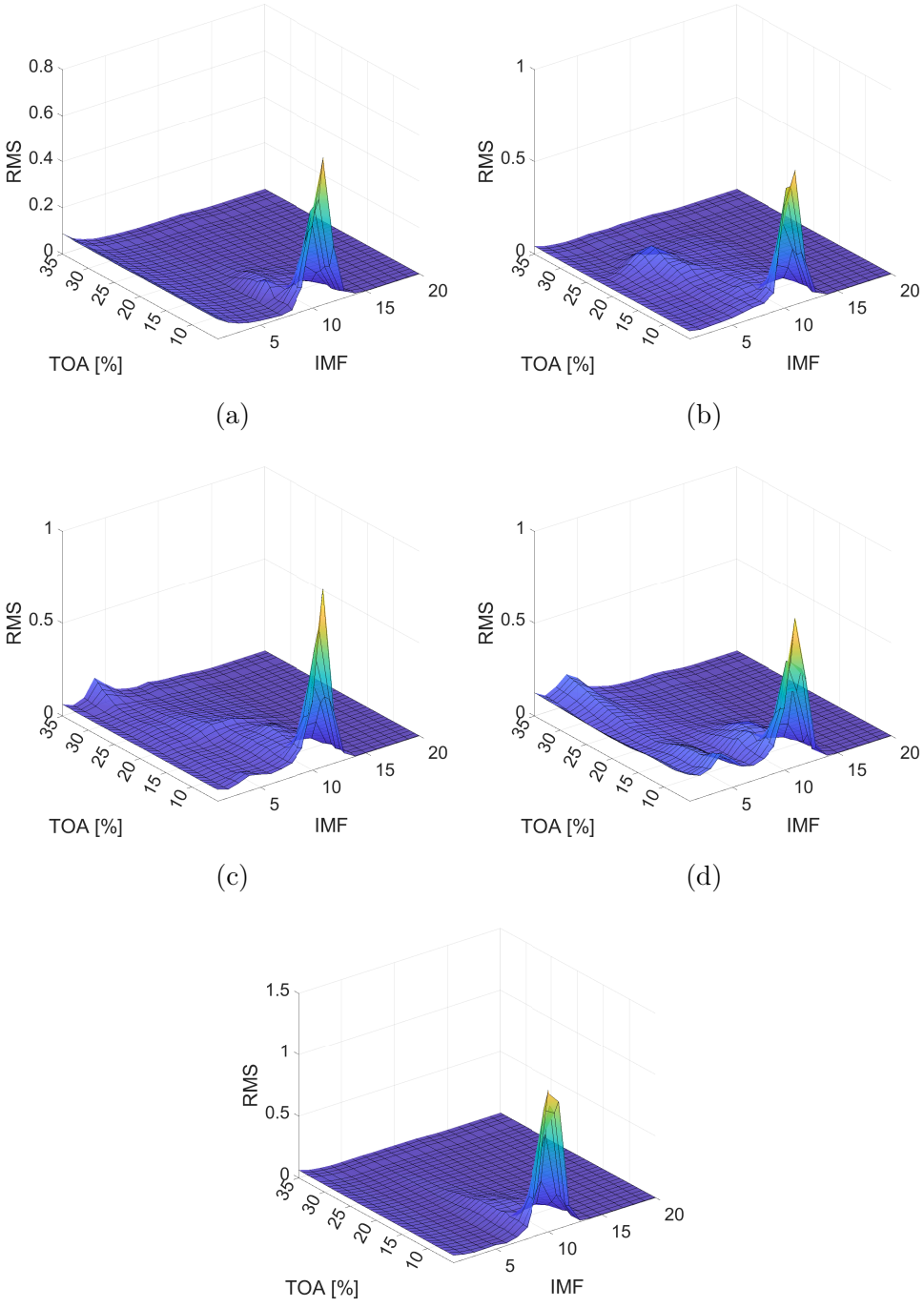
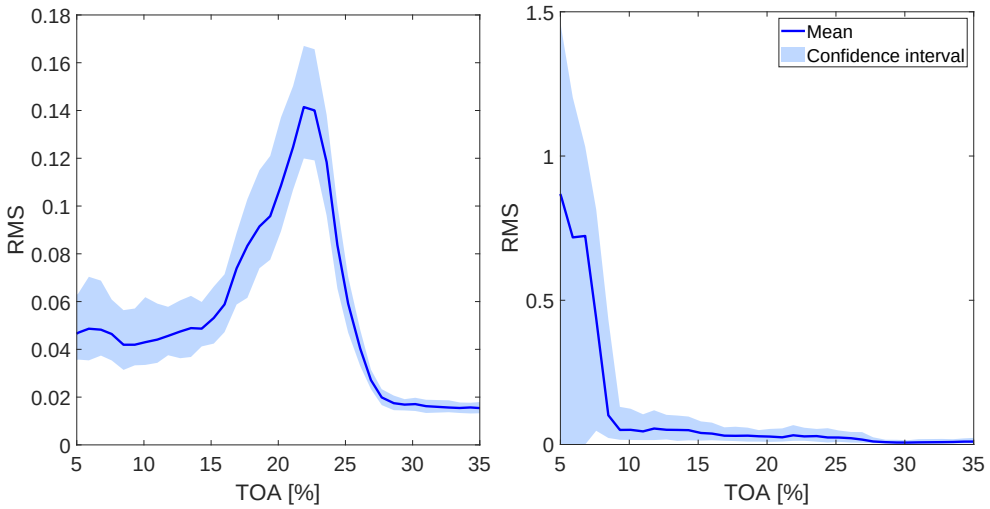


Figure 4.1: RMS values for different IMFs at different operating conditions and sensor locations in the experimental rig demonstrating differences in IMFs response to unstable structures in the compressor; a) Inlet p_{s-in}; b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out}

The IMFs from 3 to 7 for the sensor p_{s-imp1} before impeller demonstrate an important increase in RMS, correlated with local instability while being insensitive to changes of a mass flow rate which increases towards higher TOA values. High IMFs (IMF 7 and higher) for all sensors are showing sensitivity to global instability, with the highest RMS value for the outlet sensor. The indications regarding the location of instabilities provided by IMFs align with previous studies [45, 97].

When aiming for quick and robust detection, it is important to understand the dispersion of the data, not only its mean value. Figure 4.2 presents the confidence intervals around the mean value for IMFs 6 and 11 for p_{s-imp1} sensor before impeller. This sensor was chosen for visualisation purposes as it is the only one that clearly captures both local and global instabilities. The lower confidence interval was computed as $Q(0.05)$ and higher as $Q(0.95)$, making the shaded region encompass 90% of the data.



(a) (b)

Figure 4.2: Mean value of RMS and 90% confidence interval for selected IMFs from p_{s-imp1} sensor before impeller a) IMF 6; b) IMF 11

The confidence interval varies between IMFs as well as between operating conditions. It is very narrow for the stable operating range and increases with the onset of instabilities.

4.2 Influence of the decomposition parameters

For a successful EMD decomposition, it is necessary to define its intrinsic control parameters: sifting stoppage criterion and decomposition stoppage criterion. The sifting stoppage criterion influences the number of sifting iterations (SN) performed to extract each IMF, therefore it has a direct influence on the character of IMFs [90]. The decomposition stoppage criterion defines when, often based on its energy, the residual no longer contains important oscillatory components [90, 196]. This choice can affect the number of IMFs, but does not influence their content. EMD outcomes and performance are highly dependent on the input signal, mainly its length N_s and content. The signal content is an inherent property of the signal, but the input length N_s can be considered a parameter.

Figure 4.1 demonstrates that EMD can be used to highlight the changes in compressor pressure signal due to the presence of instabilities when considering the mean value of IMFs. However, as was demonstrated in Figure 4.2, the confidence interval of the components at unstable conditions may be high, hindering a quick and robust detection of instabilities despite an important change in the mean value. The mean and dispersion of the data can be affected by the selection of the decomposition parameters [188, 197]. To understand the influence of the number of sifting iterations SN and input signal length N_s on IMFs, a parametric study is performed in Sections 4.2.1 and 4.2.2. Those parameters are considered the most important in the application of EMD pursued in this thesis. The effect of the decomposition stoppage criterion is deemed insignificant in this study, therefore it is not investigated.

4.2.1 Number of sifting iterations

The choice of a sifting stoppage criterion and its influence on the extracted IMFs has been the subject of a number of studies [139, 188]. Wang [188] suggested that

keeping the number of sifting iterations constant provides the most repeatable results and that keeping the number of iterations low is important to preserve the physical character of the modes and avoid errors accumulating with every sifting step. Increasing the number of sifting iterations also increases the time needed for the decomposition. The SN influence on the outcome of the decomposition can be case-specific and it has not been proven that in certain applications, increasing the number of sifting iterations can not be beneficial. Factoring the above, the base approach employed in this study assumes keeping $SN = 8$. To investigate the influence of the number of sifting iterations on the character of the IMFs, the EMD procedure is also performed with $SN = 16$ and $SN = 32$ to investigate the potential benefit of increasing the number of sifting iterations. The signal length N_s is set to 10,000 for sifting stoppage criterion investigation.

Figure 4.3 shows the changes in mean RMS for the sensor before impeller. Increasing the number of sifting iterations results in a larger total number of IMFs. Consequently, the obtained IMFs are more narrow-banded [188, 198]. Increasing the number of sifting iterations, the information about instabilities that occupies a specific frequency range is moved to higher IMFs compared to the decomposition with the lower number of sifting iterations. The highest value of RMS for inlet recirculation region shifts from IMF 6 for 8 siftings, through IMF 8 for 16 siftings to IMF 10 for 32 siftings. The same can be observed for the peak of surge, which goes from IMF 11 through IMF 14 to IMF 17. To investigate the changes in mean and dispersion in detail, the sections of the RMS surface are taken.

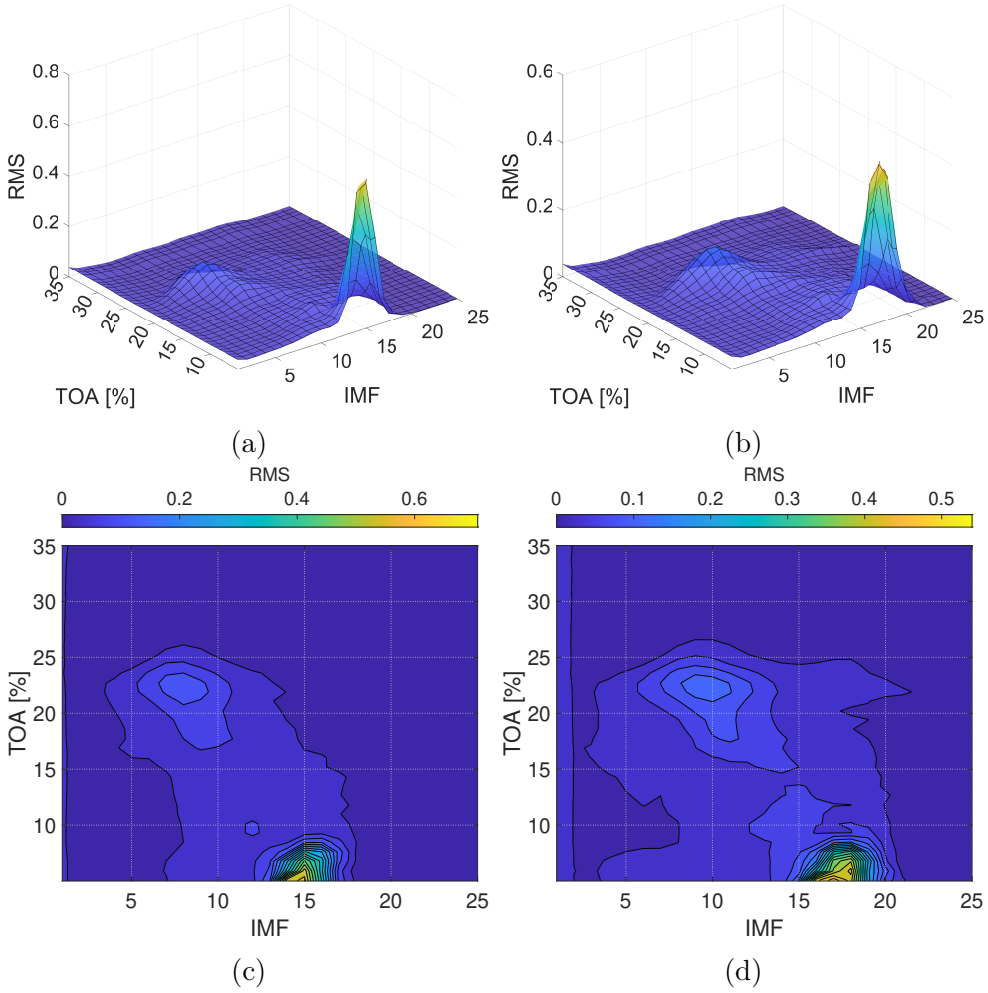
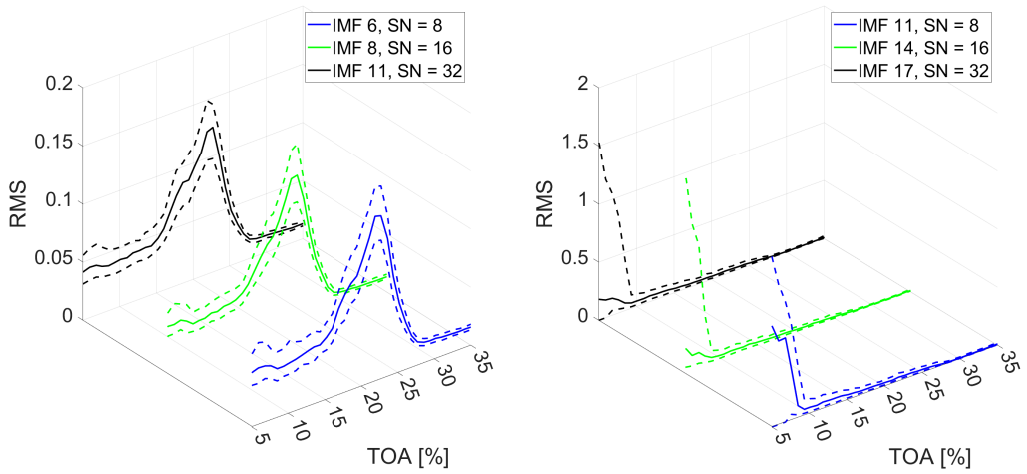


Figure 4.3: Changes in the RMS distribution of IMFs for a varying number of sifting iterations in EMD procedure for p_{s_imp1} sensor; a) surface plot for 16 sifting iterations b) surface plot for 32 sifting iterations; c) contour plot for 16 iterations; d) contour plot for 32 sifting iterations

Figure 4.4 shows the mean value of RMS along the confidence intervals for selected IMFs obtained with different SN. The mean value of the component capturing the local instability remains very similar, despite it being a different IMF for a different number of sifting iterations. The confidence interval, in this case, is also very similar between the cases. For the high IMFs sensitive to global instability, the mean value differs between the IMFs. With a more narrow spectral range of the IMFs, it is possible that the peak observed for SN = 8 is distributed between a number of IMFs. The confidence interval for all

cases is very similar. It is narrow in the stable range, starts to increase with the onset of instabilities and diverges for deep surge.



(a) (b)

Figure 4.4: Changes in mean and confidence interval of IMFs RMS for p_{s-imp1} sensor with varying number of sifting iterations; a) IMFs capturing the presence of local instability; b) IMFs capturing the presence of global instability

Considering the selected IMFs demonstrated in Figure 4.3, there is no advantage in increasing the number of sifting iterations in the investigated range. What is more, performing more iterations increases the decomposition time. Thus, for further analysis, $SN = 8$ is used.

4.2.2 Input signal length

The choice of an input signal length N_s affects both the number of points to be processed by a decomposition method and the physical time of signal acquisition. From Nyquist theorem [199], the input length also defines a maximum detectable period of the oscillations to be included in the signal. The input length also affects the content of IMFs and their number, as it influences the sifting process [141]. To investigate the influence of N_s, four different signal lengths were used: 5,000, 20,000, 50,000 and 100,000 samples, covering a physical time from 0.05 to 1 second.

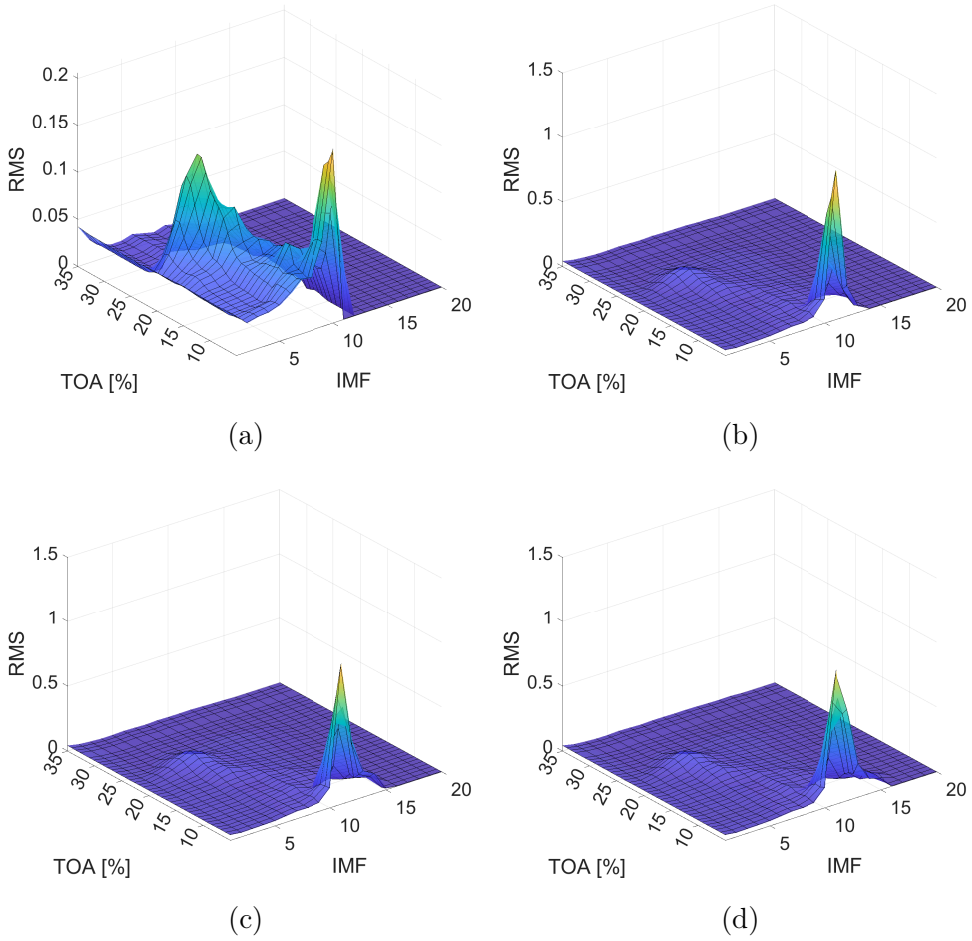
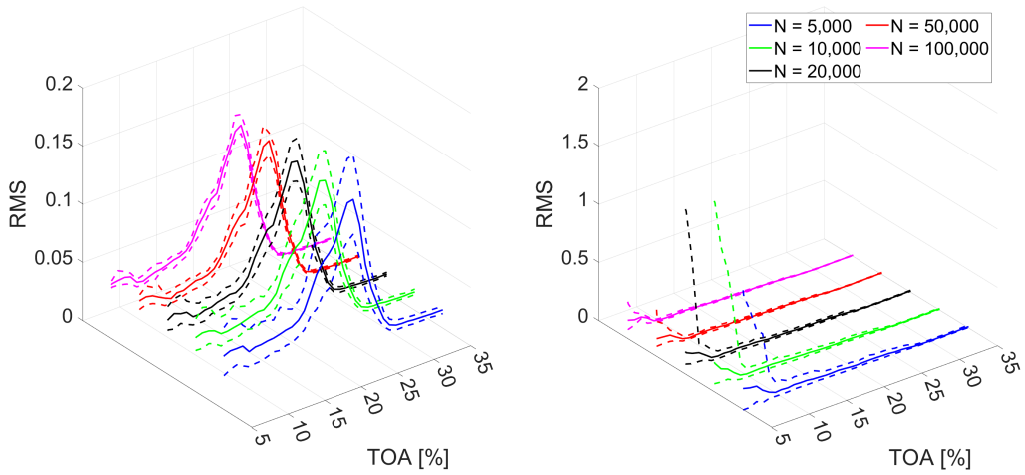


Figure 4.5: Overview of RMS distribution of IMFs for all operating conditions using varying input signal length N_s for p_{s-imp1} sensor before impeller a) N_s = 5,000; b) N_s = 20,000; c) N_s = 50,000; d) N_s = 100,000

Figure 4.5 presents the mean values of IMFs RMS. The presented data is for p_{s-imp1} sensor before impeller which held the signature of both local and global instabilities. The surfaces are very similar to each other, except for the shortest signal length N_s = 5,000, where the peak associated with the presence of surge has a lower magnitude compared to the cases with larger N_s. It is because of the surge waveform being filtered out due to the input being shorter than surge wavelength. Though, the increase in RMS is present in the globally unstable region, indicating that deep surge is not only characterised by pure Helmholtz frequency oscillations but also induces pressure oscillations at higher frequencies.

Considering the surge peak for N_s = 10,000 (Figure 4.1), its magnitude is in between that for N_s = 5,000 and those for longer input signals. This implies that by capturing the frequency range below 10 Hz, a better indication of surge can be obtained. However, the signature of surge is distinguishable with a mean value of RMS even for a short input signal.

Comparing the distribution of RMS between IMFs, the information of interest is held in the same IMFs for all the signal lengths. Both, peak for inlet recirculation and surge are captured by the same IMFs in all cases. The total number of IMFs obtained differs between the signal lengths, which is due to the extraction of oscillations having a longer period when using a longer input signal length. However, the IMFs above 15 are not sensitive to the changes in operating conditions for neither of the input lengths.



(a) (b)

Figure 4.6: Changes in mean and confidence interval of IMFs RMS for p_{s_imp1} sensor before impeller with varying input signal length for selected IMFs a) IMF 6; b) IMF 10

The choice of input signal length N_s also influences the dispersion of the IMFs, as shown in Figure 4.6 for selected IMFs capturing inlet recirculation and surge. The confidence interval, marked with dashed lines, changes greatly when

increasing the input signal length. The increase in input signal length decreases the confidence interval of IMFs RMS. Hence, the longer the signal, the better the differentiation of the conditions. However, extending the input signal length results in longer decomposition and acquisition time needed, decreasing the responsiveness of the system. As a compromise between opposing requirements, the shortest signal capturing the Helmholtz frequency expected at deep surge (N_s = 10,000) is considered a base for further analysis.

4.2.3 Timing of the decomposition

The time of instabilities detection in a compressor is crucial due to a high risk of quick damage to the machine. The required instabilities detection time can differ depending on a machine, its rotational speed and the network it operates in [7]. Two factors contribute to the time needed for obtaining stability indication. Those are:

- data acquisition time, related to the physical observation time needed for obtaining meaningful and robust indication;
- data processing time, including the signal decomposition and processing of the components.

In this analysis, only the acquisition time and decomposition time are considered as they are several orders of magnitude longer than the processing of the components.

The goal of the instabilities detection method is to provide high accuracy while minimising the time needed for detection. Depending on the physical character of instability, it might not be possible to shorten the acquisition time below a certain threshold. In each monitoring scenario the aim should be to provide an online detection. The online detection is understood as a situation where the input processing time is shorter or equal to its acquisition time.

It enables real-time processing of the data and obtaining stability indication, without causing a backlog.

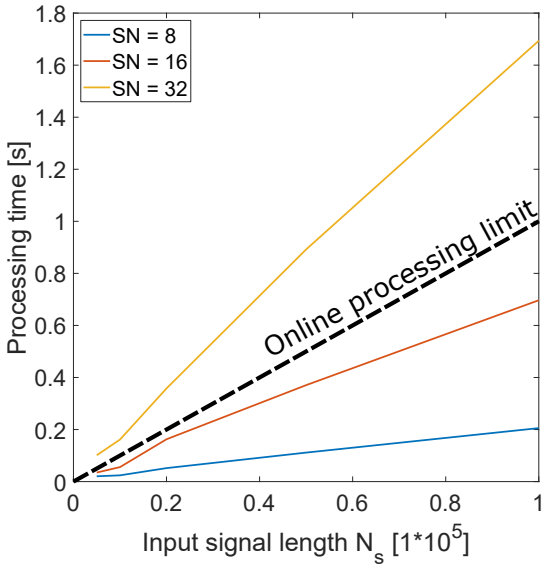


Figure 4.7: Time of processing a signal of different length through EMD for different number of sifting iterations; the dashed black line is an online processing limit above which the method no longer can provide real-time indication

Figure 4.7 shows the timing of decomposition for different input signal length N_s and the number of sifting iterations SN. The decomposition of the same input signal is performed on a PC class computer for a variable input signal length, in the range from 5,000 to 100,000 samples and the number of siftings of 8, 16 and 32. The number of IMFs extracted is limited to 15. The decomposition for each data point was performed ten times and the results shown are the average value. The figure also depicts an online processing limit, where the time of acquisition is equal to the processing time. The region below this line allows for online detection.

The time needed for processing increases almost linearly with the input signal length N_s for each SN. An increase in slope is observed for an increasing number of sifting iterations. When 8 siftings are performed, the processing time is almost five times lower than the acquisition time. For 16 siftings, it is only about 20%

lower than the limit, while for 32 siftings, the processing time is above the online processing limit. Therefore, keeping the number of sifting iterations at 8 or 16 should enable online processing with the EMD-based approach, provided that a combination of hardware and software used is at least as quick as the tested one. For performing a larger number of siftings in an online manner, a more powerful processing unit or a different implementation of algorithm could be used. One of the possibilities is a field-programmable gate array (FPGA) [200] architecture that is suited for quick computations. In the investigated case, there is no gain coming from an increase in the number of sifting iterations, as was shown in Section 4.2.1. Hence, further analysis is performed on IMFs obtained with $SN = 8$.

4.3 General compressor stability identification

The first level of EMD-based method is to investigate the differentiation of general stability conditions, defined to be: *stable*, *locally unstable* and *globally unstable* *. The goal is to find features that can be used for quick and robust detection of conditions. The IMFs showing detection potential are selected and their performance is validated. The physical interpretability of those components is discussed and, based on selected features, a feature space representation is built. The feature space representation is used for quick and robust identification of operating conditions of a compressor.

4.3.1 Locally unstable conditions

The first unstable condition appearing in the system when increasing the throttling and decreasing the TOA value is a local instability, identified to be inlet recirculation [9]. In the EMD-based method, its presence is reflected in an increase of the RMS of IMF 3 to IMF 9 for the sensor p_{s-imp1} before impeller, with

*The *italic* notation is used here to highlight the types of conditions investigated; this notation is omitted further in this chapter for clarity

the highest peak for IMF 6 (see Figure 4.1). The signature of inlet recirculation is not exclusively captured by a single component but rather distributed in between several of them due to a broadband excitation caused by this instability [7]. To understand the detection potential of the components, IMFs 4 to 7 are analysed in detail.

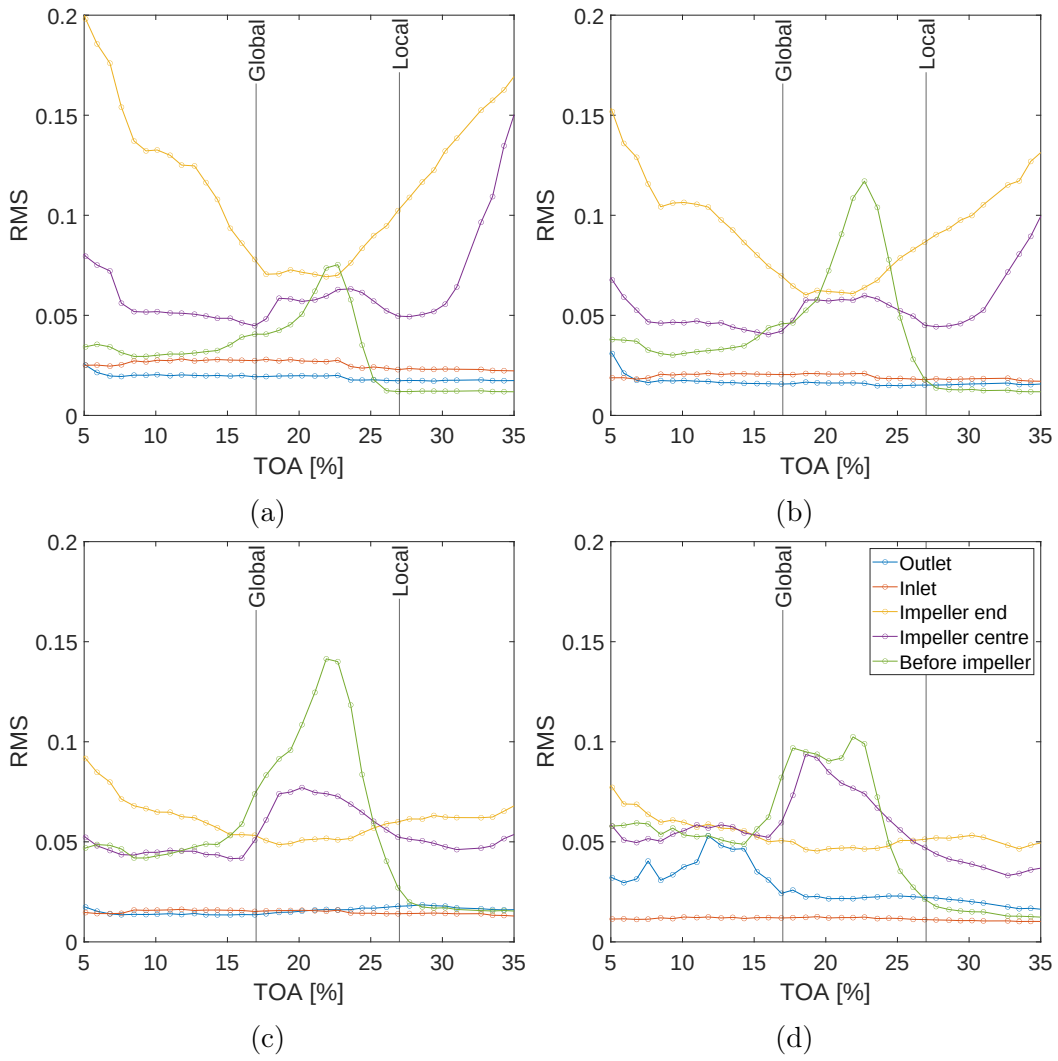


Figure 4.8: RMS of IMFs for all sensors and all operating conditions; a) IMF 4; b) IMF 5; b) IMF 6; c) IMF 7

Figure 4.8 demonstrates the changes in RMS of IMFs 4 to 7 for all sensor in the range of TOA from stable operation to surge. The RMS for inlet sensor p_{s-in} shows no important change for local instability region for all investigated

IMFs. The recirculation zone was probably not big enough to reach p_{s-in} sensor, located far upstream of the impeller. The sensor p_{s-imp1} before impeller demonstrates a substantial increase in RMS for the region of inlet recirculation for investigated IMFs. For each of those IMFs, the region of inlet recirculation can be clearly distinguished from stable operation. For IMF 5 and 6 there is a single peak in the inlet recirculation region, while two peaks can be observed for IMF 7. This may be related to the growth of inlet recirculation in upstream direction or change in its characteristics.

The effects of inlet recirculation are not exclusive to p_{s-imp1} sensor before impeller, as some of its effects are also captured by other sensors. The RMS for p_{s-imp2} sensor at impeller centre shows a subtle change for IMF 5, more important one for IMF 6 and the biggest for IMF 7. The beginning of the increase is correlated with that observed for p_{s-imp1}, but the peak is shifted towards lower TOA values. The RMS for impeller end sensor p_{s-imp3} does not demonstrate any increase over the inlet recirculation zone. For IMF 5, a decrease is observed with a minimum in the recirculation region. It is possible that the observed minimum is correlated with inlet recirculation, but it is not proven in this study. The p_{s-out} sensor does not show any change in RMS over the inlet recirculation region for IMF 4 to 7. It is probably due to strong mixing of the flow and a long distance from the impeller to the outlet sensor that the signature of inlet recirculation is not distinguishable. Overall, the best indication of inlet recirculation can be obtained with IMF 6 obtained using data for p_{s-imp1} sensor before impeller, therefore it is used in further analysis.

To understand the physical meaning of IMF 6, its spectral content is investigated. Figure 4.9 shows the averaged spectrum from sensors p_{s-imp1} and p_{s-imp2}. The values come from an algebraic averaging of 150 frequency spectra obtained for independently for 150 input signals. It can be seen that the signal amplitude is significantly higher in the range of frequencies between 300 and 2000 Hz for TOAs where the inlet recirculation was observed. Therefore, it

may be concluded that the IMF 6 for p_{s-imp1} sensor before impeller is related to the physical signature of inlet recirculation [9]. Overall, the amplitude of frequencies below 1000 Hz remains higher in the unstable region than it was for the stable region, even past the inlet recirculation region. The spectrum for p_{s-imp2} sensor at the impeller centre shows a similar increase in amplitude for the locally unstable range. Contrary to p_{s-imp1} before impeller, the level of energy for frequencies below 1000 Hz remains similar throughout the whole range.

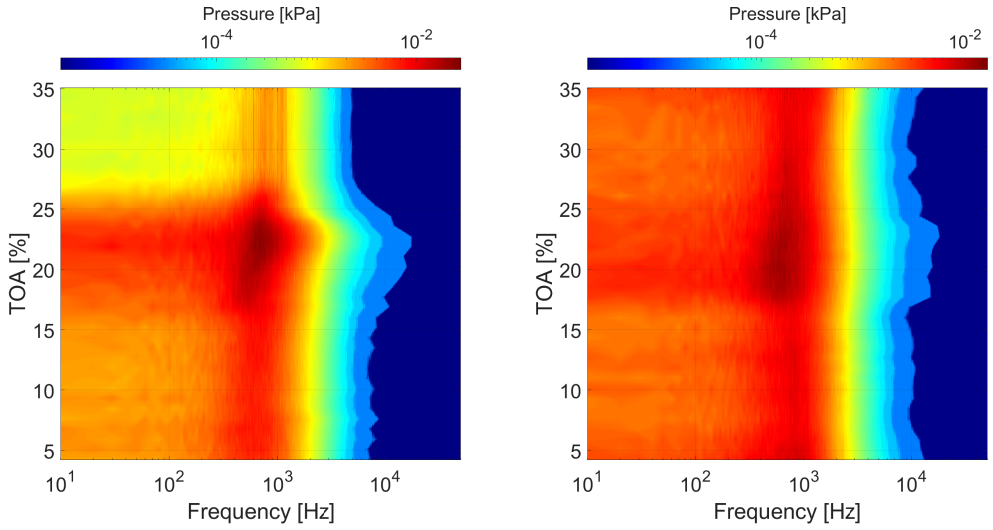
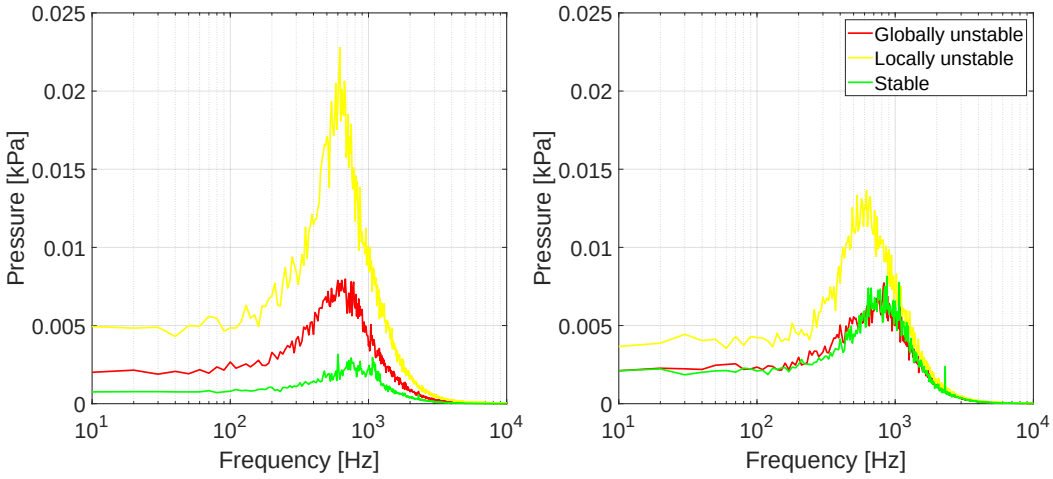


Figure 4.9: Contours of averaged spectral content of IMF 6 for a) sensor p_{s-imp1} before impeller; b) sensor p_{s-imp2} at impeller centre

Figure 4.10 shows the averaged spectra of IMF 6 for p_{s-imp1} and p_{s-imp2} sensors at stable, locally unstable and globally unstable conditions. The spectrum peak is located at around 700 Hz, for both p_{s-imp1} and p_{s-imp2} sensors. The central frequency tends to shifts to lower values for more unstable conditions. For p_{s-imp1} sensor, the lowest amplitude is present for stable operation. The inlet recirculation data shows an important jump in amplitude, but it also remains high for the globally unstable conditions. For IMF 6 from p_{s-imp2} sensor, the difference in amplitude in reference to stable conditions can only be detected

for locally unstable conditions. The amplitude for the globally unstable region is almost identical with stable operation. Comparing the relative levels between sensors, the amplitude for inlet recirculation is significantly higher for the sensor before impeller, implying much higher effect of inlet recirculation in that location.



(a) (b)

Figure 4.10: Averaged spectral content of IMF 6 for a) sensor p_{s-imp1} before impeller; b) sensor p_{s-imp2} at impeller centre for different operating conditions

The mean value of the IMFs RMS can be used for differentiation of the conditions. However, in a detection system which aims to be quick and robust, the confidence interval of the features is very important. The width of the confidence interval in the investigated case is dependent on the conditions, choice of IMFs and input signal length. The mean along with confidence interval of RMS of IMF 6 for p_{s-imp1} sensor obtained using different signal lengths N_s is demonstrated in Figure 4.11. The confidence interval is computed as $Q(0.05)$ and $Q(0.95)$ for lower and upper bounds respectively. The interval is the narrowest in the stable region and its width increases when instabilities start to be present. It remains similar for the whole unstable region. The confidence interval is wider for $NS = 5{,}000$ than for $NS = 10{,}000$. It results in its lower extremity getting closer to the threshold defined based on the stable conditions

as well as a slight shift of the threshold location. A further decrease in signal length may increase the risk of misclassification in the globally unstable region due to an upward shift of the threshold and wider confidence interval in the globally unstable region.

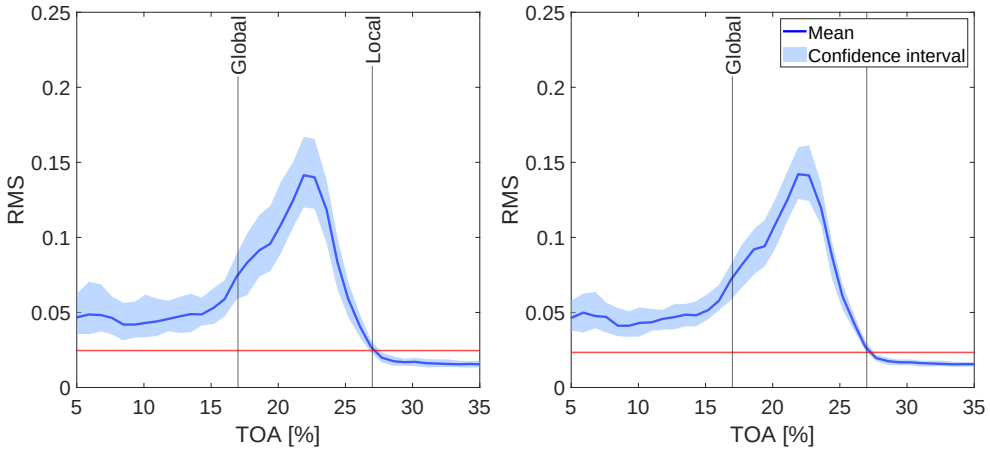


Figure 4.11: Mean and confidence interval of RMS for IMF 6 at before impeller sensor for different signal lengths; a) N_s = 5,000; b) N_s = 10,000; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation

To quantify the detection performance, the accuracy of classification is computed for all prospective features based on selected IMFs obtained for different input signal lengths. The lengths range from 1,000 to 100,000 samples. The results are shown in Figure 4.12. The accuracy is defined as a total number of true classifications divided by a total number of classifications. The classification is considered correct if the feature remains above the threshold in the whole unstable range. For the shortest input length, detection with accuracy above 80% is possible with IMFs from 3 to 7, while decreasing importantly for IMF 8. When increasing the input signal length, the accuracy increases. It remains above 90% for most of the IMFs and lengths. The accuracy over 99% is obtained for a number of IMFs and lengths, but it is the highest for IMFs 5 to 7. The shortest signal input demonstrating the accuracy of over 99% is IMF 6 for N_s = 10,000.

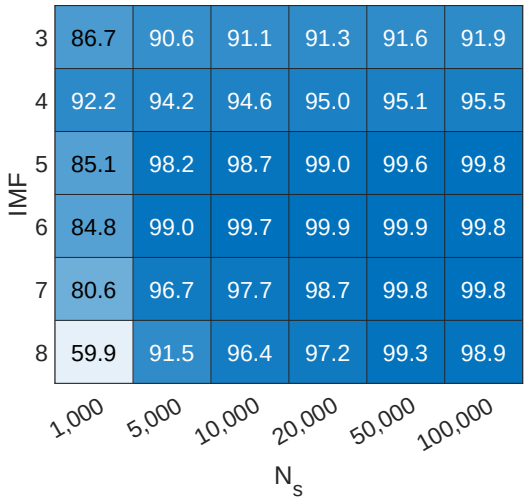


Figure 4.12: Accuracy, in percent, for the features based on different IMFs and using different input signal length

The input length $N_s = 10{,}000$ enables a good indication of locally unstable conditions from stable. **Therefore μ_L based on IMF 6 extracted for the length of 10,000 samples is defined as an EMD-based feature sensitive to local instability.**

4.3.2 Globally unstable conditions

The second type of conditions appearing in the system when increasing the throttling level are globally unstable conditions. They are characterized by a change in the readouts for all sensors inside the machine. The globally unstable conditions were not identified as one specific type of instability, but rather as two instabilities labelled transient and deep surge. The presence of the globally unstable conditions is marked with an increase in RMS of high IMFs as presented in Figure 4.1.

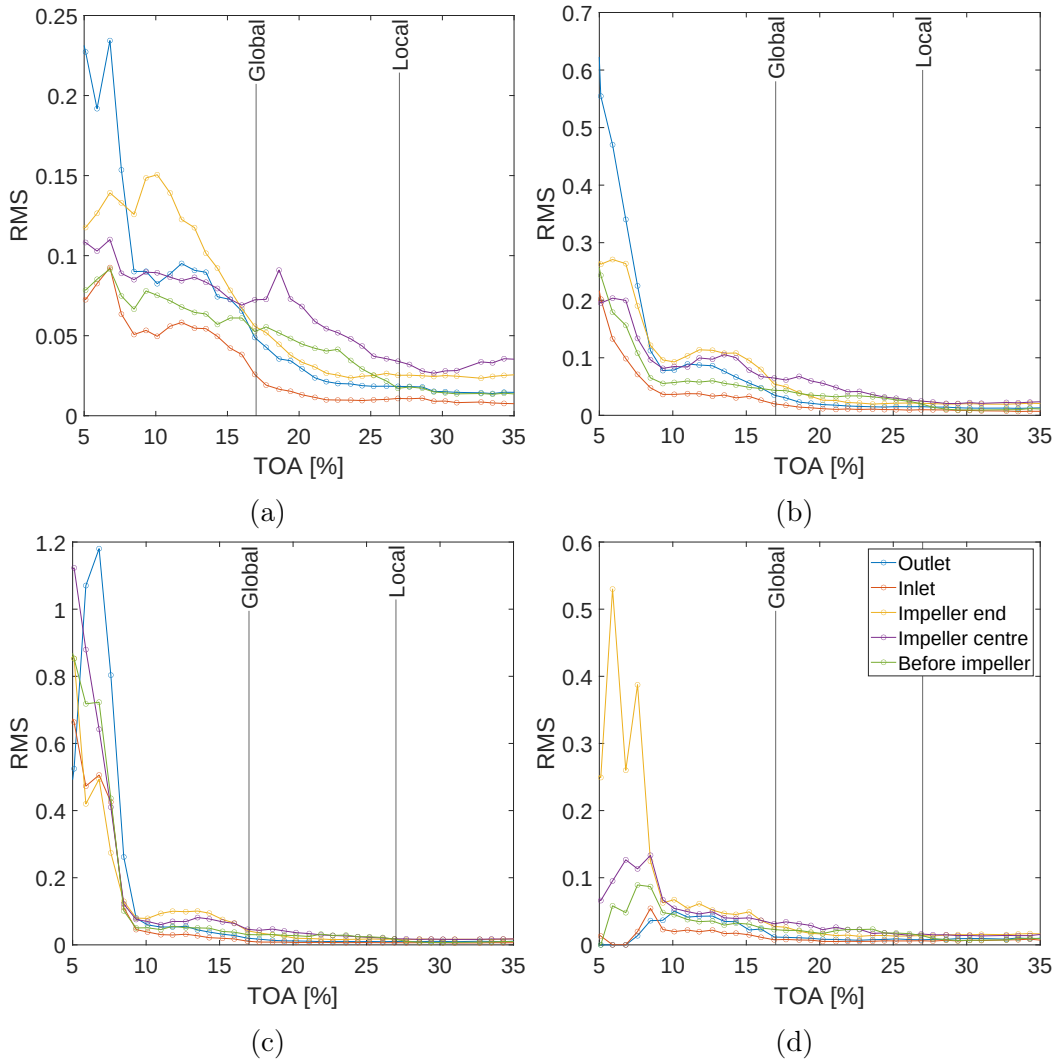


Figure 4.13: RMS of IMFs for all sensors and all operating conditions; a) IMF 9; b) IMF 10; c) IMF 11; d) IMF 12

Based on the overview, IMFs from 9 to 12 are selected as potential features and their RMS is plotted in Figure 4.13. For stable conditions, the RMS of all IMFs for all sensors is constant and low, compared to the globally unstable region. The behaviour for unstable operation differ between sensors and IMFs.

For IMF 9, the global instability is best detected with p_{s-in} sensor. The increase of RMS for other sensors takes place for higher TOA values than the defined global instability limit. The earliest increase is observed for p_{s-imp2}, as it starts at the beginning of the locally unstable range. The RMS for all

sensors increases in a globally unstable range, but the increase is not consistent, despite the increasing intensity of the instability [9]. IMF 10 behaves similarly to IMF 9, but with more consistency and a greater increase in RMS for globally unstable conditions. The overall character of IMF 11 is very similar to IMF 10.

The scale of the plot is dictated by the RMS increase for the deep surge region, starting at around 10% of TOA. The behaviour of IMFs for deep surge region is not consistent between the sensors ond IMFs. Sensor p_{s-imp1} presents a steady increase over the deep surge region. Sensor p_{s-out} demonstrates a sharp increase up to around 8% of TOA, followed by a sharp decrease with further decrease of TOA. The rest of the sensors show fluctuations in the globally unstable region, with a sharp increase at the beginning of deep surge, followed by a stagnation or decrease, followed again by an increase towards the end of the plotted range. The decrease of RMS for p_{s-out} past the middle of deep surge region may indicate that the signature of deep surge is leaking to other IMFs, as the surge intensity increases till the end of the studied TOA range [9].

IMF 12 for most of the sensors shows a very similar behaviour for stable, inlet recirculation and transient, while being completely different for deep surge region. The sharp increase at surge is demonstrated only for sensor p_{s-imp3} at the impeller end. Strong RMS fluctuations are observed throughout deep surge conditions, but the absolute value remains high above that for any other conditions. For the rest of the sensors, maximum is obtained in transient region and a gradual decline is observed for surge conditions. The mean value for p_{s-in} and p_{s-out} sensors even drops to zero for certain points, indicating that IMF 12 was no longer obtained for those sensors and conditions.

The behaviour of high IMFs for globally unstable conditions may result from a significant change in the spectral content of the signal. Decreased mass flow makes the overall energy of the signal at different frequencies lower, hence a lower number of IMFs is extracted from the signal. This way, the signature of surge can be present in different IMFs and the total number of IMFs extracted

can differ. It can influence the mean value of the IMF RMS as well as its dispersion.

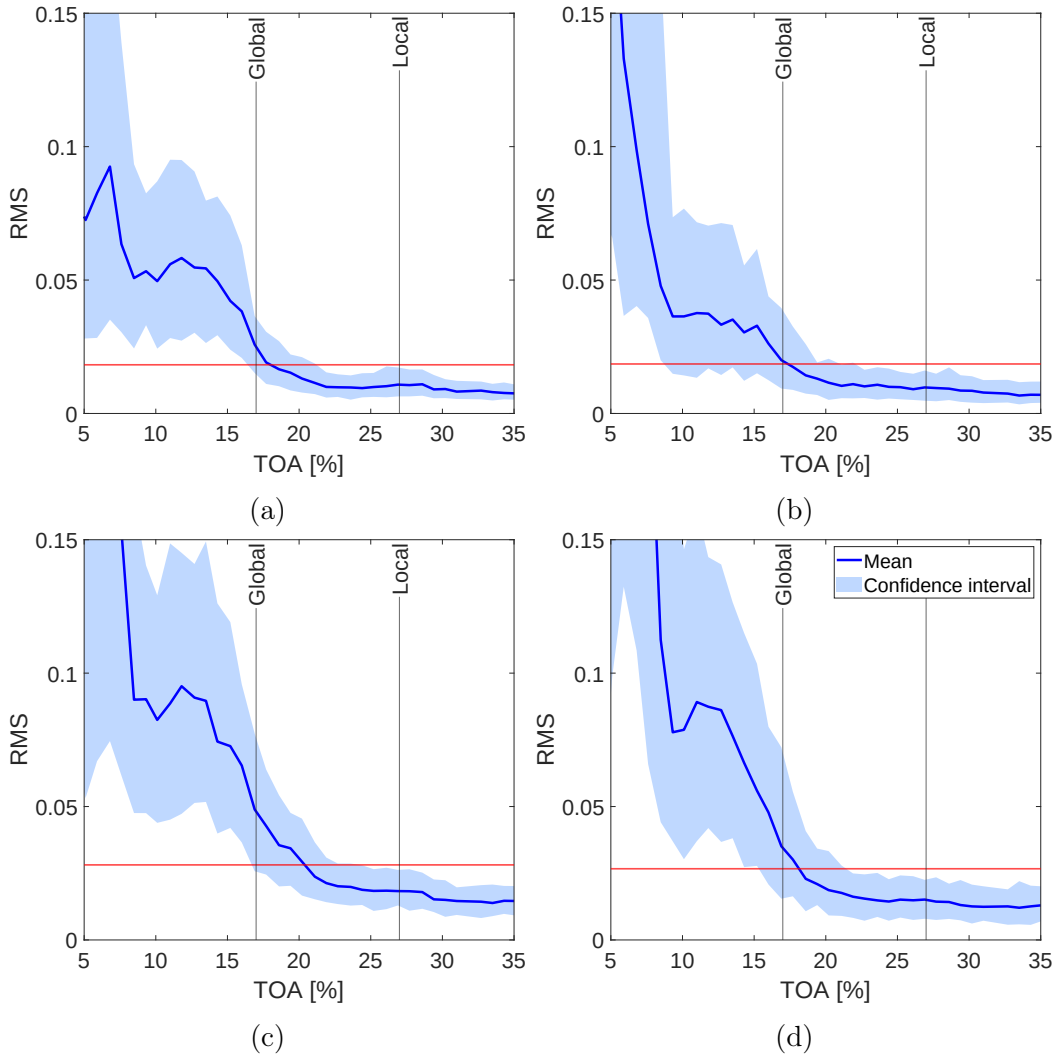


Figure 4.14: Mean and confidence interval for selected IMF obtained for data from selected sensors a) IMF 9, p_{s-in}; b) IMF 10, p_{s-in}; c) IMF 9, p_{s-out}; d) IMF 10 , p_{s-out}; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation

The changes in IMFs 9 and 10 are the most consistent, therefore these components are investigated in detail. Figure 4.14 shows the mean RMS and confidence interval of IMFs 9 and 10 for inlet and outlet sensors. In each figure, a threshold defined as $Q(0.995)$ of data from compressor stable operation is drawn to demonstrate the performance of a threshold-based classification. The

plotting range of y-axis is limited to better visualise the changes in RMS values at the beginning of the globally unstable range.

In general, all selected IMFs could be successfully used for distinguishing globally unstable conditions. However, they differ in classification of the conditions in the locally unstable zone. It is expected of a feature to remain below the stable threshold for locally unstable region and above for globally unstable one. Failing to meet the first condition would result in an increase in false positive indications, which is the case for both IMFs obtained for outlet sensor. The appearance of the global instability would be detected earlier than using the inlet sensor, resulting in false positive indications.

For outlet sensor, both IMFs display an important increase in the mean value for the globally unstable region and the confidence interval is above the threshold, causing no false negatives. For the inlet sensor, the increase in mean value of IMFs 9 and 10 is in line with reference indication of globally unstable conditions [9]. Therefore, a lower number of false positives is produced compared to the outlet sensor. However, due to a gradual change in value and the dispersion of the data, some errors are still present. The globally unstable region is marked with an increase in the mean value, however not as important as in case of the outlet sensor. In conjunction with large dispersion, a number of false negatives can be observed in the globally unstable region for IMF 10. IMF 9 performs better in this matter, as the lower bound of the confidence interval is above the threshold.

Figure 4.15 demonstrates the accuracy of classification of globally unstable conditions using different IMFs, input signal lengths and sensors. The accuracy is computed as a ratio of correct indications to all indications. The indication is assumed correct when the points are above the threshold in the globally unstable region and below for other regions.

The highest accuracy of 97% and is observed for IMF 11 with $N_s = 50{,}000$ at inlet and IMF 12 with $N_s = 100{,}000$ at inlet and outlet. For shorter input

lengths, IMF 9 with N_s = 10,000 at inlet performs best with 95% of accuracy. For both inlet and outlet sensors, increasing the input length does not lead to an increase in accuracy. What is more, the accuracy of IMF 9 decreases and accuracy of higher IMFs, such as IMF 12, increases. To explain this observation, the physical understanding of the IMFs should be built, for example through investigation of their spectral content. Despite a high overall accuracy score, the need for a very specific choice of parameters to obtain it raises concerns about the generality of the approach.

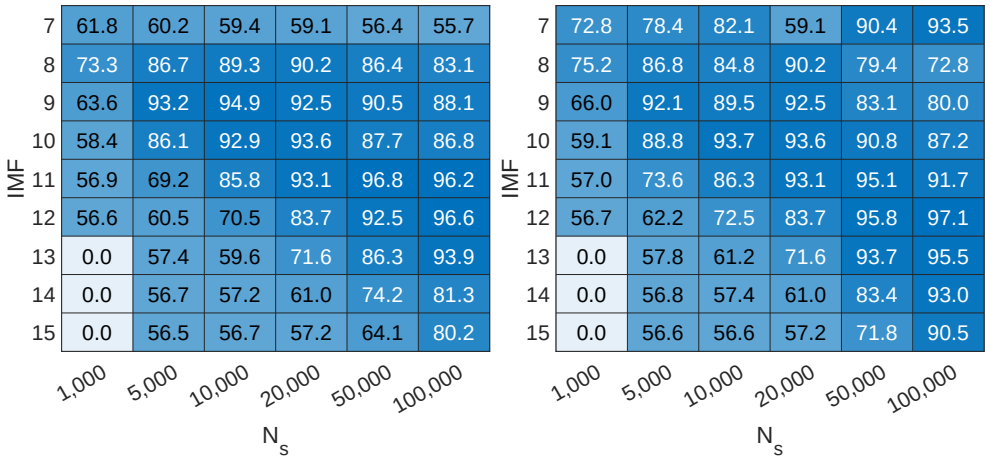


Figure 4.15: Accuracy of detection, given in percent, for selected IMFs and input signal lengths a) p_{s-in}; b) p_{s-out}

Figure 4.16 demonstrates the mean frequency of IMF 9 from sensors at the inlet and outlet for stable, locally unstable and globally unstable conditions. The amplitude for globally unstable conditions is much higher than for locally unstable or stable. There is a shift in the peak of the spectrum towards lower frequencies when globally unstable conditions onset. The peak for the globally unstable conditions is present for 60 Hz for inlet and 40 Hz for the outlet. None of those spectra show an important increase for surge frequency of 10 Hz. Therefore, it might be concluded that globally unstable conditions can be

accurately captured focusing on a frequency range higher than that of deep surge.

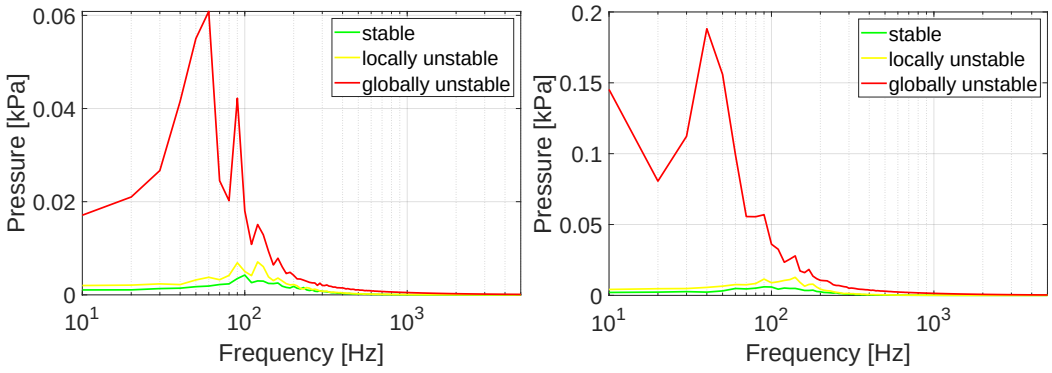


Figure 4.16: Averaged spectral content of IMF 9 for different operating conditions and sensors a) p_{s-in}; b) p_{s-out}

Despite the high accuracy demonstrated for selected components, the distance between the threshold and the bottom of the confidence interval for globally unstable conditions is small (Figure 4.14). This is caused by a wide confidence interval, the extent of which connected with two factors.

The first factor is the variability inherent to the compressor operation and captured in the signal. The compressor operation may not be stationary, hence the signal collected from the compressor is not stationary. This might be reflected in the important fluctuations of the signal power at different frequencies, captured by the RMS of selected IMFs. The pressure signature is also highly noisy, therefore even if the same global structures are present in the machine, some fluctuations of pressure are expected.

The other factor is the variability introduced by the decomposition method, that can increase the dispersion of the feature for given conditions. Investigating the outcomes of the decomposition in detail, part of the variability can be attributed to a mode mixing phenomenon, reported in EMD applications [90]. The IMFs obtained for subsequent input signals are not covering the same frequency bands and the same physical phenomenon can be captured by different

IMFs. The shift in IMFs content is due to a variation of the signal frequency spectrum in time, which results in different decomposition of the input signals. EMD extracts the components sequentially, starting from the highest frequency ones. Hence, the change in signal structure at high frequencies can affect not only the content of low IMFs, but also high IMFs. Changes in signal spectrum also influence the number of IMFs obtained [90].

The mode mixing effect observed for the compressor data is demonstrated in Figure 4.17, where the spectral content of IMFs from 8 to 15 is presented for selected input signals obtained for the same TOA from the globally unstable region.

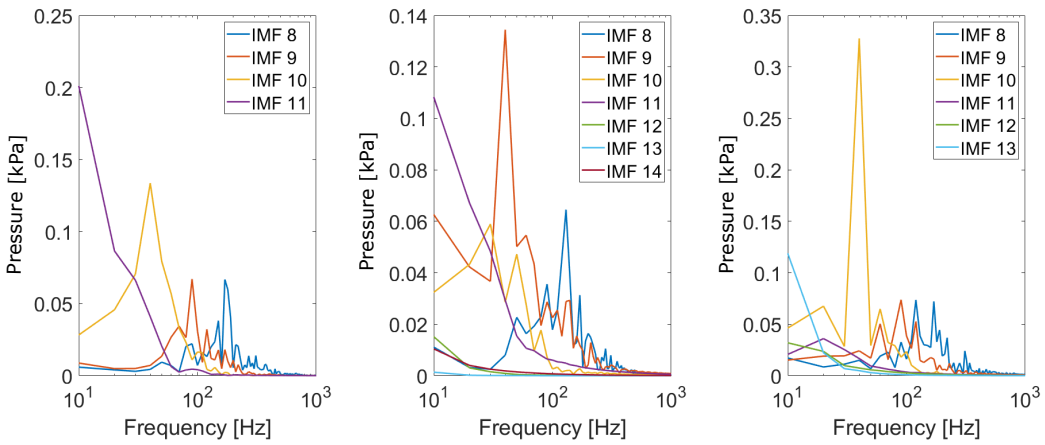


Figure 4.17: Spectral content of selected high IMFs for three consecutive input signals obtained for N_s = 10,000 at TOA = 10%

For three selected input signals, both the number of IMFs and their content differed. In the two first cases, the surge component was captured by IMF 11, however, in the third case, it was IMF 13 that held such information. The magnitude of the components also differed between the input signals, even though they were collected for the same conditions. The mode mixing could possibly be diminished by increasing the input signal length, as the fluctuations of the pressure signal spectrum are more averaged. Nevertheless, the mechanism of mode

mixing could still be present in the data. This brings a question of repeatability and generality of a global instability-sensitive feature based on a selected IMF.

High IMFs hold the signature of the globally instability, however, due to mode mixing, it is not possible to define an exact IMF that would always capture all physical aspects of the globally unstable conditions. The signature can migrate between the components, increasing the dispersion of the feature. To ensure that the entire signature of globally unstable conditions is captured, a sum of high IMFs rather than a specific component is used as a feature. By combining a number of high IMFs, it is possible to ensure that the globally unstable character of the flow is captured. What is more, the confidence interval of a feature can be significantly decreased.

Figure 4.18 demonstrates the mean and dispersion of the sums of IMFs from 8 to 12 and from 8 to 15 for N_s = 10,000. The mean follows the same trend that was present for IMF 10 alone (Figure 4.14), but the confidence interval is more narrow.

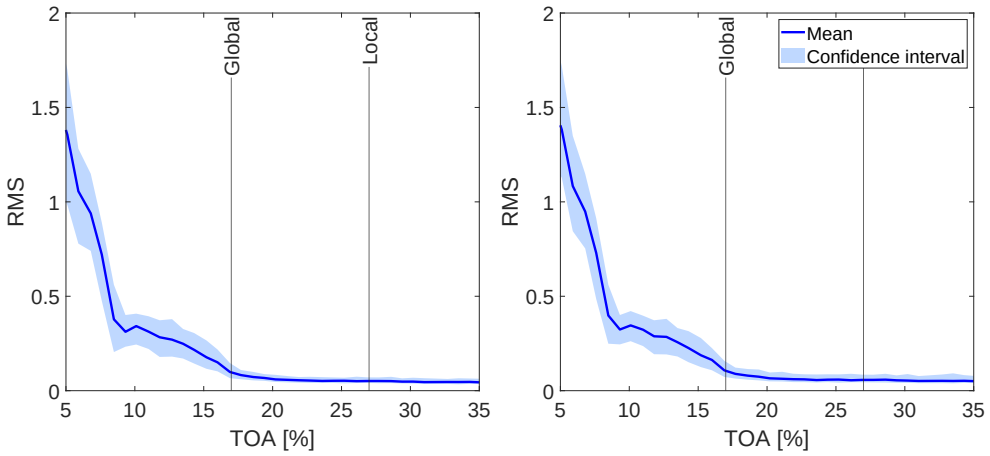


Figure 4.18: Sum of selected IMFs RMS for signal length N_S = 10,000 samples; a) IMFs from 8 to 12; b) IMFs from 8 to 15

It can be seen that up until the globally unstable conditions, the RMS value is low. It starts to grow in the transient region and spikes in the surge region.

In both of those regimes, the increase is monotonic and related to the overall severity of instability.

To investigate the effects of the components choice on the accuracy of detection, the accuracy was computed for different sets of IMFs. It was assumed that the sum consists of consecutive IMFs, with changing starting and ending IMF, i.e. the sum starting IMF 8 and ending with IMF 12 contains IMFs from 8 to 12. The results for different sums are shown in Figure 4.19.

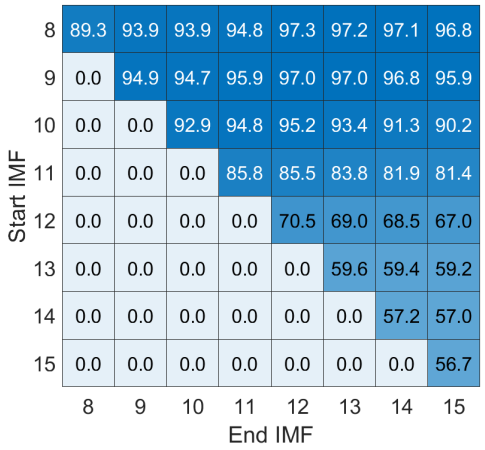


Figure 4.19: Accuracy of globally unstable conditions detection with summation of IMFs obtained for signal of input length N_s = 10,000 for inlet sensor p_{s-in}

The accuracy of detection is high when the start IMF is 8, 9 or 10. When starting from higher IMF, the accuracy drops. When investigating the influence of the end IMF, the accuracy seems to peak for the IMF 12 and decrease slightly when including higher IMFs. The overall accuracy is over 97%, which is comparable to using a single IMF for N_s = 10,000 samples. The gain is capturing all physical aspects of the globally unstable conditions. Using the sum of high IMFs, the signature of both transient and deep surge are incorporated in the feature. The two regions, transient and deep surge, can be differentiated visually when investigating Figure 4.18 due to an important change in the slope of the RMS changes with TOA.

The use of a sum of IMFs helps to diminish the influence of the mode mixing and varying number of the IMFs. It enables using an input signal of 10,000 samples, while maintaining high accuracy of the indication. **Therefore, a sum of IMFs from 8 to 15 is designated μ_G feature, to be used for detection of globally unstable conditions**. The sum of IMFs is physically interpretable and the differentiation between transient and surge conditions is possible based on the magnitude of RMS, as approaching surge the RMS would increase significantly, compared to other conditions. However, to enable the approach where the thresholds are not defined solely based on stable conditions, the data for unstable operation has to be obtained.

4.3.3 Feature space representation

Using the features sensitive to local and global instabilities, it is possible to define a feature space and project compressor data onto it. Figure 4.20 demonstrates a projection of the data onto the features for stable operation, locally unstable conditions and globally unstable conditions. The thresholds T_{μ_L} for μ_L and $T\mu_G$ for μ_G are computed as $Q(0.995)$ of data from compressor stable operation and plotted as horizontal and vertical lines respectively. The plot axes are logarithmic to increase the readability.

The points for different conditions occupy different locations in the feature space. The clusters of points for specific conditions have different dispersion. The points for stable conditions form a concentrated cluster in the low left corner of the space, representative of low values of both instability features. The cluster for locally unstable conditions extends upwards from the stable region, reaching a maximum and moving to the right in the feature space. The points for globally unstable conditions are also more scattered than those for stable operation, extending to the right in the feature space. Considering the logarithmic scale of the plot, their extent is much larger than that of stable data. Two large sub-clusters can be found, where the first cluster centres at μ_G

of around 0.3, while the other at μ_G of 1. The presence of those clusters may indicate the existence of two different modes that could possibly be separated. What can also be noted it that a path can be defined in the feature space when using the mean values of features for each TOA.

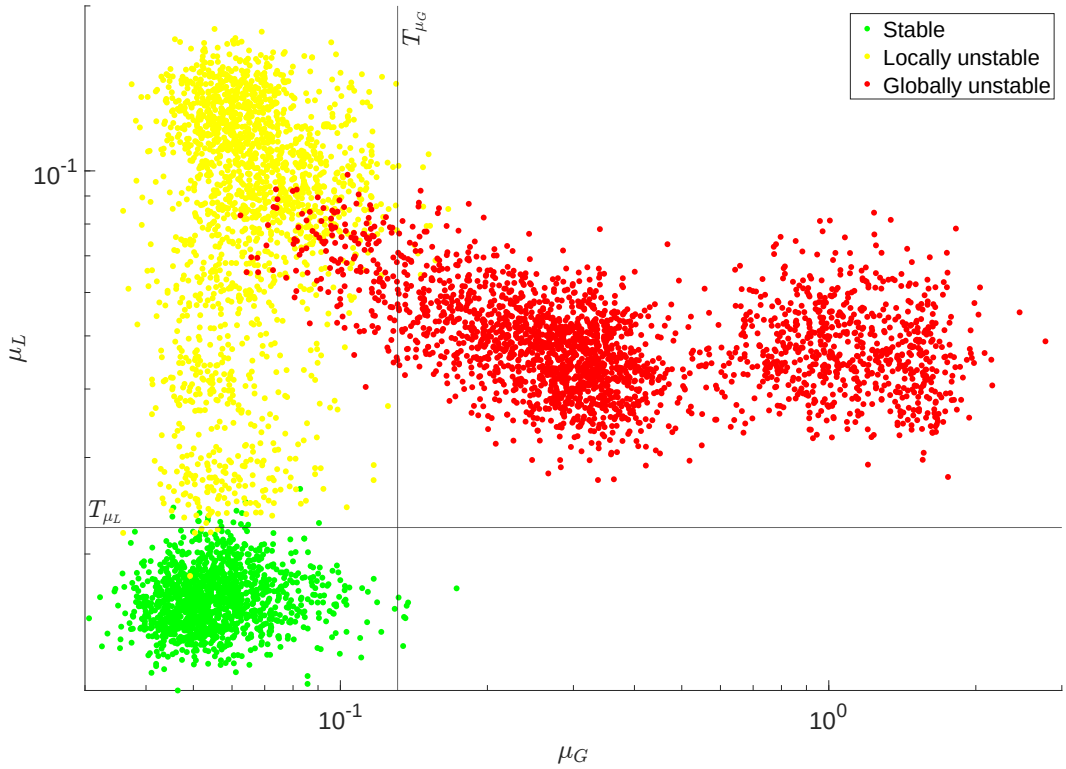


Figure 4.20: Feature space representation of compressor data using μ_L and μ_G features to capture general operating conditions of a compressor; thresholds T_{μ_L} and T_{μ_G} defined based on stable operating conditions

The feature space can be used for classification of the points obtained from the unknown conditions. The proposed approach is based on comparing the feature values for the new operating point against two thresholds, defined using stable data. The first threshold, T_{μ_L}, when surpassed, indicated the presence of instability, while the second threshold T_{μ_G} defines a type of instability. If surpassed, the global instability is present.

The thresholds T_{μ_L} and T_{μ_G} represent the data well and provide good differentiation between all three states. However, some misclassifications are present

in the region where locally unstable conditions transition into globally unstable. The great advantage of this approach is that the compressor does not have to be operated in a dangerous, unstable region to define the threshold. The accuracy of the classification with a feature space approach is validated in Chapter 6.

4.4 Globally unstable conditions differentiation

To ensure safety of operation while maximally extending the compressor operating range, the control system should be capable of approaching the dangerous operating conditions as closely as possible. The globally unstable conditions observed for the compressor analysed in this study can be furtherly divided into transient and deep surge [9]. Differentiating these conditions could allow for greater flexibility in setting the point of stoppage for the machine, should operation in the transient region be allowable from the structural and dynamic point of view. Thus, this section aims to demonstrate the possibility of transient and surge conditions differentiation with EMD-based features.

4.4.1 Transient and deep surge features

Differentiation of transient and deep surge conditions could be based on features that are sensitive to one, but not to the other instabilities. The very first step is to investigate specific IMFs that can provide a clear indication of the transient operation. Based on the overview of the RMS of the IMFs (Figure 4.1), the detection of transient conditions could be enabled using the RMS of IMFs 7 and 8. The change for different sensors and conditions is demonstrated in Figure 4.21.

The character of the RMS of IMF 7 is different for each sensor. Transient conditions differentiation could only be possible with the sensor at the outlet. The rest of the sensors are either reacting to other instabilities or do not show any consistent change in the transient region. For IMF 8, more sensors demonstrate the reaction to transient conditions. The increase in the transient region

is observed for sensor p_{s-in}, with the beginning of the increase coinciding perfectly with the defined boundary and a drop in value at the end of the transient zone. Sensors p_{s-imp3} and p_{s-out} also show a similar reaction, but the increase starts earlier, and the decrease is not as significant. The clearest indication of transient conditions is offered by IMF 8 from the sensor at the inlet, therefore it is investigated in detail.

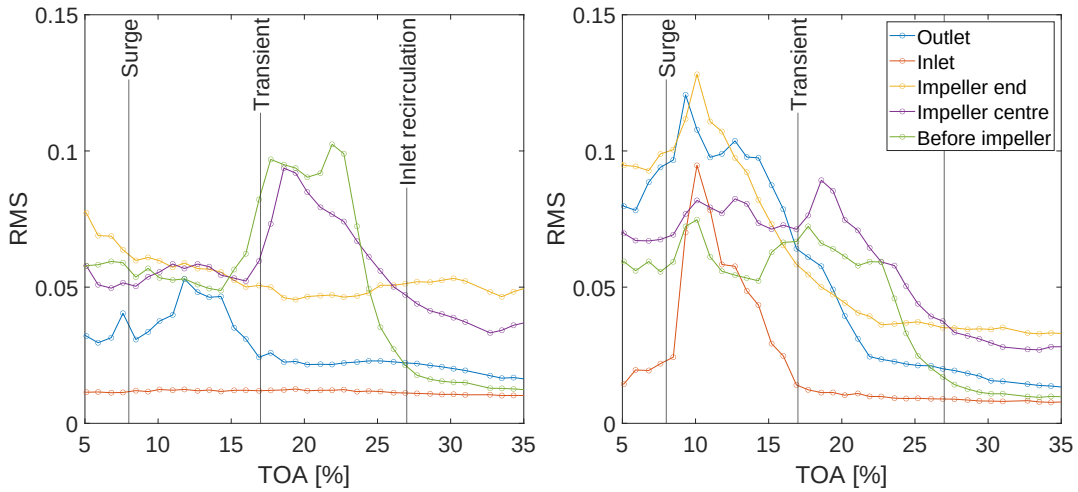


Figure 4.21: RMS of IMFs for all sensors and all operating conditions; a) IMF 7; b) IMF 8

Figure 4.22 shows the confidence interval of IMF 8 for p_{s-in} sensor for input signal lengths of 10,000, 50,000 and 100,000 samples. The confidence interval of the data increases with increasing intensity of the transient conditions but remains high for deep surge region. The interval is generally decreasing with increasing input length. For the base length of 10,000 samples, the distribution for the transient range highly overlaps with the stable region distribution. Increasing the input length five times, the confidence interval decreases and the distributions no longer overlap for the majority of the transient range. An increase in input length to 100,000 decreases further the confidence interval, but the change is not very important compared to 50,000 samples. However, despite the drop in the mean RMS value at the end of the transient region, it

remains above the threshold for the deep surge region. To gain a better understanding of its physical interpretation, the spectral content of the components is investigated.

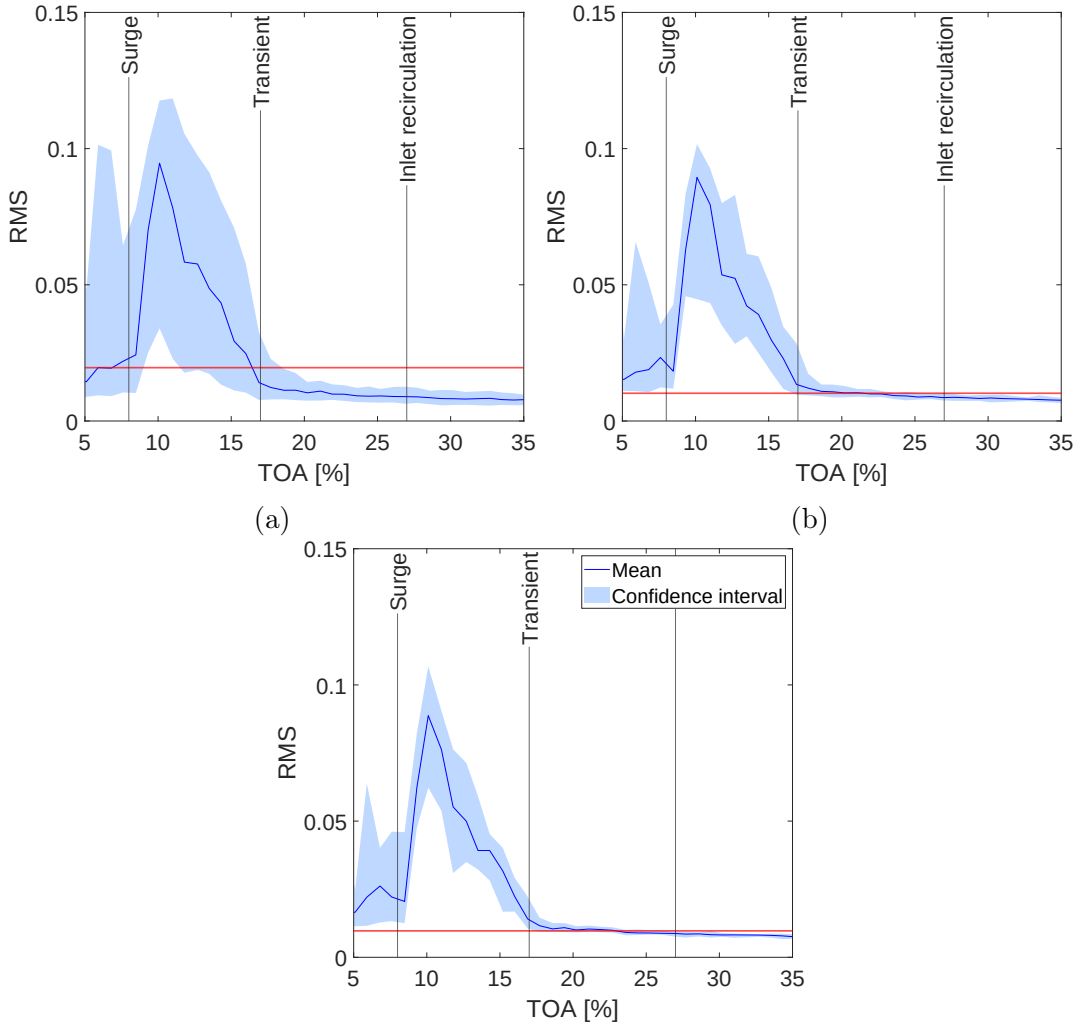
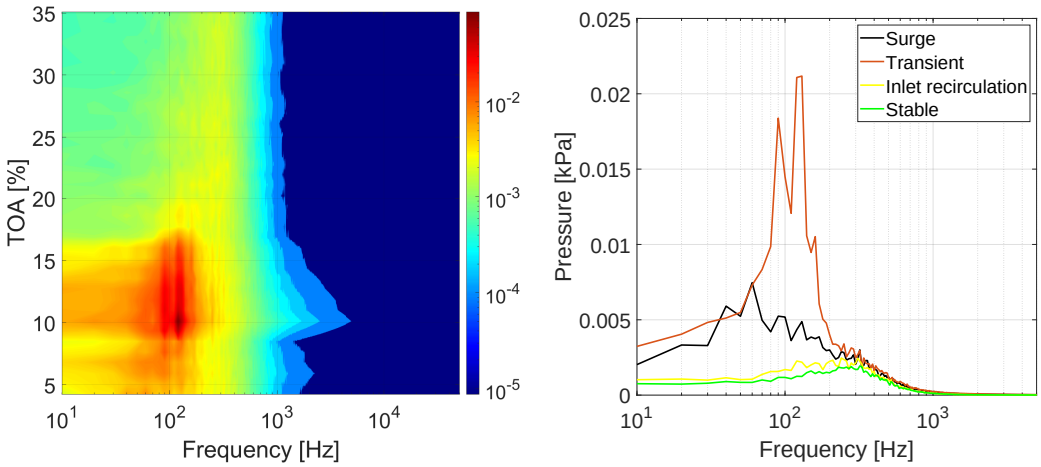


Figure 4.22: Mean and confidence interval of RMS for IMF 8 at inlet sensor for different signal lengths; a) N_s = 10,000; b) N_s = 50,000; c) N_s = 100,000; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation

Figure 4.23 demonstrates the averaged spectral content of IMF 8 for p_{s-in} inlet sensor. The frequency range of 80 to 200 Hz is the most amplified during transient operation, exhibiting two distinct peaks around the compressor's ro-

tational speed of 100 Hz. Such peaks are not observed for inlet recirculation or stable region, while they remain partly present for deep surge. Increasing the level of throttling, the energy of the signal shifts towards lower frequencies. For deep surge, the most important peak is observed at 60 Hz, but a sub-peak is also visible at 40 Hz. IMF 8 does not contain a clear peak for the surge frequency of 11 Hz, as it is contained in a higher IMF. The increased amplitude of low frequencies suggests that pressure oscillations observed during transient conditions also persist during deep surge. It is similar to inlet recirculation signature, which remains present past the peak of this instability. Therefore, the RMS of IMF 8 remains consistently above the threshold during deep surge conditions. Consequently, it becomes impossible to distinguish between transient operation and deep surge using a threshold based only on stable data.



(a) (b)

Figure 4.23: Averaged spectral content of IMF 8 for p_{s-in} sensor a) contour for all operating conditions; b) line plots for selected TOAs

Since it is not possible to isolate the signature of transient operation, it may be possible to focus on defining a feature isolating solely deep surge. With the EMD method of operation, such a feature should be based on high IMFs, holding the lowest frequency components. Based on Figure 4.13, IMF 11 should be capable of distilling the signature of deep surge. Considering the mode

mixing phenomenon, it is also expected that the dispersion of the data for surge conditions may be high, which could be altered by increasing the length of the input signal.

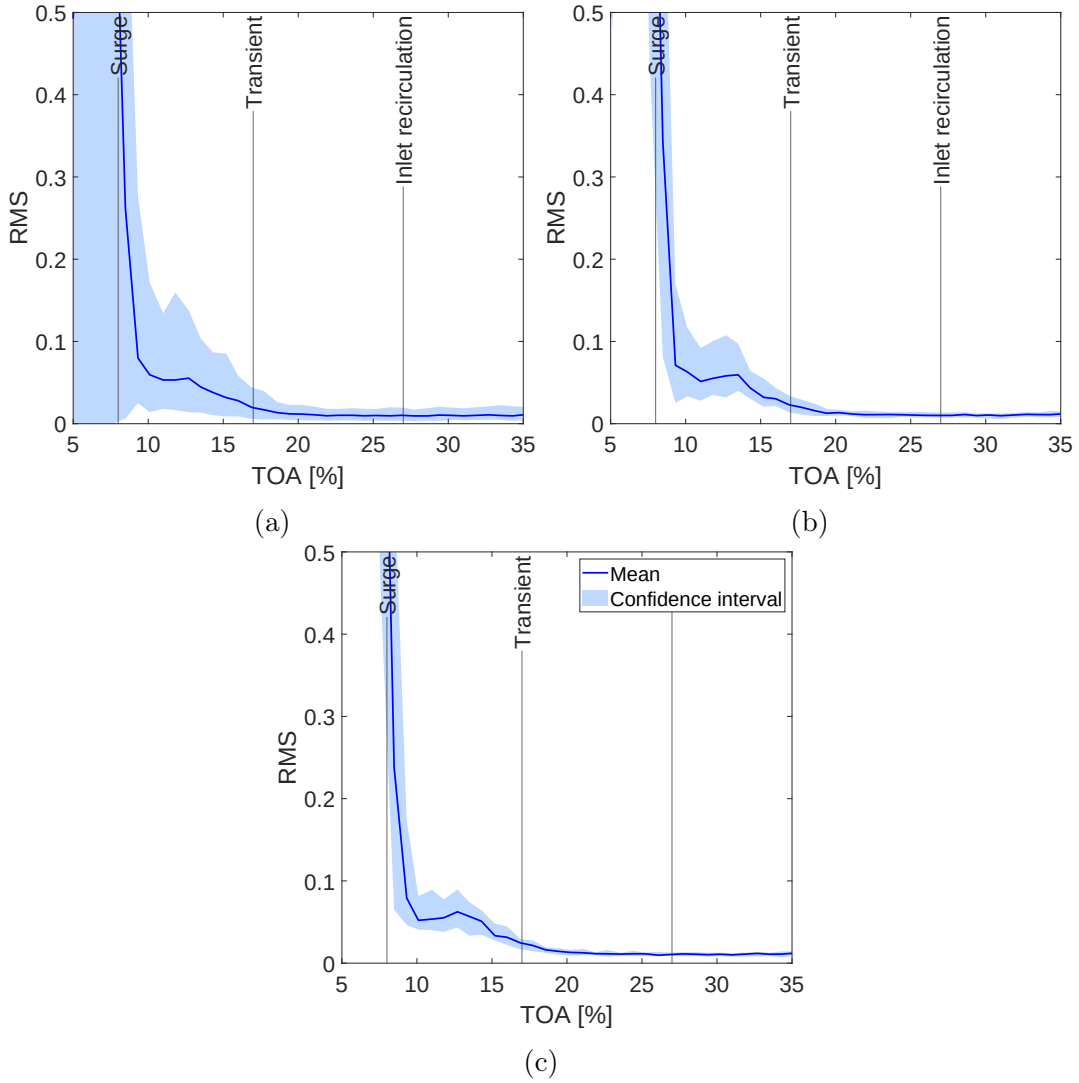


Figure 4.24: Mean and confidence interval of RMS for IMF 11 at inlet sensor for different signal lengths; a) N_s = 10,000; b) N_s = 50,000; c) N_s = 100,000

Figure 4.24 demonstrates the mean and confidence intervals for different input signal lengths from the p_{s-out} sensor at the outlet. The increase of RMS is observed for each input length in the region of transient and deep surge. This increase can be associated with the amplification of the low-frequency range, as

shown in Figure 4.25. The increase in amplitude for low frequencies, including the frequency of surge, is present before the deep surge region.

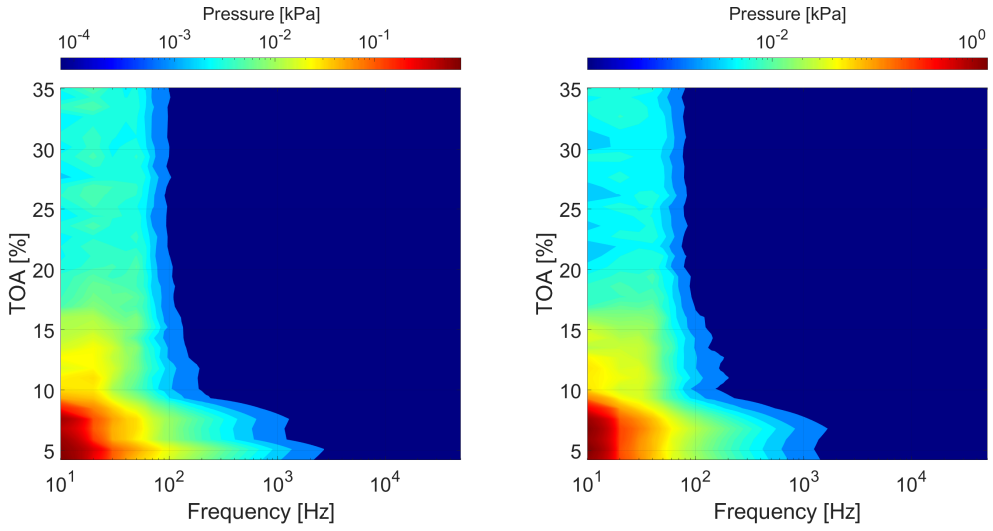


Figure 4.25: Contour of mean frequency spectrum of IMF 11 from selected sensors a) p_{s-out} sensor; b) p_{s-in} sensor

It is not possible to obtain a clear differentiation of transient and deep surge conditions with a threshold based on stable conditions with EMD-based features. It is possible to obtain features providing differentiation of the conditions, but they do not fully decouple the signature of both instabilities. Therefore, to use those features for indication, the threshold RMS level has to be set based on the feature values obtained for the unstable regime. Consequently, the possibility of building the model solely using data from stable conditions is lost. What is more, obtaining the features partly decoupling transient and deep surge conditions requires a significantly longer data acquisition period. The accurate indication is only possible when the input length was equal to 50,000 samples. It is five times more, compared to the feature for general stability conditions. Since defining features that allow one to fully differentiate the exact instability based on stable conditions is not possible, a different approach can be proposed.

4.4.2 Feature space representation

The feature μ_G, based on the sum of IMFs, is holding a signature of global instability. It incorporates both, transient and deep surge conditions. The presence of two regimes is visibly distinguishable, as per Figure 4.18 and the confidence interval is narrow for all conditions, compared to the mean value of the feature but also individual IMFs. Therefore, it is possible to take advantage of μ_G for the differentiation of transient and deep surge conditions. The differentiation is no longer based on a threshold derived from stable conditions. However, the gain with respect to the exact transient feature is a lower acquisition time needed. Figure 4.26 shows the feature space representation using μ_L and μ_G, but labelling the points according to the exact instabilities observed [9]. The non-transparent points represent the peaks of instabilities, as defined in Section 3.3.2, while semi-transparent points demonstrate data for all TOAs.

The clusters for peaks of conditions are well separated, occupying different locations in the feature space which enables their differentiation. When considering transparent points, the overlap of the points from different classes is visible, especially in the region between transient and deep surge conditions. What can also be noted is that the points form a path, from stable, through inlet recirculation and transient conditions to surge. It could be used to explore a regression approach to the detection of instabilities, however, this is not pursued in this thesis.

Overall, the feature space representation based on local and global instability feature μ_G allows for a good differentiation of the conditions, including separating transient and deep surge. There is a potential for improvement in addressing the regions of the conditions overlap in the feature space. It can be attempted with a probabilistic approach, which is demonstrated in Chapter 7.

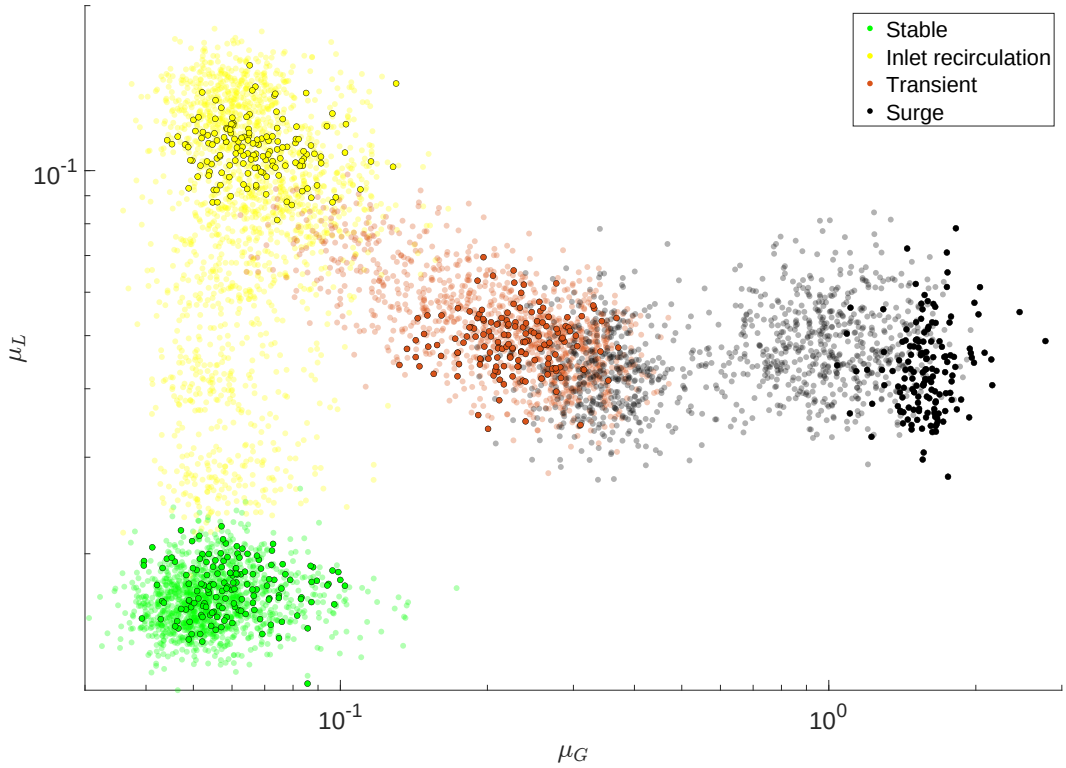


Figure 4.26: Feature space representation of compressor data using μ_L and μ_G features to capture exact instabilities present in the compressor system; scatter points with black edge represent peaks of conditions, transparent points represent remaining data

4.5 Summary, discussion and conclusions

This chapter demonstrated the possibility of using EMD-based features for quick and robust detection of instabilities in centrifugal compressors. Using a MATLAB-based EMD implementation and PC-class computer, it is possible to provide an online detection of the general operating conditions of a compressor.

The robustness of the indication, related to the dispersion of the data for given operating conditions, is affected mostly by the input signal length N_s. The longer the signal, the less dispersed the data. The reason for this may be two-fold. Firstly, a longer input signal can average out the variation that is present in the shorter portion, affecting the final indication. Secondly, the more averaged data will demonstrate less of a spectral content variation, which affects

the mode mixing in EMD. Changes in the number of sifting iterations does not seem to affect the dispersion of the data in the investigated range. The increase in the number of sifting iterations increases the decomposition time, which can make the system unfit for online operation.

The features defined based on selected IMFs can extract the signature of locally unstable and globally unstable conditions. A single IMF (IMF 6) is chosen for identifying locally unstable conditions and its choice is based on the outlook of the whole range of operating conditions. However, it could be possible to define a group of components fit for the purpose. The feature for identification of globally unstable conditions is composed of a number of high IMFs (IMFs 8 to 15), holding a low-frequency range of the signal. Such a feature is capable of capturing different physical conditions characteristic of the globally unstable operation of the studied machine.

The globally unstable conditions indicator is similar to what was proposed by Bianchini [58] for rotating stall detection. The difference is in the location of the sensor to be used. Nevertheless, both approaches acknowledge the diagnostic information coming from changes in the energy of sub-synchronous frequencies. The compressor investigated in the case study does not display clear symptoms of rotating stall in its operation, therefore no firm conclusion can be made about the detection of this instability. However, it could be expected that rotating stall would be captured in high IMFs, mostly for the sensor at impeller end. Similarly to transient conditions, it could be difficult to define a component clearly distilling the presence of rotating stall without a significant extension of the data acquisition period. The same goes for establishing a meaningful threshold value based on stable conditions. However, if such a feature can be found, it is easily incorporable into the feature space approach.

The globally unstable conditions can be furtherly differentiated into transient and surge. Detecting the exact instability present in the system may be useful when the aim is to maximally extend the operating range of the machine

and the operation in transient region is acceptable from the structural standpoint. It was demonstrated that for robust identification of transient conditions, a much longer sampling period is required than for differentiating general globally unstable conditions. For a short input signal, the confidence interval for a feature is very high, hindering the detection.

The threshold-based approach is no longer valid using the transient feature, as its RMS remains above the threshold derived from stable conditions. A similar limitation is observed for a feature capturing deep surge. Its value surpasses the stable benchmark in the transient zone, therefore one should set a different threshold to take advantage of this approach. The robustness of deep surge feature is also low for the base input signal of 10,000 samples, therefore a longer acquisition period is required. Thus, a system based on such a solution would require a priori mapping of the unstable conditions and it would be significantly more reactive due to the requirement of a longer acquisition time.

An alternative approach to exact instabilities detection is proposed, based on the general feature space approach. Based on the features of local instability μ_L and global instability μ_G is possible to define the stability of the system. The thresholds based on stable conditions represent the data well. A good indication of general stability conditions can be obtained for the $N_s = 10{,}000$, corresponding to 0.1 s of wall clock time. The features based on selected IMFs are interpretable. The increase in their value can be tied to an increase in the intensity of an instability. It can be included in a decision-making process, accounting for the severity of the unstable conditions.

The quantification of the accuracy of the EMD-based approach and a detailed discussion of its advantages and disadvantages for practical implementation in an instabilities detection system are presented in Chapter 6.

5 SSA-based detection of instabilities

This chapter demonstrates the SSA-based approach for the detection of aerodynamic instabilities in centrifugal compressors. Firstly, the overview of the RCs RMS changes related to different operating conditions is demonstrated to understand the reaction of the components to changes in stability. Subsequently, a parametric study is performed concerning how parameters of the decomposition affect the mean value of components RMS as well as their distribution. The dispersion of selected components is quantified and a preliminary selection of the components of interest for each instability is made. Then, a detailed study is performed for selected components to validate their performance as instability-sensitive features. The detection of global conditions (*stable*, *locally unstable* and *globally unstable*) with the feature space approach is demonstrated and discussed. Subsequently, to increase the accuracy of detection and flexibility of a potential system, *globally unstable* conditions are furtherly investigated to differentiate transient and deep surge states, defined in the source data. The feature space representation is leveraged in each case to perform the classification of the conditions. This chapter is based on a manuscript co-authored by the author of this thesis [201].

5.1 Compressor data decomposition

The compressor pressure data from all operating conditions was decomposed with SSA and processed, as defined in Section 2.3. For the initial analysis, the signal length N_s was set to 10,000, equivalent to 0.1 seconds of a wall clock time. It is marginally longer than the estimated waveform of surge, characterized with the frequency of 11 Hz [9]. The initial window size L was set to 50, similar to what was suggested by Logan [47]. The number of analysed RCs was limited to 15 as the contribution of higher RCs to the overall variance was marginal.

The overview of RCs RMS is shown in Figure 5.1. For better visualisation, the data is scaled to capture the variability of each component with respect to the stable conditions. The scaling was done by dividing RMS values for each RC by the mean RMS value of this RC for stable operating conditions (TOA = 30%).

The biggest change in the RMS level with respect to the stable conditions is observed for RC 1. This component increases significantly for each sensor at globally unstable conditions. The relative change is the biggest for inlet and outlet sensors. Some increase of RC 1 is also visible for locally unstable conditions for p_{s-imp1} and p_{s-imp2}. For the inlet, outlet and impeller end sensor, a slight increase in RC 1 prior to the spike in RMS value is present. Compared to the variation in RC 1, the changes in other RCs for most sensors are far less significant. A clear response to locally unstable conditions can be seen for RCs 2 to 4 for sensor p_{s-imp1} before impeller as they experience an important change in value for the inlet recirculation region. Such an increase is observed neither for p_{s-imp2} impeller centre sensor, nor for p_{s-imp3} sensor at impeller end, where potentially inlet recirculation could extend [82]. The changes in RMS past RC 5 are not significant for the selected L. It is expected, as often the first few components hold most of the information [91], however by increasing the window length it is possible that information is also distributed to higher RCs.

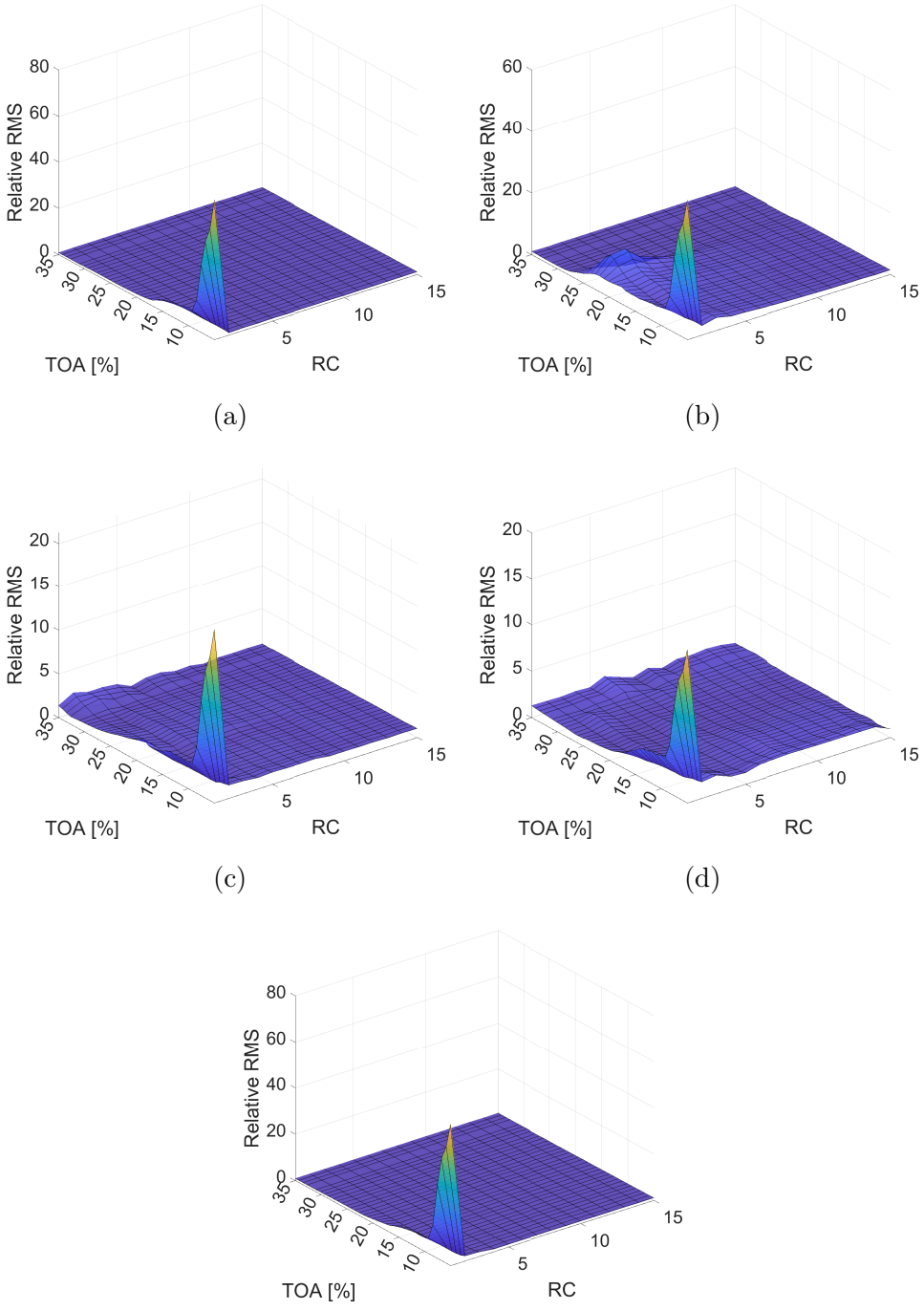


Figure 5.1: Relative RMS of RCs at different operating conditions and sensor locations in the experimental rig demonstrating differences in RCs response to unstable structures in the compressor; a) Inlet p_{s-in}; b) Before impeller p_{s-imp1}; c) Impeller centre p_{s-imp2}; d) Impeller end p_{s-imp3}; e) Outlet p_{s-out}

The changes in the RMS of RCs 1 to 4 for all sensors are shown in Figure 5.2. The data in this case is not scaled to demonstrate the relation between the same component for different sensors. The behaviour of RC 1 is very similar for all sensors. It is low and steady in the stable operating range, increases slightly for the local instability range for selected sensors and increases more importantly in the globally unstable region. The increase in the globally unstable region has two modes, where the change with decreasing TOA is less steep and more steep. The biggest increase in relative RMS value for globally unstable conditions takes place for the outlet sensor.

The behaviour of other RCs is more varied between sensors. For p_{s-imp1} sensor before impeller, the RMS level is very stable and low in the stable range. An important increase takes place in the locally unstable region for RCs from 2 to 4. The RMS level decreases for the globally unstable zone, but it remains higher than in the stable region. For p_{s-imp1} sensor at impeller centre, the decrease in RMS takes place with decreasing TOA value in the stable range. It is followed by an increase in the inlet recirculation zone, which is not as consistent and important as for p_{s-imp1} sensor before impeller.

In the globally unstable range, the RMS remains relatively stable for the transient part and increases in the deep surge region. For p_{s-imp3} sensor at impeller end, RCs from 2 to 4 generally decrease for decreasing TOA, up to the inlet recirculation zone, where they remain stable and then increase with decreasing TOA. Certain local variations are observed for the region between stable and inlet recirculation in RC 3 and in the transient zone for RC 4. For inlet and outlet sensors, their RMS remains very stable for most of the operating range. The difference in value can be observed for the deep surge region. An increase takes place for RCs 2 and 3 for the outlet sensor, while a decrease is observed for RC 2 from the inlet sensor.

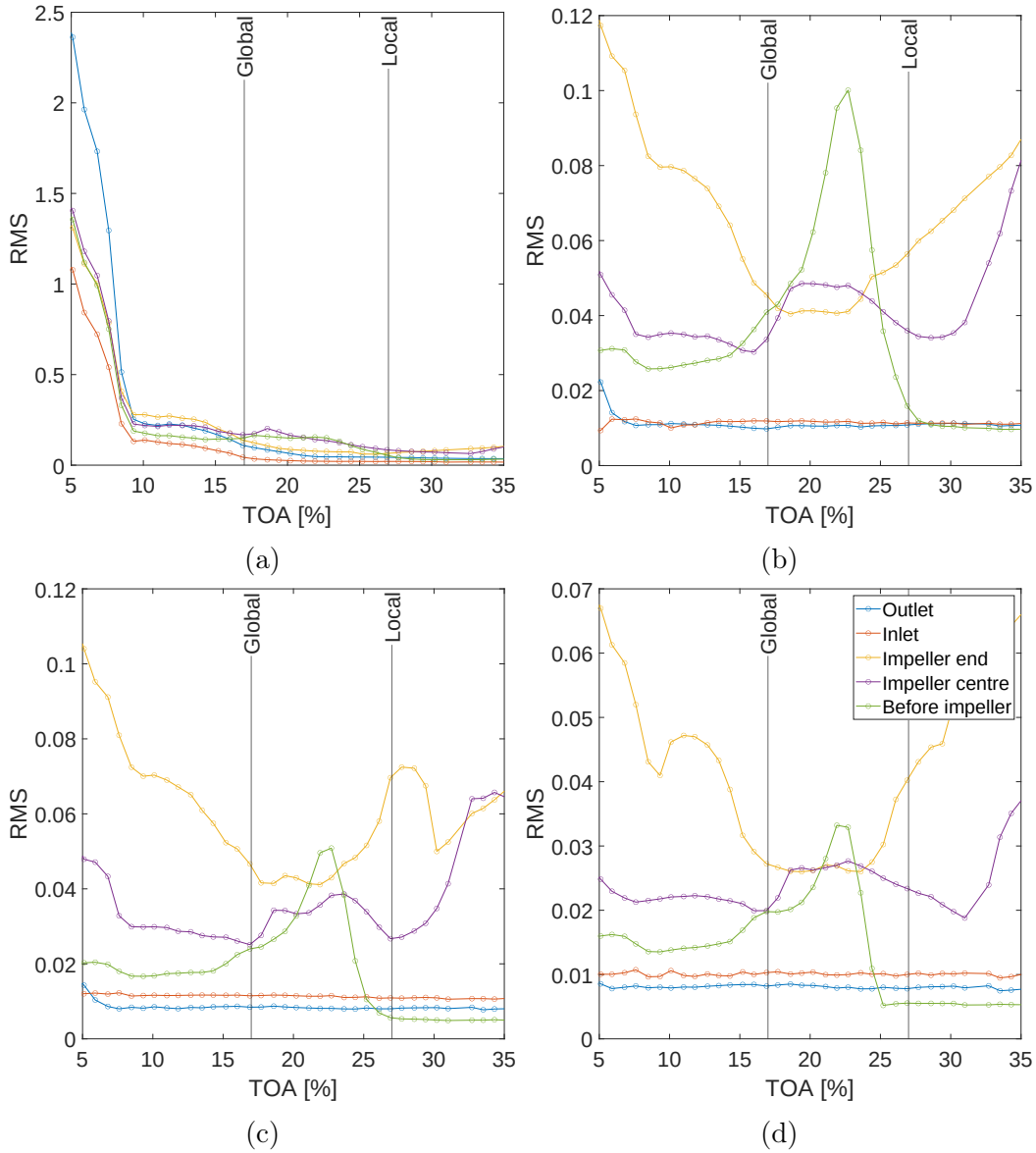


Figure 5.2: RMS of RCs for all sensors and all operating conditions demonstrating their response to different unstable flow structures; a) RC 1; b) RC 2; c) RC 3; d) RC 4

The changes in RMS of selected RCs can be associated with the appearance of instabilities in the compression system. The indications of instabilities in terms of TOA value and sensor locations are in line with previous research studies [9, 45] presented in Table 3.2 in Chapter 3. Therefore, the SSA-based features can be considered prospective indicators for instabilities in centrifugal compressors.

The possibility of quick and robust detection depends on the dispersion of the RMS for a given component and conditions. Figure 5.3 presents the mean value of RMS and its dispersion for RC 1 and RC 2 obtained from p_{s-imp1} sensor before impeller. The dispersion for the selected sensors and components is the lowest in the stable range and increases in the unstable range. Overall, it is low in relation to the mean and does not interfere with the identification of the operating conditions based on the selected components. A narrow confidence interval indicates that SSA-based features could have a high diagnostic potential.

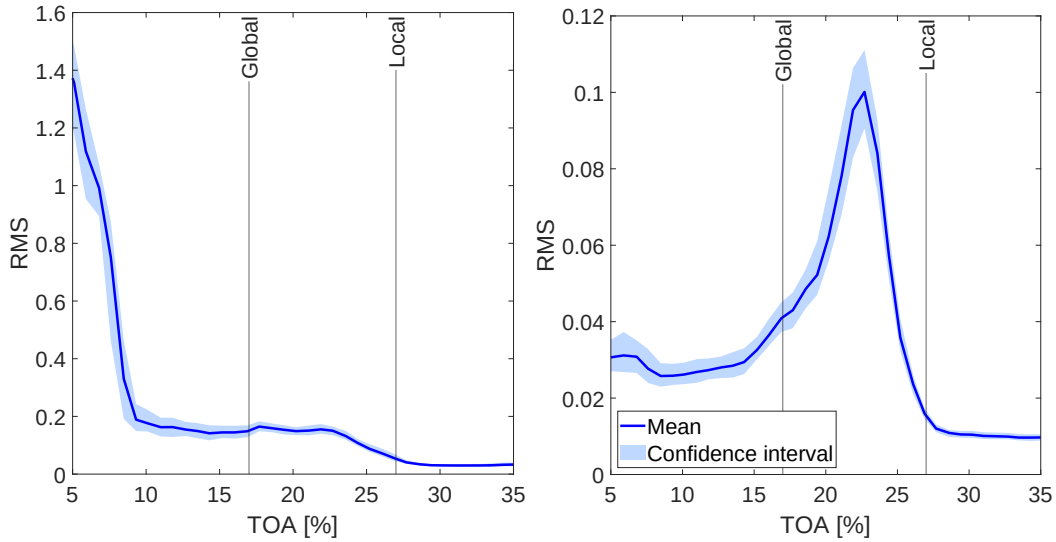


Figure 5.3: Mean and confidence interval of the RMS of selected RCs from p_{s-imp1} sensor before impeller a) RC 1; b) RC 2

SSA is a data-driven method, but it requires the selection of intrinsic parameters to perform the decomposition. A crucial parameter for SSA decomposition is window length L. In the implementation pursued in this study, the input signal length N_s is also important as it might affect the dispersion of the data and the overall predictive performance of the method. To understand the influence of the decomposition parameters on the RCs, a parametric study is performed in the next section.

5.2 Influence of the decomposition parameters

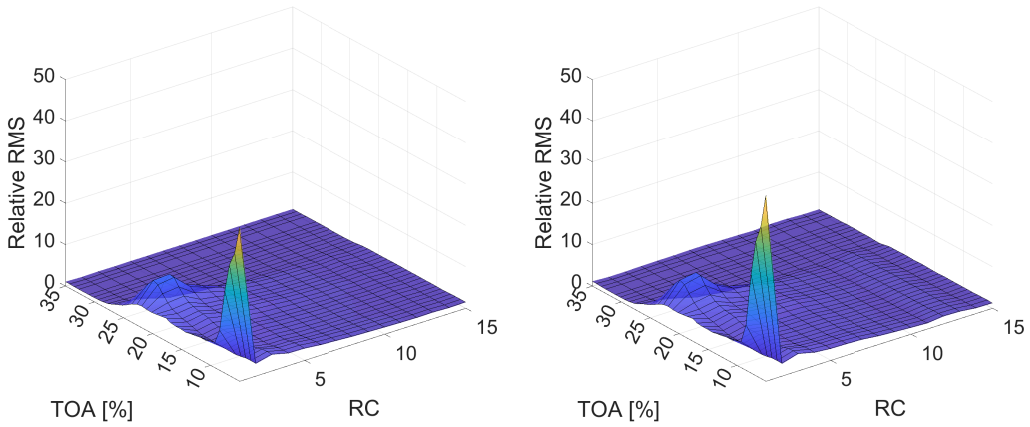
To investigate the influence of the decomposition parameters, the data from sensor p_{s-imp1} before impeller is used. It is chosen due to sensitivity to both local and global instability. Two parameters investigated in this study are signal length N_s and window length L.

5.2.1 Input signal length

Figure 5.4 demonstrates the relative RMS of selected RCs for input signal lengths N_s = 5,000 and N_s = 100,000 which are both shorter and significantly longer than the base input length N_s = 10,000 (Figure 5.1). For each decomposition, the window length was kept constant at value $L = 50$. It can be seen that the character of the first five RCs is not affected by the selected signal length as their general shape is very similar. The first two RCs were shown to capture both locally unstable and globally unstable conditions signatures, thus they are investigated in detail.

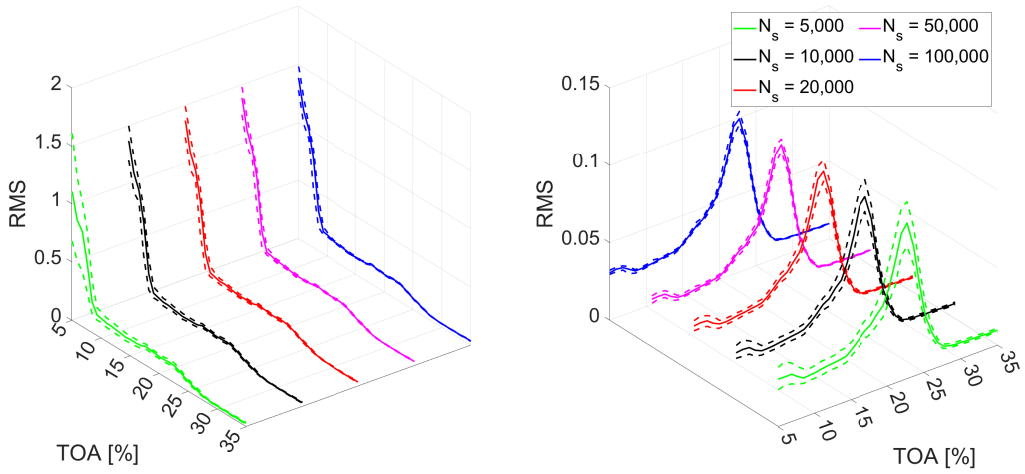
The effect of signal length N_s on RC 1 and RC 2 RMS mean value and dispersion is shown in Figure 5.5. The mean values of RC 1 are almost identical in most cases, except for the shortest input signal. For N_s = 5,000, the signature of deep surge is not as strong as it is for longer signal inputs. This is again due to the limitations coming from Nyquist theorem [199]. The dispersion of the data decreases with increasing input length, being significantly larger for N_s = 5,000 than for other input lengths.

Considering RC 2, the mean values for all lengths are almost identical and there is no important difference for the shortest input length. Considering that RC 2 captures the signature related to inlet recirculation, the cut-off of low frequencies does not affect the RMS of this RC. The influence of the input length on dispersion is the same, with a clear decrease in magnitude with increasing input length.



(a) (b)

Figure 5.4: Relative RMS of selected RCs from p_{s-imp1} sensor before impeller for different lengths of the input signal; a) N_s = 5,000; b) N_s = 100,000



(a) (b)

Figure 5.5: Mean and dispersion of RMS for selected RCs obtained using different input signal length N_s for p_{s-imp1} sensor before impeller a) RC 1; b) RC 2

5.2.2 Window length

The choice of the window length in SSA has an important influence on the RCs [91, 159]. There is no universal rule defining the optimal window length. The selection may differ depending on the character of the signal and the expected

decomposition outcomes. In general, the window length defines the longest period that is decomposed from the original signal [91]. This may be used as a guidance on the minimum window length value, in reference to the physics of the phenomena captured in the signal. Considering the upper length limit, the window length should be lower than half of the input signal length N_s [91]. This range is often wide, therefore the window length selection is commonly preceded by a sensitivity study [47, 158].

To understand the RCs sensitivity to window length changes, a parametric study of window length was performed. Windows of length from $L = 100$ to $L = 1000$ were investigated for an input signal of length $N_s = 10{,}000$. The data from p_{s-imp1} sensor before impeller was used as it contained both the signature of locally and globally unstable conditions.

Figure 5.6 demonstrates the changes in RMS repartition with changing window length L. The relative RMS is used, which is obtained by dividing the RMS for each RC by the RMS of this RC for the stable conditions (TOA = 30%). The effect of increasing the window length is mostly visible for the region of inlet recirculation. The initially concentrated peak observed in RC 2 for the base window length $L = 50$ (Figure 5.1) is getting distributed between higher RCs for $L = 100$, or between most of the RCs in the investigated range for $L = 200$ or $L = 500$. For the longest window, the increase in RMS in the inlet recirculation range is less significant and more distributed between the RCs. The character of peak in the globally unstable region, observed for RC 1 is very similar for all window lengths. The magnitude of the peak is the highest for the longest window. Since a relative RMS is investigated, it may result from the highest RMS value in unstable conditions, the lowest value in reference stable conditions or the combination of both. If the information to be extracted is not a component of specific frequency, then increasing the window length can cause the distribution of its signature between a number of RCs, which often is not beneficial to the detection potential of a method.

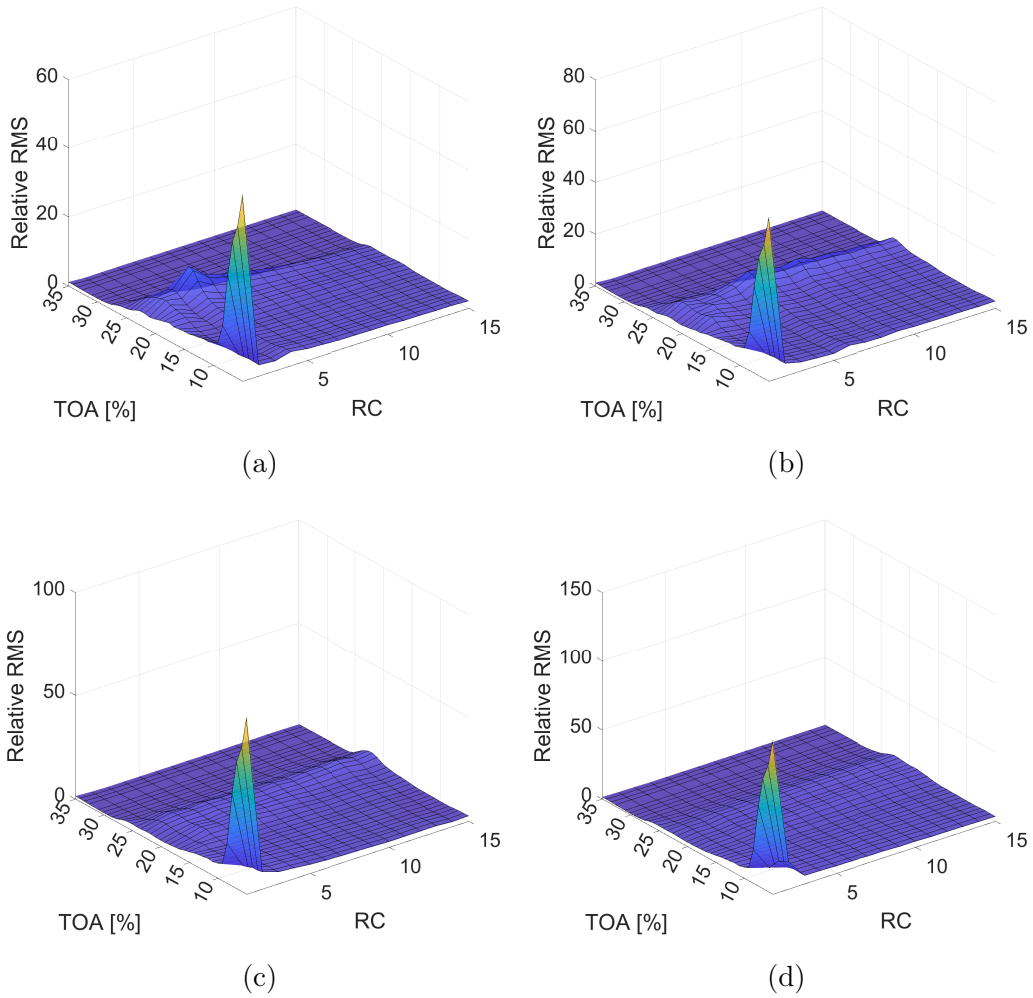
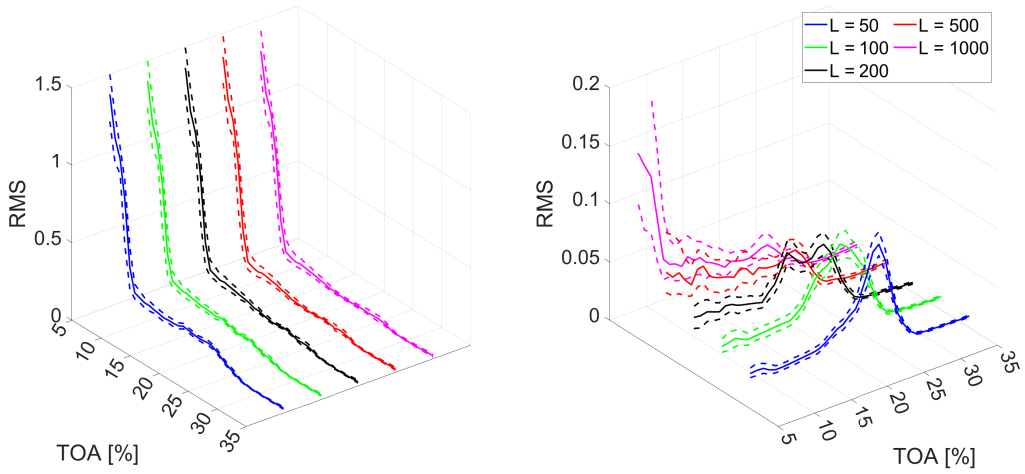


Figure 5.6: Relative RMS of RCs for p_{s-imp1} sensor before impeller for different window lengths; a) $L = 100$; b) $L = 200$; c) $L = 500$; d) $L = 1000$

Figure 5.7 demonstrates mean values and dispersion of RC 1 and RC 2 RMS for p_{s-imp1} sensor before impeller. For RC 1, there seems to be very little difference in between the window lengths concerning both the mean and dispersion. A slight difference can be observed in the locally unstable region (TOA 20%), where for $L = 50$ there seems to be an increase in the RMS value, while it is less significant or none for the longer windows. This is probably due to the signature of inlet recirculation being captured in RC 1 because of a short window length L.

For RC 2, the difference between the RMS for different window lengths is much more significant. For the shortest window, the increase of RMS in the locally unstable region is clear. By increasing the length of the window, the difference in RMS for the locally unstable range changes. For $L = 100$, it has the same shape but a lower magnitude than for $L = 50$. For $L = 200$, two peaks are observed and the RMS does not show its maximum where the highest intensity of locally unstable conditions was observed [9]. The RC 2 for $L = 500$ does not show a clear reaction to any of the unstable conditions, while for $L = 1000$, the RC 2 clearly captures the signature of global instability. This is related to the decomposition of the signal into more components, which makes the signature of a single instability to be more dispersed between the RCs.



(a) (b)

Figure 5.7: Mean and dispersion of RMS of selected RCs for p_{s-imp1} sensor before impeller for different window lengths L a) RC 1; b) RC 2; solid line represents the mean value, dashed line demonstrates the confidence interval of 90 %

Overall, the choice of window length has to be adapted to the instability to be captured. When using a long window, some of the features might be over-decomposed, meaning that their signature is shared between a number of components. This might result in the need for grouping of the components [91],

which makes the feature extraction procedure even more specific to a given case and introduced another source of variability. Extending the window length increases the decomposition time, which makes the system less responsive. Therefore, in principle, it should be long enough to sufficiently decompose a signal for a given purpose, but no longer to limit the acquisition and processing time.

5.2.3 Timing of the decomposition

In SSA decomposition, the timing is influenced by the input length and window length. To understand the window and signal length influence on timing limitations coming from the choice of window length, a timing study is performed. The decomposition of the same input signal is performed on a PC-class computer for a variable input signal length N_s in the range from 5,000 to 100,000 and window length L from 50 to 500.

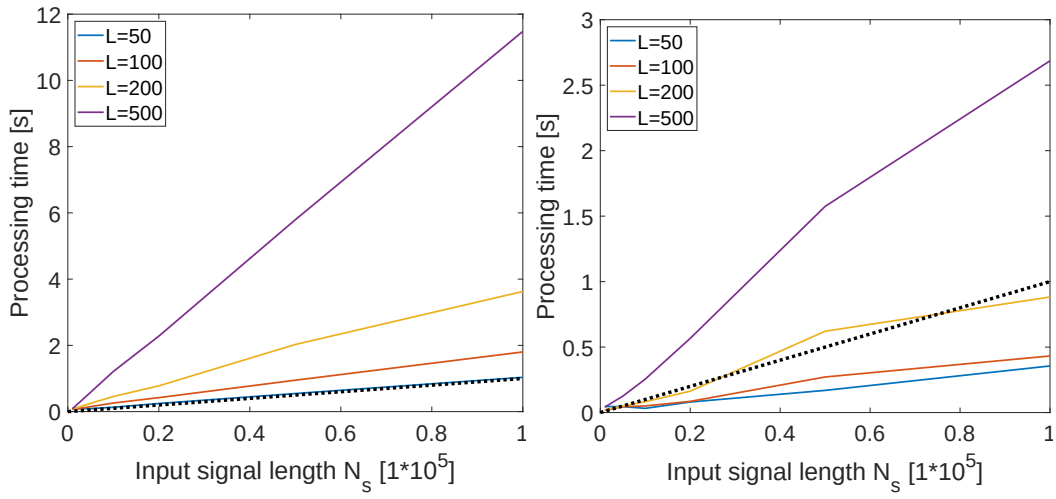


Figure 5.8: Time of SSA decomposition depending on the input signal length N_s and window length L for obtaining different number of RCs a) 15 RCs; b) 2 RCs; the dashed black line is an online processing limit above which the method no longer can provide real-time indication

Two scenarios regarding the number of RCs are investigated. In the first, the number of RCs extracted is limited to 15. In the second case it is limited

to 2, as they were sufficient to obtain a comprehensive indication of globally and locally unstable conditions. The advantage of SSA is that the components can be independently reconstructed, unlike for EMD where they have to be sequentially extracted from the lowest to the highest. The decomposition for each data point was performed ten times and the results for averaged timing are shown in Figure 5.8.

The time needed for processing almost linearly increases with the input signal length N_s for extracting 15 RCs. The line limiting the online decomposition is almost identical to the decomposition for $L = 50$. All other window lengths result in decomposition taking a longer time than the acquisition time. For the extraction of 2 RCs, the linear character of the change in timing is not observed. With a lower number of computations, the effects of background processes are not fully controlled on Windows machines, therefore they can affect the timing. Limiting the number of extracted RCs, the decomposition time is over 4 times shorter for each window length, allowing one to obtain an online decomposition while using a longer decomposition window.

The timing study in this thesis is used to understand the order of magnitude of timing and investigate the feasibility of online detection. SSA decomposition could possibly be accelerated when using different hardware and software, for example, based on FPGA architecture [202] Therefore, despite the limitation coming with the choice of a longer window, the windows above 200 are not excluded from further analysis.

5.3 General compressor stability identification

This section demonstrates how the SSA-based features can be used for quick and robust identification of general stability conditions. Three general states are defined, being *stable*, *locally unstable* and *globally unstable* *. The features

*The *italic* notation is used here to highlight the types of conditions investigated; this notation is omitted further in this chapter for clarity

indicative of each unstable condition are defined to enable the detection. Those features can be based on selected components or groups of components that best highlight both the presence and intensity of an instability. In this thesis, the focus is on finding specific RCs that hold clear information about instability and avoid the need for grouping of the components.

5.3.1 Locally unstable conditions

The first unstable condition appearing in the system when increasing throttling and decreasing TOA level is a local instability, defined to be inlet recirculation. Inlet recirculation for the initial SSA decomposition parameters was mostly reflected by an increase in RMS of RC 2 for sensor p_{s-imp1} before impeller, however, certain traces of this instability were observed for subsequent RCs as well as for selected RCs for the downstream sensor at impeller centre. Figure 5.9 demonstrates the RMS of RCs 2 to 6 for sensors p_{s-imp1} and p_{s-imp2}.

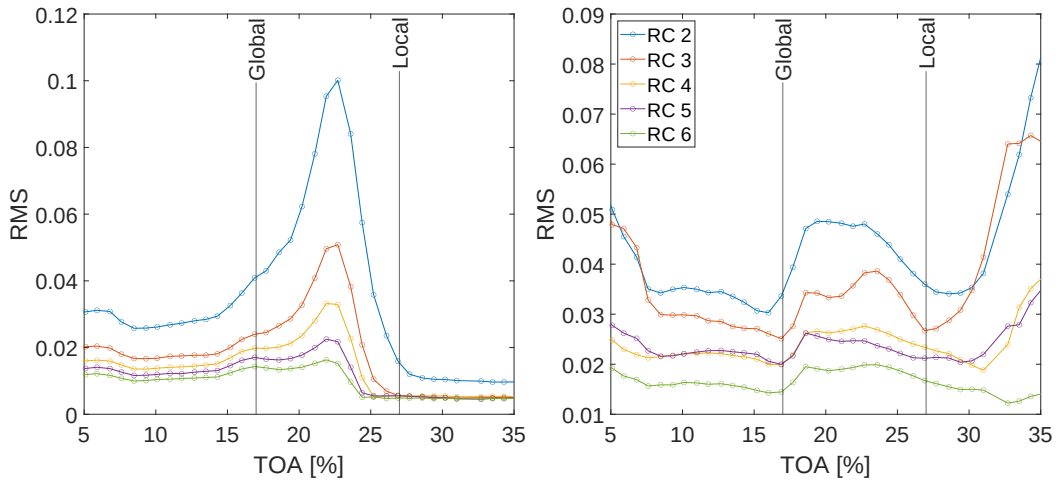


Figure 5.9: RMS of selected RCs for all operating conditions obtained for sensors a) p_{s-imp1}; b) p_{s-imp2}

The RMS of RCs 2 to 6 for p_{s-imp1} sensor is generally low and stable for normal compressor operation, increases for local instability and decreases for

the globally unstable region. The magnitude of the peak for the locally unstable conditions is lower for each consecutive RC. RCs from 3 to 6 have almost identical magnitude and the increase starts at ever higher TOAs. Each consecutive RC reacts to the presence of local instability for lower value of TOA, demonstrating an increase in its RMS value. The RMS of each RC for globally unstable conditions remains above the level established for stable operation. Since the magnitude of RC 2 is the highest in the locally unstable region and its increase coincides with the begining of the locally unstable zone, it might be considered a prospective indicator for the local instability presence.

The RMS of RCs 2 to 6 for p_{s-imp2} sensor demonstrates some potential for locally unstable conditions differentiation, but it is not as clear as for the sensor before impeller. In the stable region, the RMS of RCs from 2 to 5 seems to be related to the mass flow, which decreases with increasing throttling. An increase in RMS of RCs 2 and 3 can be seen in the region of locally unstable conditions, followed by a decrease at the beginning of the globally unstable region. Another increase can be observed for the end of the globally unstable region. As much as it might capture certain traces of inlet recirculation, the RCs from the sensor at impeller centre seem not to be exclusively tied to the local instability.

To consider a feature robust, it is important to evaluate its dispersion for varying conditions and understand its sensitivity. Figure 5.10 demonstrates the changes in the dispersion of RC 2 for p_{s-imp1} sensor before impeller for varying input signal length N_s. The dispersion of the data is overall low and similar throughout the whole range of operating conditions. The increase in input signal length results in a lower level of dispersion, however, the change is not very important and the inlet recirculation zone can still be distinguished from stable conditions even if the signal length $N_s = 1{,}000$. The presence of inlet recirculation brings far less risk to machine operation and can be sustained in many cases, hence decreasing the input length for the cost of robustness is not beneficial and the input length $N_s = 10{,}000$ is considered appropriate.

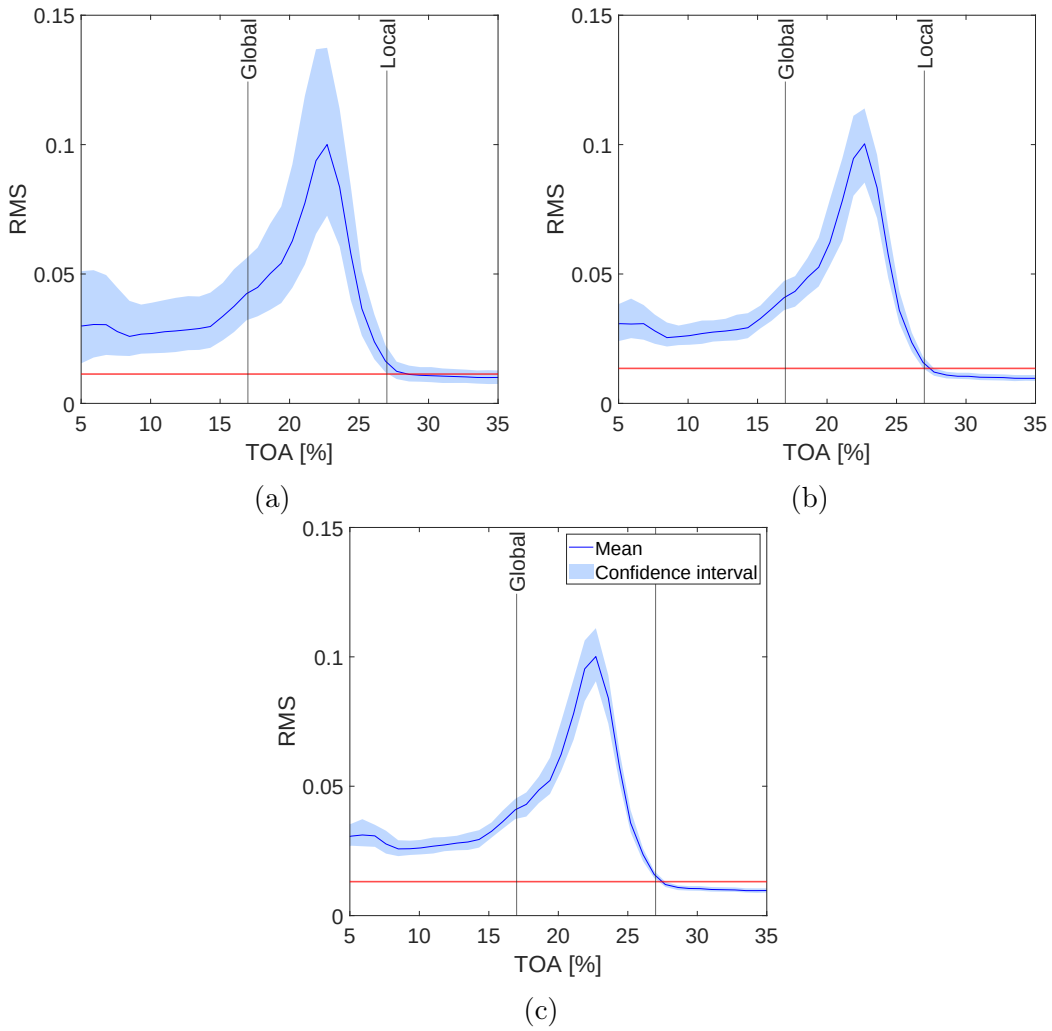


Figure 5.10: Mean and confidence interval of RC 2 obtained for sensor p_{s-imp1} for different input signal length a) N_s = 1,000; b) N_s = 5,000; c) N_s = 10,000; the red horizontal line represents a threshold defined as $Q(0.995)$ of data from compressor stable operation

To understand the physical character of the components holding the signature of inlet recirculation, their spectral content is investigated. It is obtained through the computation of a Fourier transform for the components obtained for independent inputs and averaging such spectra over 150 input signals for the same conditions. The results for RC 2 are presented in Figure 5.11. The spectral content of the RC 2 demonstrates a wide, single smooth peak centred at around 1000 Hz. This is in line with previously observed inlet recirculation

characteristics for this machine [9] and remains within the range defined for other machines [43]. The averaged spectral content of RC 3 has a similar shape, except for a peak resulting from blade passing frequency at 2300 Hz. The centre of the peak is shifted towards a higher central frequency of approximately 2000 Hz, where the inlet recirculation signature could also extend [9]. For TOA values of 30 and above, RC 3 displays also some peaks at higher frequencies, over 10,000 Hz. In the inlet recirculation zone, the increase in RMS at low frequencies is also present, which is not observed in other regions. The analysis of spectral content demonstrates that the content of components is highly data-driven and the same RCs can focus on different frequency ranges, depending on the conditions.

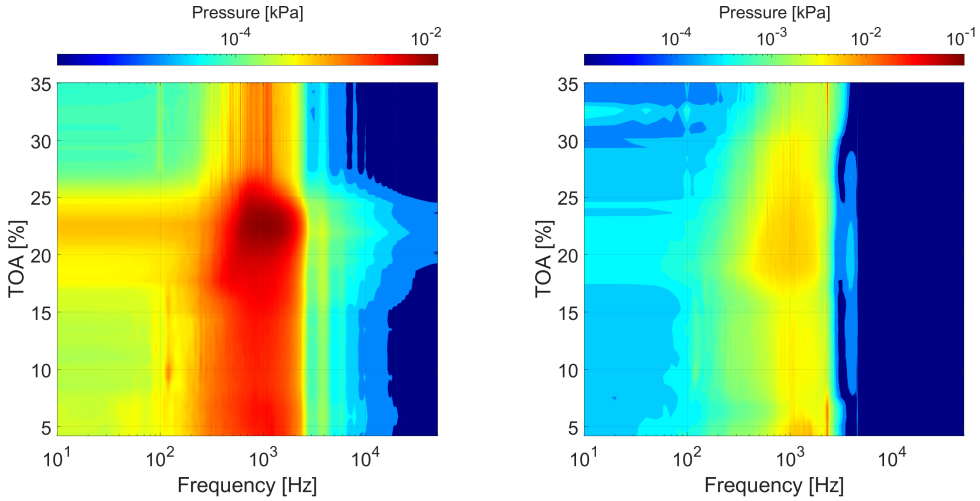


Figure 5.11: Contour of averaged spectral content of selected components of RC 2 obtained from different sensors a) p_{s-imp1}; b) p_{s-imp2}

Figure 5.12 demonstrates the accuracy of detection for varying window length and number of RC. The accuracy is defined as the number of correct indications of conditions divided by the total number of data points. The correct indication is assumed to be below the threshold for stable conditions and above the threshold for both locally and globally unstable conditions. It can be noted that using the shortest window length $L = 50$, a very accurate detection

can be obtained with both RC 1 and RC 2, reaching almost 100 %. Increasing the window length does not result in higher accuracy for neither of RCs, but more components tend to provide accurate indication. It is related to the signature of local instability being shifted to higher RCs and shared with more RCs. With no increase in accuracy and an evident increase in the time needed for the decomposition, extension of the window past $L = 50$ is not recommended in this particular case.

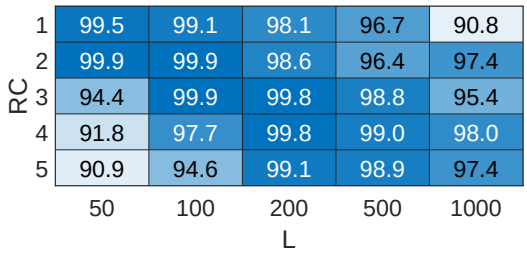


Figure 5.12: Accuracy of local instability detection with selected RCs obtained for different window lengths

Considering both the mean values of RMS for different conditions as well as their dispersion, it is possible to define a feature that can robustly capture the presence of the locally unstable conditions. **This feature μ_L is designated as RMS of RC 2 obtained from sensor p_{s-imp1} before impeller for input signal length $N_s = 10{,}000$ and window length $L = 50$.**

5.3.2 Globally unstable conditions

With a further increase in throttling past the locally unstable conditions, the compressor enters the globally unstable regime [9]. Figure 5.13 shows the behaviour of the RMS of RC 1, both in the global range and in a limited range to better visualise its behaviour close to the boundary. The overall character of RC 1 is similar across the sensors, but the increase related to the onset of instabilities begins at different TOAs. It is the earliest for p_{s-imp1} and p_{s-imp2}, which seem to indicate the influence of locally unstable conditions. A delayed reaction is observed for p_{s-imp3} sensor and p_{s-out} sensor. They start to increase

in the middle of the locally unstable range, indicating that the signature of global instability starts to be present earlier for those sensors. The latest reaction is observed for p_{s-in}. The increase is exactly in line with the boundary of a globally unstable range [9]. The RMS of RC 1 for sensors at the inlet and outlet are prospective indicators and they are investigated in more detail.

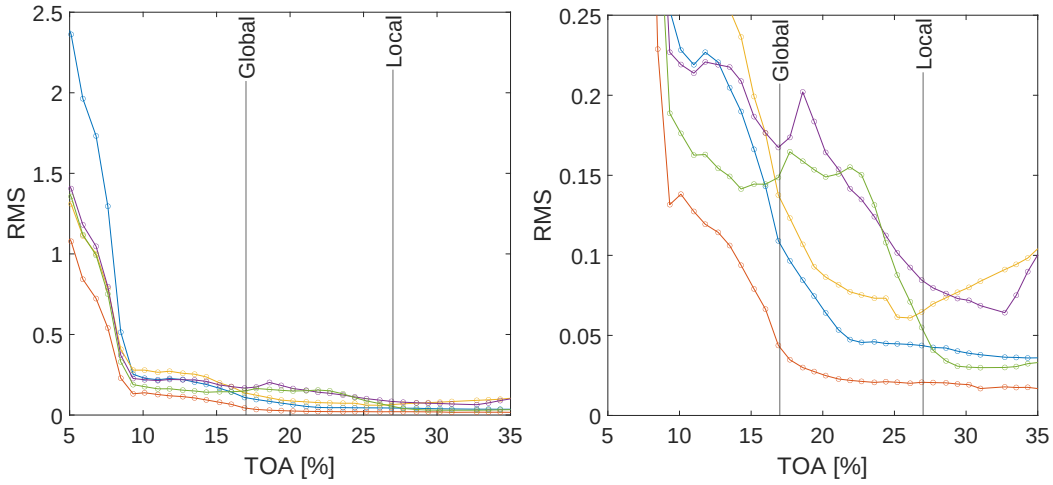
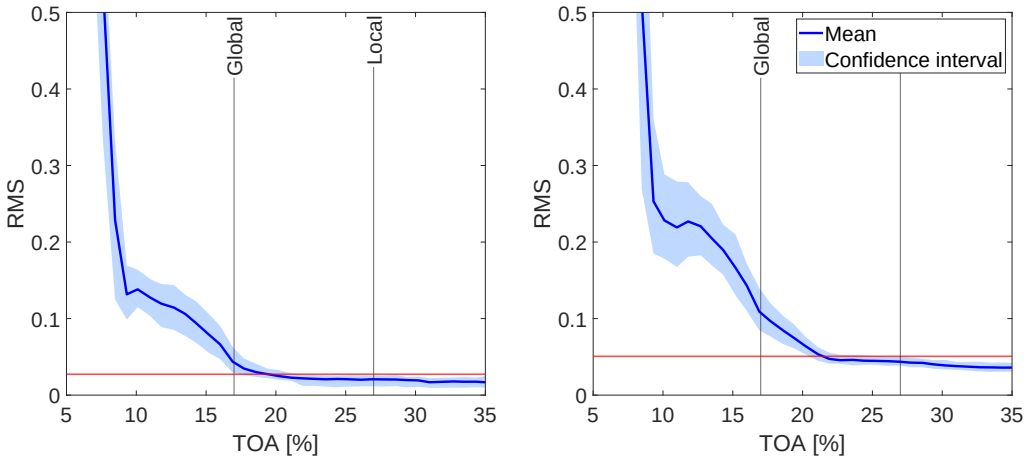


Figure 5.13: RMS of RC 1 for all sensors and conditions a) full range; b) limited range to highlight the changes in locally unstable region

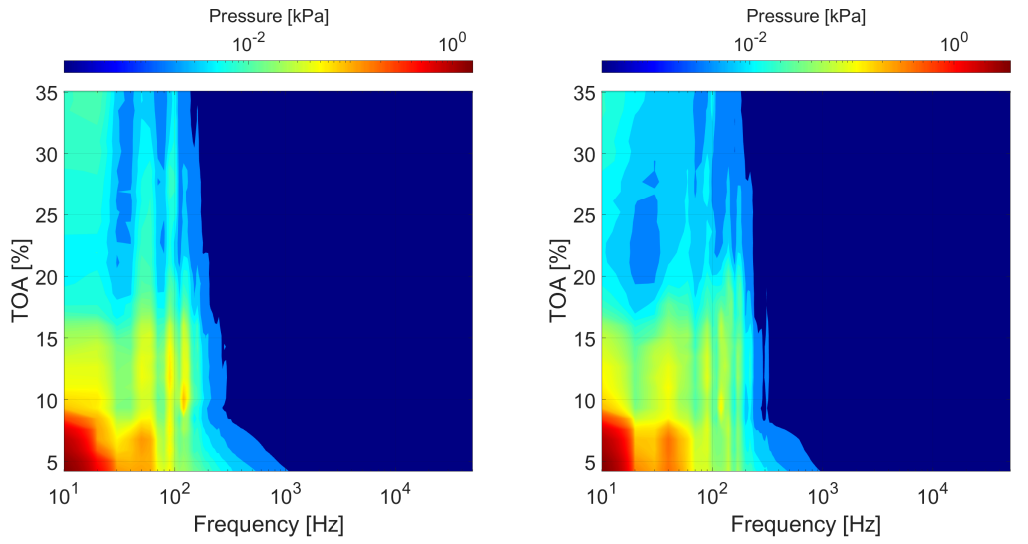
Figure 5.14 demonstrates the mean and dispersion of the RC 1 obtained from p_{s-in} and p_{s-out} sensors. The plot also includes the threshold obtained from stable conditions for each sensor. The dispersion is low for the stable region and increases with the onset of unstable conditions. The confidence interval follows the changes in the mean and the dispersion of the data does not hinder the possibility of globally unstable conditions detection. To investigate the physical meaning of the RCs and the reason for their different behaviour, their spectrum is investigated. Figure 5.15 demonstrates a spectral content of RC 1 for inlet and outlet sensors. The overall spectral content of RCs is very similar in its range. A visible increase in amplitude is observed for low TOA values for both sensors, however, it starts earlier for the outlet sensor. Therefore, the earlier

increase in RMS of RC 1 for the outlet sensor comes from the physical character of the signal registered at that location.



(a) (b)

Figure 5.14: Mean and confidence interval of RMS of RC 1 obtained for window length $L = 50$ from selected sensors a) p_{s-in} sensor at the inlet; b) p_{s-out} sensor at the outlet; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation



(a) (b)

Figure 5.15: Contour of the mean frequency spectrum of RC 1 for all operating conditions for selected sensors and window lengths a) p_{s-in}, $L = 50$; b) p_{s-out}, $L = 50$;

The shift in the reaction to globally unstable conditions can be dependent

on the selected window length. The changes in RC content due to variations in the decomposition window length are not trivial. To investigate how first RCs for different selection of the window length could be used for detection, the accuracy of prospective features is investigated.

The accuracy is defined as a ratio of correct indications to all indications made in the whole operating range. A correct indication is assumed to be over the threshold for the globally unstable range and below the threshold elsewhere. The results are shown in Figure 5.16 for both inlet and outlet. For the base window length, the accuracy of detection for inlet sensor is higher than for the outlet. It is caused by a too early increase in feature value for the outlet sensor compared to the data labels, resulting in errors in the locally unstable range. When increasing the window length, more RCs perform well in detecting globally unstable conditions. This is due to the signature of global instability being shared by those components.

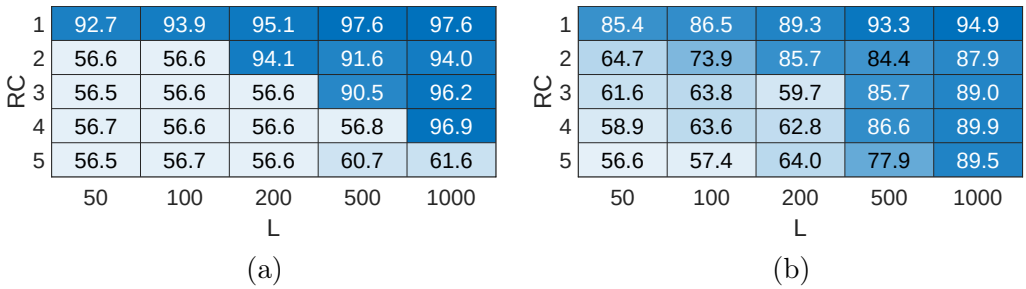


Figure 5.16: Accuracy of indication based on threshold derived from stable conditions for a) p_{s-in}; b) p_{s-out}

The accuracy of indication for both sensors increases with an increase in window length, as then the number of misclassifications in the locally unstable range decreases (see Figure 5.21). The extension of window length results in the RMS increase at higher value of TOA. The increase in accuracy is at about 5% for the inlet sensor, but it comes at the cost of increased decomposition time. The errors affecting the accuracy are made near the defined stability line. Since the RMS of RC 1 is related to the intensity of instability, such errors does not

have to be misclassifications, which is discussed in Section 6.2.2. Due to the overall high accuracy and short decomposition time, the shortest decomposition window is considered appropriate.

For further investigation, an RC 1 obtained for window length $L = 50$ and sensor p_{s-in} is designated μ_G feature, to be used for detection of globally unstable conditions.

5.3.3 Feature space representation

Using the features μ_L and μ_G, it is possible to define a feature space related to SSA-based features and project the compressor data onto it. Figure 5.17 demonstrates a projection of the data for stable operation, locally unstable conditions and globally unstable conditions. The thresholds for T_{μ_L} and T_{μ_G}, computed as $Q(0.995)$ of data from compressor stable operation, are plotted as horizontal and vertical lines respectively. The plot axes are logarithmic to increase the readability of the data.

The points for different conditions occupy different locations in the feature space. Stable points form two clusters in the low left corner of the space. The clusters have their means shifted in the horizontal direction. For the local instability, some sub-clusters are observed. The clusters have their means shifted in both horizontal and vertical directions. The vertical shift can be associated with data obtained from separate TOAs, for which the value of μ_L feature differs. Along with limited dispersion, the points for subsequent TOAs do not overlap in the feature space, forming separate clusters. For each such sub-cluster, a smaller cluster offset to the left of the feature space can be observed. The reason for the presence of a smaller, shifted cluster is explored in Chapter 6.

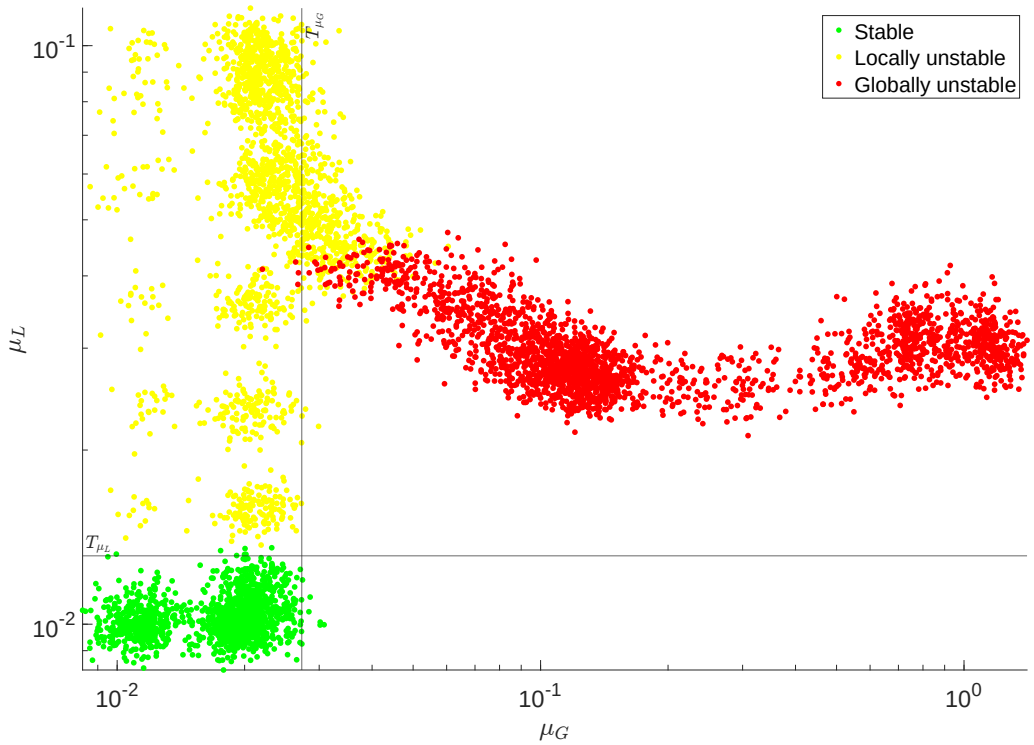


Figure 5.17: Feature space representation of compressor data using μ_L and μ_G features with thresholds T_{μ_L} and T_{μ_G} defined based on stable operating conditions

The points for globally unstable conditions are also more scattered than those for stable operation, extending to the right in the feature space. Considering the logarithmic scale of the plot, their extent is much larger than that of stable data. The clusters for individual TOA values are not observed for globally unstable conditions. With a much higher dispersion, the difference in mean value is not significant enough to demonstrate separate clusters. However, two large sub-clusters can be seen, where the first cluster centre is located at μ_G of around 0.1, while the other at μ_G of 1. The presence of those clusters reflects two different modes included in the globally unstable conditions.

The feature space representation and thresholds derived from the stable operation can be used for the classification of the unknown conditions. To define current stability, two thresholds have to be used. Being over the first threshold T_{μ_L}, the data is considered to come from unstable operation. The second

threshold T_{μ_G} defines a type of instability. If the point is below, it is a local instability. If it is above, the instability is considered global. The thresholds T_{μ_L} and T_{μ_G} derived from stable operation represent the data well and provide good differentiation between all three states. To obtain the thresholds, only the data from stable operation is required. It is considered an advantage because it limits the risk associated with driving the compressor towards an unstable operation. However, it can be noted that some points are misclassified in the region where locally unstable conditions transition into globally unstable. The detailed study of the accuracy of the classification with an SSA-based feature space is demonstrated in Chapter 6.

5.4 Exact instabilities detection for increased accuracy

Using features based on the components from SSA decomposition, it is possible to differentiate the general conditions of the compressor using thresholds derived from stable data. The globally unstable conditions can be furtherly divided into transient and deep surge [9]. The aim of this section is to investigate the possibility of defining SSA-based features capable of robust differentiation of those two states. Such a possibility could increase the flexibility and accuracy of the instabilities detection system.

5.4.1 Transient and deep surge features

Transient conditions are a part of the globally unstable range. When increasing the level of throttling, they are present after inlet recirculation and prior to deep surge. Thus, detection of transient conditions can be taken advantage of to predict deep surge onset, while maximally extending the operating range of the compressor. With SSA decomposition using window length $L = 50$ (Figure 5.2), there is no component reacting solely to transient conditions. Investigating the RCs in more detail, a change in the behaviour of RCs from 2 to 4 for sensors

p_{s-imp2} at the impeller centre can be seen, as well as a local increase in RMS for RC 4 for sensors p_{s-imp3} at the impeller end. The transient conditions are also marked with an increase in RMS of RC 1 for all sensors, but not as clear or important as for deep surge. With RC 1 holding both, the signature of transient and deep surge, it should be furtherly decomposed in order to isolate those signatures. It is expected that increasing the window length L, a part of the information contained in RC 1 will be distributed to subsequent components and possibly the modes of transient and deep surge could be separated.

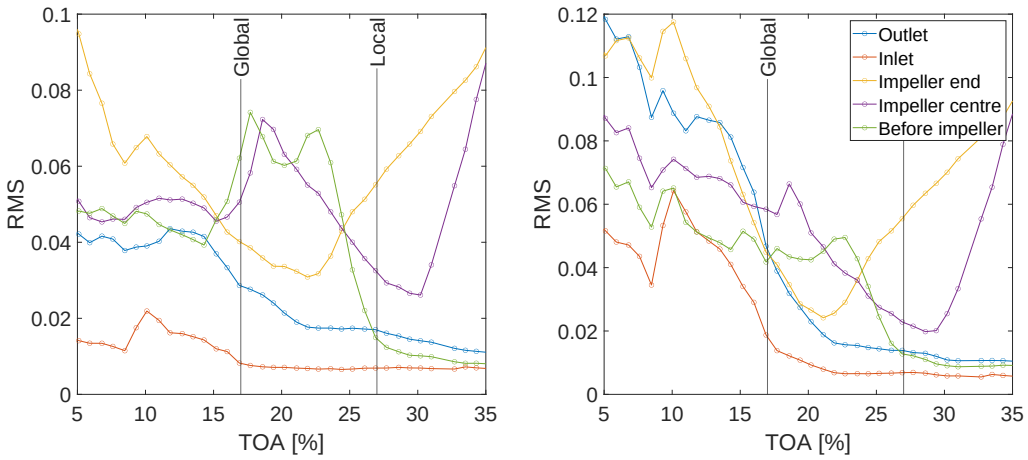
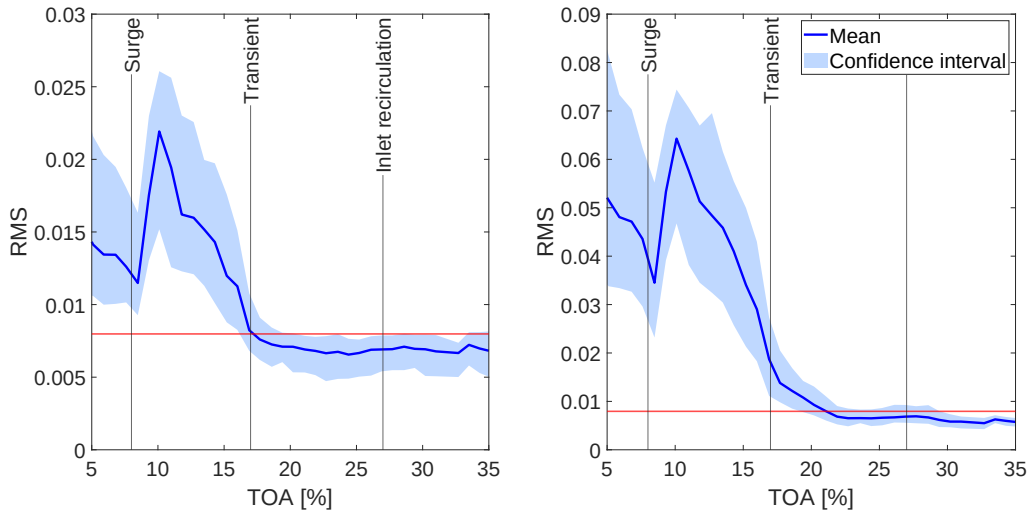


Figure 5.18: RMS of RC 2 for all sensors and all operating conditions for different window lengths; a) $L = 200$; b) $L = 500$;

Figure 5.18 demonstrates the RMS of RC 2 for the window lengths $L = 200$ and 500. For $L = 200$, it has a constant value in the stable region, followed by an increase in the transient region. The increase at the outlet starts for a higher TOA value and does not demonstrate any distinct behaviour for the transient region. The RMS for the inlet sensor starts to grow at the edge of the transient region, peaks there and drops prior to the onset of deep surge, where it remains stable. With a longer decomposition window, the behaviour is very similar. The peak in the transient region has a higher magnitude and the component shows an increase in the deep surge region instead of stagnation. Offering the clearest

indication of the transient region, the sensor at the inlet will be considered in detail. For either window length, the RMS level in the surge region remains above the threshold that could be obtained with stable data.

Investigating the dispersion of the data, decomposition for $L = 200$ and $L = 500$ is considered and presented in Figure 5.19. As much as the mean value should allow isolating the transient region from stable conditions, the dispersion is high, both for stable and transient conditions. Therefore, robust detection for those conditions is not possible.



(a) $L = 200$ (b) $L = 500$

Figure 5.19: RMS of RC 2 for all sensors and all operating conditions for different window lengths; a) $L = 200$; b) $L = 500$; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation

The change in RMS of RC 2 obtained for $L = 500$ at transient conditions indicates that it captures signatures of this instability. However, it is not possible to define a threshold based on stable conditions that could differentiate transient from deep surge based on this feature. The RMS level for RC 2 in the deep surge region does not drop below the threshold defined with stable data. It might imply that the extracted feature is not fully representative of the physics of transient compressor behaviour or that the transient behaviour is

still present at deep surge. To gain a deeper understanding of the system, the spectral content of RC 2 is investigated.

Figure 5.20 shows mean spectral content for RC 2 obtained with $L = 200$ and $L = 500$. The transient signature is associated with the presence of three frequency bands centred at around 60, 90 and 105 Hz. Those bands are amplified over the background level for most of the operating range and get furtherly amplified for the transient region. The amplification of low frequencies also takes place, especially for the deep surge region. This results in an increased level of energy for that region, despite a lower intensity of the phenomena associated with transient operation.

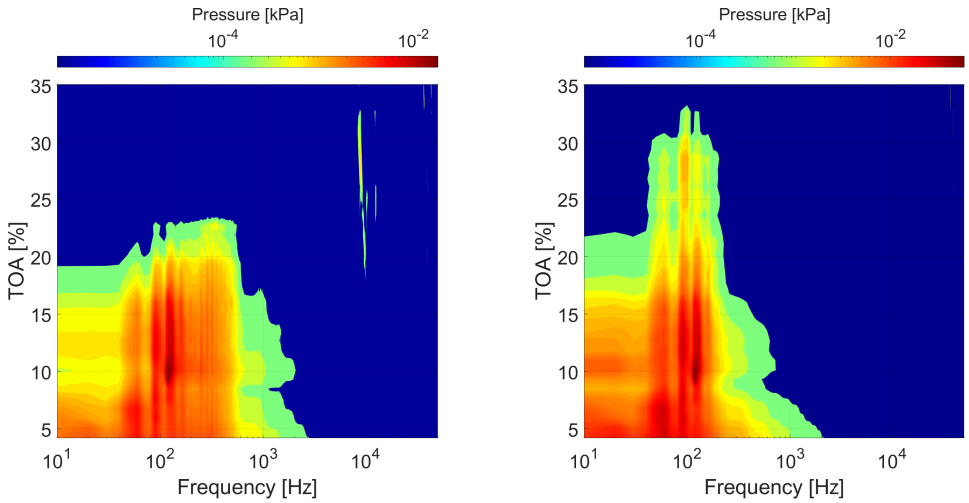


Figure 5.20: Contour of mean frequency spectrum of RC 2 for all operating conditions for p_{s-in} sensor and different window lengths a) $L = 200$; b) $L = 500$;

With RC 2 having a value above the benchmark for both transient and deep surge region, a feature sensitive only to deep surge should be seeking to make a differentiation between the two zones. It is expected that such a feature would be based on RC 1, as surge dominates the signal characteristics. However, investigating RC 1 for various window lengths (Figure 5.21), the increase in their RMS above the benchmark level starts for the transient region. The rise

in RMS level for the transient region is observed for all window lengths. For higher length values, it is less significant, but still results in a feature value above the threshold. Therefore, it is not possible to fully isolate the signature of deep surge with this feature. Investigating the spectral content of RC 1 for different window lengths (Figure 5.22), it can be noted that the presence of the transient signature is limited when extending the window length. However, the increase in energy for low frequencies, below 30 Hz starts in the transient region. Therefore, should a window length be furtherly increased to remove the presence of transient conditions peaks located around 100 Hz, the feature based on RC 1 would still increase above the stable threshold in the transient region. Hence, defining a feature differentiating deep surge from transient conditions based on a threshold derived from stable data may not be possible.

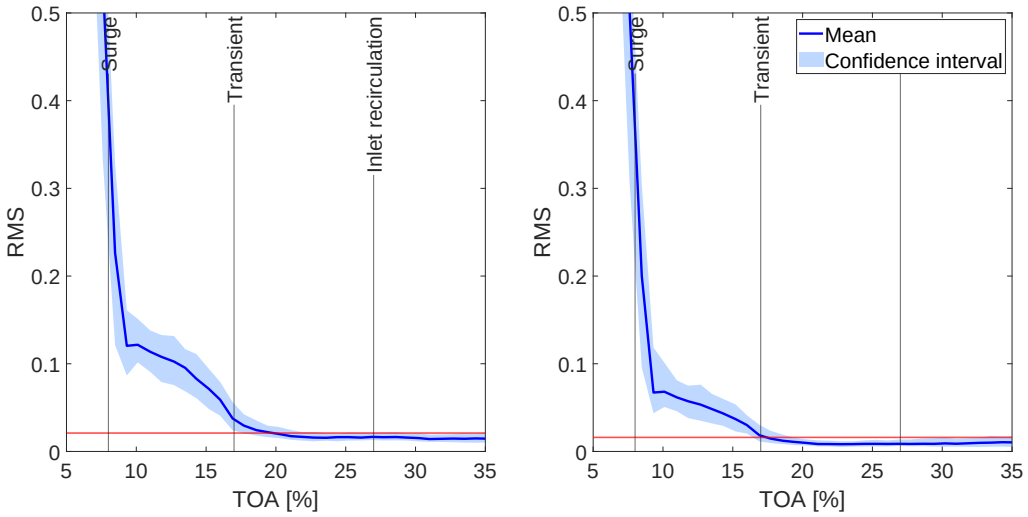


Figure 5.21: Mean and confidence interval od RMS of RC 1 obtained for selected sensors with different window length a) p_{s-in}, $L = 200$; b) p_{s-in}, $L = 1000$; the red line is a threshold defined as $Q(0.995)$ of data from compressor stable operation

Concluding, the differentiation of transient and deep surge conditions based on the threshold defined using stable conditions is not possible with SSA-based

features. It is possible to define features sensitive to those instabilities, however they are not capable of fully separating their signature. To use them in a threshold-based detection system, the values of thresholds have to be set using data from unstable conditions. What is more, the cost of obtaining those features is high, as they require using a long decomposition window. The differentiation of transient and deep surge conditions may be possible with a global instability feature μ_G. In this context, the data from unstable compressor operation has to be obtained, but a short window length can be used to ensure system responsiveness.

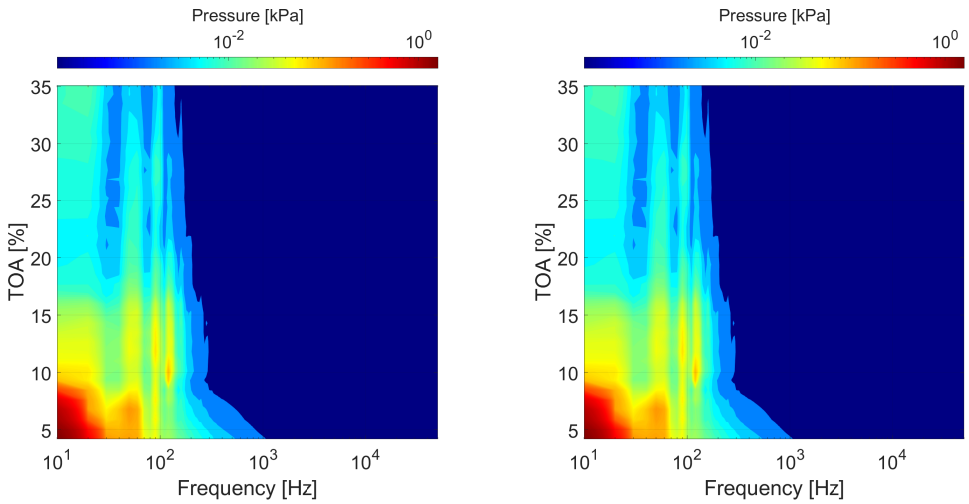


Figure 5.22: Contour of mean frequency spectrum of RC 1 for all operating conditions for p_{s-in} sensor and different window lengths a) $L = 200$; b) $L = 1000$

5.4.2 Feature space representation

The feature μ_G, based on RC 1 incorporates a signature of both, transient and deep surge conditions. The presence of two regimes is visibly distinguishable, as per Figure 5.14. Therefore, it is possible to take advantage of μ_G for the differentiation of transient and deep surge conditions. The differentiation is no longer based on a threshold derived from stable conditions. Figure 5.23 demonstrates the mapping of the feature space using μ_L and μ_G features and

compressor data. The labels of the points in the figure are based on the exact instabilities observed [9]. The points with a black border represent the values obtained for peaks of instabilities, while transparent points demonstrate the remaining data.

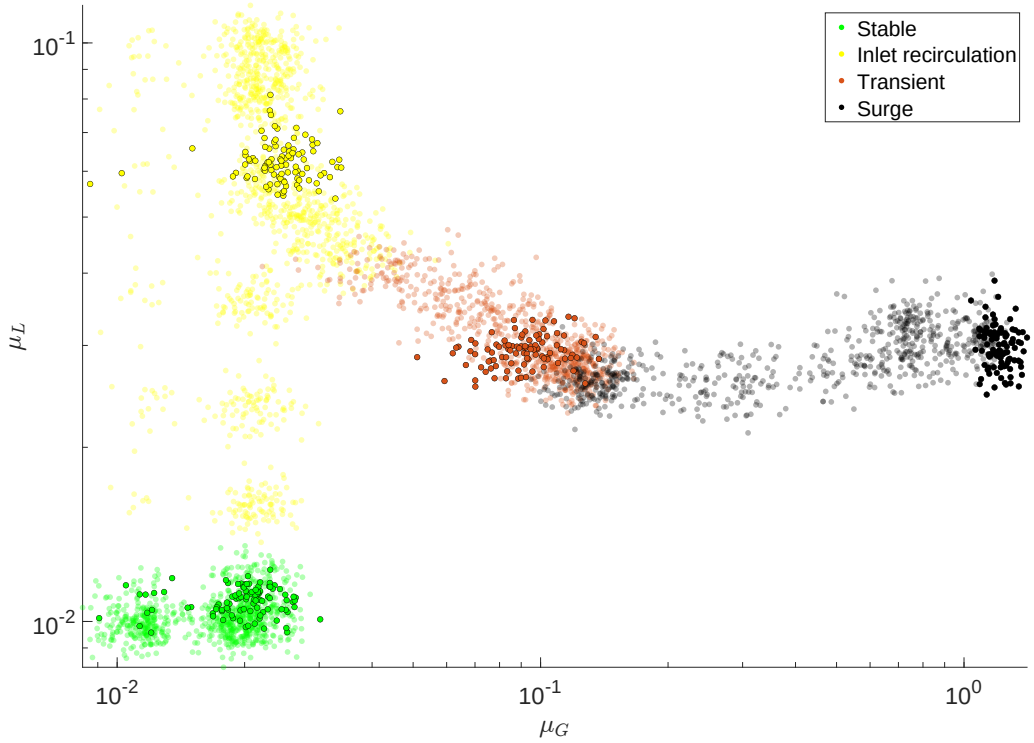


Figure 5.23: Feature space representation of compressor data using μ_L and μ_G features to capture exact instabilities present in the compressor system; scatter points with black edge represent peaks of conditions, transparent points represent all data

Overall, the clusters for peaks of conditions are well separated in the feature space. The data for different conditions also generally occupy different locations in the feature space, although those regions tend to overlap at some locations. The part of transient points overlaps with deep surge data for a region where a transition between those conditions takes place. A similar overlap can be observed between inlet recirculation and transient conditions, suggesting that the change in stability is gradual when the TOA changes are small.

With the presented data, it is even more clear that the points form a path in the feature space for changing TOA value. The features defined based on RCs are interpretable, hence the location of a point in the feature space should be related to the intensity of an instability. This feature space representation can be used for differentiation of the exact instabilities present, however, it is necessary to obtain a mapping of the unstable conditions first. This might be disadvantageous compared to using thresholds derived from stable operation.

Overall, the feature space representation based on local and global instability features allows for good differentiation of the conditions, including separating transient and deep surge. A potential for improvement lies in addressing the regions of the conditions overlap in the feature space. It can be attempted with a probabilistic approach, which is demonstrated in Chapter 7.

5.5 Summary, discussion and conclusions

This chapter demonstrated the possibility of defining SSA-based features that can be used for quick and robust detection of the operating conditions of the compressor. Using MATLAB implementation of SSA algorithm and a PC-class computer, it is possible to provide an online detection of the general operating conditions of a compressor. The robustness of indication, related to the dispersion of the data for given operating conditions, is affected by the input signal length N_s and decomposition window length. The longer input signal decreases the dispersion of the data. Increasing the window length changes the repartition of the data between the components and the influence on the confidence interval of features is not trivial. If the decomposition window is too long, it might be possible that the components have to be grouped to extract a meaningful physical signature. The need for grouping makes the selection of the components more difficult and adds variability to the system, making it less general. The

extension of window length importantly affects the time needed for the decomposition, hence the window length should be sufficiently long to decompose the data properly, but not longer to limit the acquisition and processing time.

The features defined based on selected RCs are sensitive to locally unstable and globally unstable conditions. Each feature is based on a single RC to avoid the components grouping after the decomposition. The signature of interest is held within the two first RCs, but obtained from different sensors. The global instability feature is based on RC 1 for the p_{s-in} sensor at the inlet, while the local instability feature is based on RC 2 from the p_{s-imp1} sensor before impeller. Both features are obtained for the input signal length $N_s = 10{,}000$ and window length $L = 50$. Such a combination of decomposition parameters and components allows obtaining physically meaningful features that can capture and be used to differentiate general operating conditions.

Globally unstable conditions can be furtherly differentiated into transient and deep surge. Detecting the exact instabilities might be useful when the aim is to maximally extend the operating range of the machine. It is possible to define SSA-based features sensitive to transient conditions. However, such a feature requires increasing the window length, extending the time needed for the decomposition. What is more, differentiation of the transient conditions based on a threshold derived from stable conditions is not possible. The feature level grows importantly in the transient range, but remains above the threshold for deep surge conditions. A similar observation can be made when aiming for a feature extracting the pure signature of deep surge. Its value surpasses the stable benchmark in the transient zone, followed by an important increase in the deep surge region. Consequently, using the threshold defined solely on stable data is not possible. What is more, extracting such features requires significant extension of the window length, negatively affecting the pace of the detection.

An alternative approach to exact instabilities detection is proposed, based on the general feature space approach. Based on the features of local instability

μ_L and global instability μ_G is possible to define the stability of the system, taking advantage of the values of the features. The thresholds derived from stable conditions do not differentiate the transient and deep surge. Therefore, obtaining the data for unstable conditions is required to differentiate transient and deep surge.

The quantification of the accuracy of the SSA-based approach and a detailed discussion of its advantages and disadvantages is presented in Chapter 6.

6 Comparison of EMD and SSA for instabilities detection

In this chapter, the instabilities-sensitive features based on EMD and SSA are compared. Firstly, the features capturing general operating conditions are compared to explore their similarities and differences. Subsequently, the models built with quasi-dynamic data are validated with dynamic data to investigate the representativeness of features obtained using data from different experimental protocols. Next, the accuracy of both methods is investigated and the regions of misclassifications are analysed in detail to understand the reasons for incorrect classification. The chapter is finished with a summary of the methods performance, as well as identification of strengths and threats for implementation to a real-time instabilities detection system.

6.1 Comparison of instabilities-sensitive features

Using both EMD and SSA, features for local and global instabilities were defined, that proved to operate best for the detection of different instabilities. The features were termed μ_L and μ_G respectively and were used for the detection of local and global unstable operating conditions, but the value of μ_G could also be used for differentiation of transient and surge conditions. With the similar performance of features obtained with both methods, a detailed comparison is presented in this section. The goal it to understand the physical characteristics and differences between the features. Such analysis can help to understand how the features could be used for monitoring other compressors.

6.1.1 Local instability feature μ_L

Figure 6.1 demonstrates the changes in local instability features for two methods obtained using EMD and SSA. The behaviour of local instability features μ_L obtained with both methods is very similar throughout the whole investigated range. Their value is low and stable for stable operating conditions, starts to increase in the locally unstable region and falls down for globally unstable conditions. The magnitude of the EMD-based feature is higher for the whole operating range, suggesting that both features are not identical when it comes to their content. The dispersion is higher for EMD-based features, especially in the unstable region.

To investigate the character of the features, their averaged spectral content for the entire operating range is plotted in Figure 6.2. The character of the spectrum is very similar, with a peak located at a frequency of around 1000 Hz. The centre of the peak for EMD is at a slightly lower frequency than for SSA. In the locally unstable conditions zone, the amplitude of lower frequencies is also increased. Frequencies higher than 10,000 Hz do not seem to significantly contribute to the spectrum of the feature.

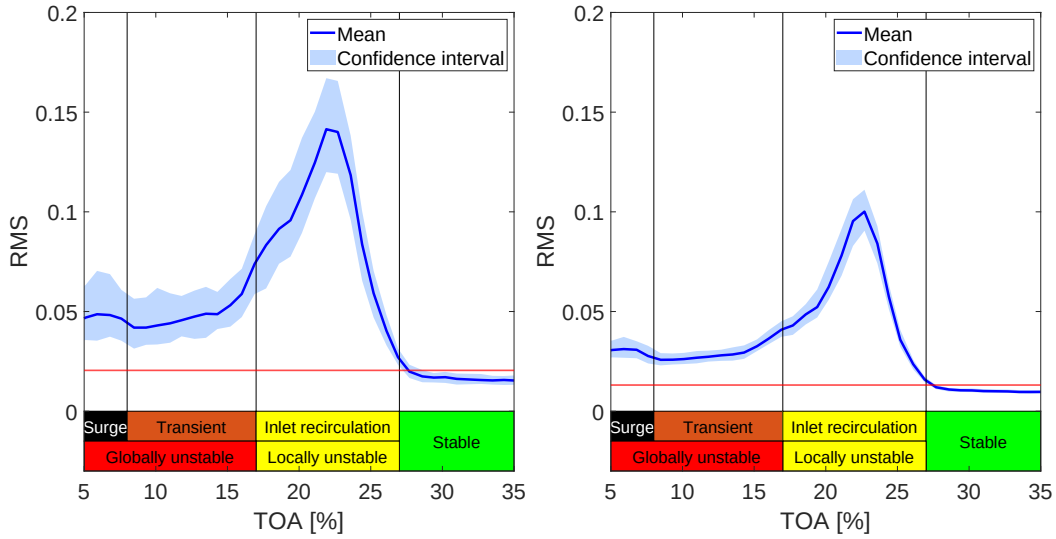


Figure 6.1: Features sensitive to locally unstable conditions obtained with different decomposition methods a) EMD; b) SSA; the red horizontal line represents a threshold defined as $Q(0.995)$ of data from compressor stable operation

The spectra for selected conditions are plotted in Figure 6.3. They are almost identical for both methods, with the spectrum for SSA showing less oscillations around the mean value. The difference in spectra can be observed for low frequencies, where the spectra for SSA have a very similar amplitude, while there is a higher difference in value between conditions for EMD.

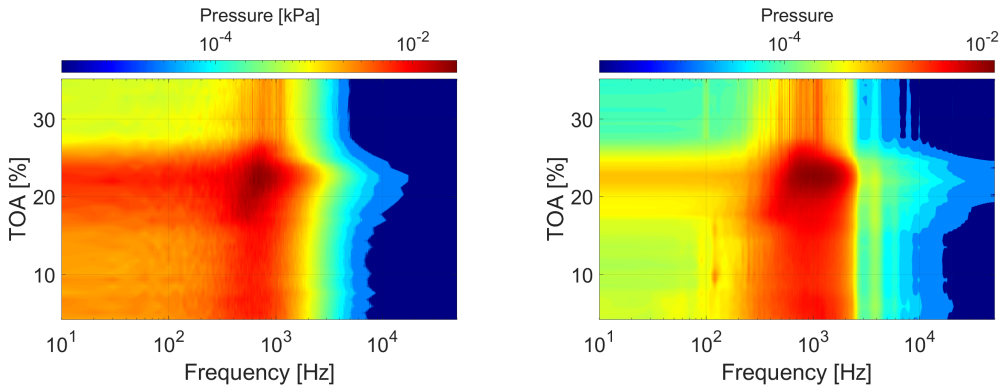


Figure 6.2: Contour of an averaged spectral content of μ_L feature obtained with different decomposition methods a) EMD; b) SSA

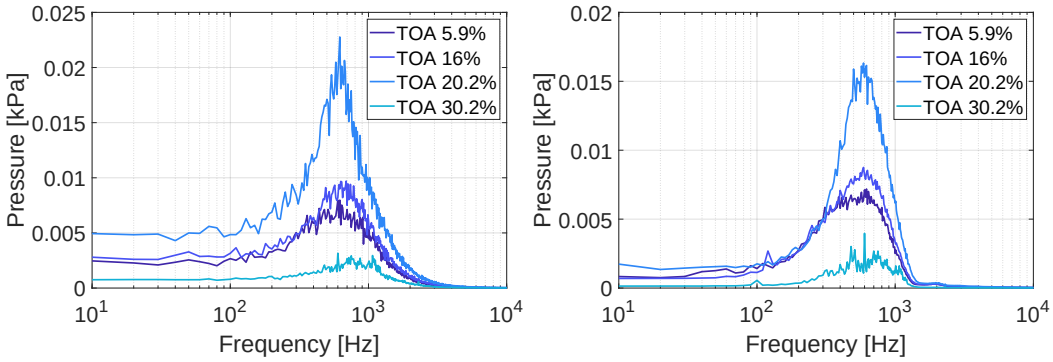


Figure 6.3: Averaged spectral content of μ_L feature for selected operating conditions, obtained with different decomposition methods a) EMD; b) SSA

Both extracted features are very similar in their character and seem to represent well the presence of local instability. The SSA-based feature seems to have a slight advantage due to a lower confidence interval.

6.1.2 Global instability feature μ_G

The global instability features based on EMD and SSA components are shown in Figure 6.4. Similarly to the local instability feature, they exhibit almost identical behaviour throughout the entire operating range. The absolute value of the feature is lower for SSA. The confidence interval for SSA feature is also narrower, mostly in the globally unstable range. Both features work well for the detection of globally unstable conditions and demonstrate an important change in value and the slope for the deep surge conditions.

Figure 6.5 demonstrates the averaged frequency spectrum for both of the features in the whole operating range. For EMD, a high level of energy is observed for frequencies below 100 Hz. With an increasing level of instability, they grow in magnitude, which translates into an increased energy level observed in Figure 6.4. The spectrum is visibly scattered for most of the operating range, with local decreases in the amplitude observed in a wide range of frequencies.

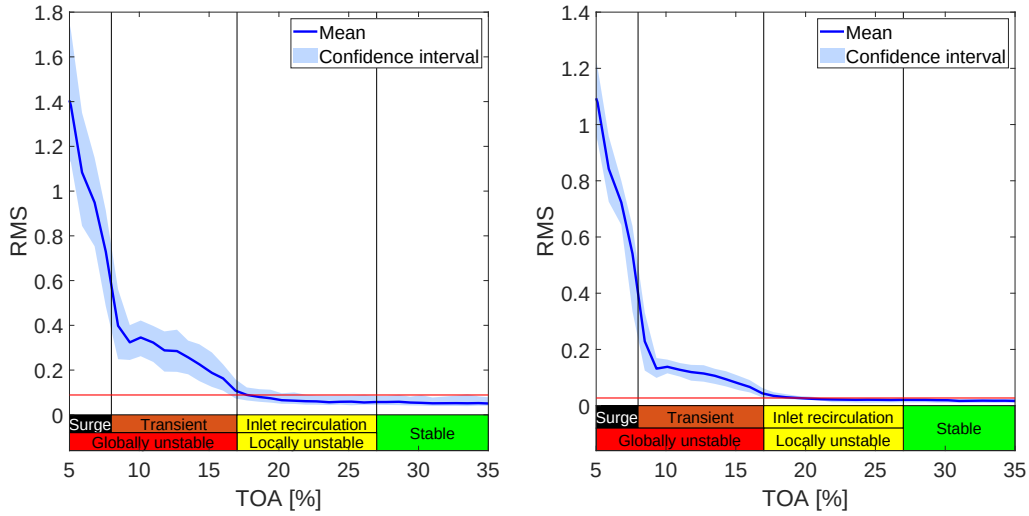


Figure 6.4: Features sensitive to globally unstable conditions obtained with different decomposition methods a) EMD; b) SSA; the red horizontal line represents a threshold defined as $Q(0.995)$ of data from compressor stable operation

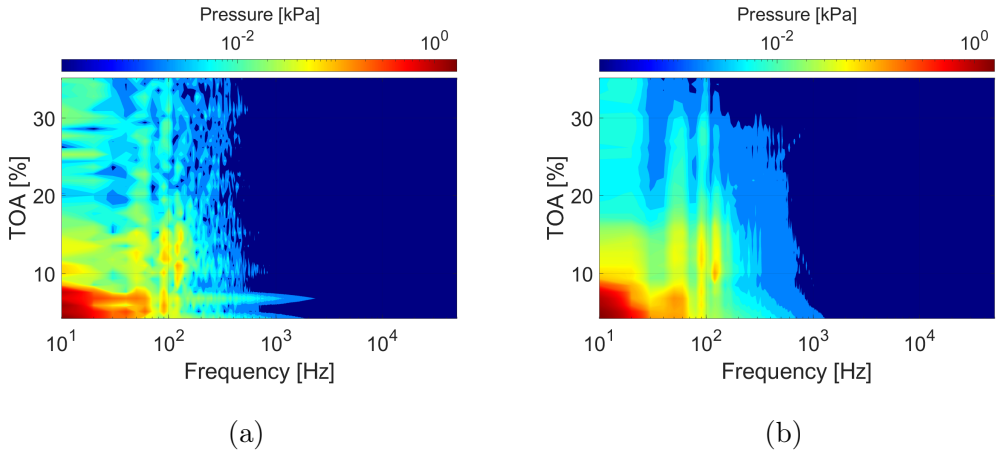


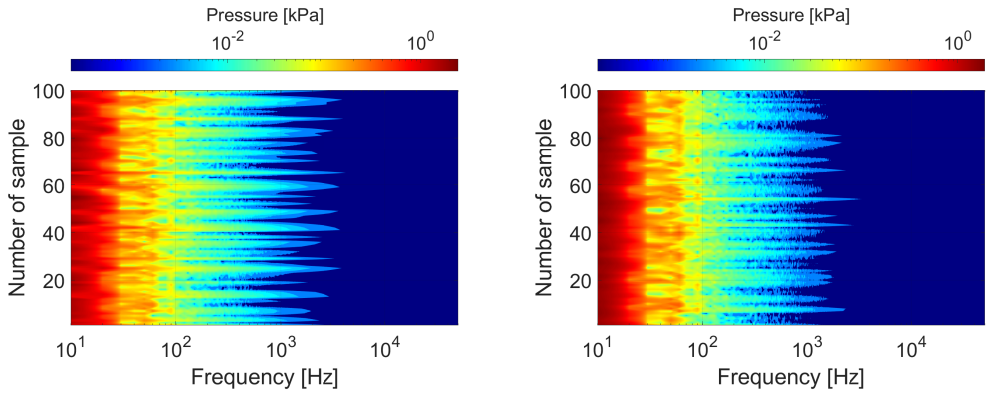
Figure 6.5: Contour of an averaged spectral content of μ_G feature obtained with different decomposition methods a) EMD; b) SSA

For SSA, the spectrum is more consistent and less variations between TOAs are present. At stable conditions, an increased energy level is present for the lowest frequencies as well as three distinct ranges - around 60 Hz, 90 Hz and 110 Hz. In the locally unstable conditions, the amplitude of those frequencies is increased. For transient operation, a further amplification is present and the

bands of frequencies get wider. The onset of transient conditions also increases the level of energy for the lowest frequencies. A visible change is present for the TOA of around 15 %. For deep surge, the low frequencies get furtherly amplified in the range below 30 Hz.

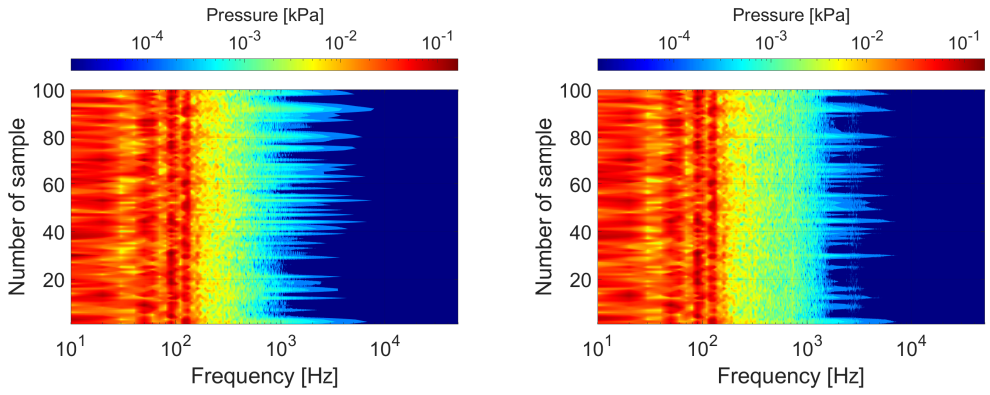
Both spectra are similar in their overall shape, however, the spectrum for SSA is clearer, more coherent and less scattered. Both features capture the range of frequencies to around 2000 Hz, below the blade passing frequency of 2300 Hz [9]. The extent of the spectrum increases with decreasing TOA value, which is probably due to the higher overall power of the specific components used to construct the features. The spectrum of EMD is more fragmented and less continuity between subsequent TOAs is observed. It might imply the leakage of some information into the IMFs not used for building the μ_G feature.

The average demonstrated in Figure 6.5 is based on 100 samples of an input signal. To investigate how stable is the content of the components, the spectra for each individual sample are plotted in Figures from 6.6 to 6.9. The presented data is for TOAs representing the peaks of instabilities, as defined in Chapter 3 starting from deep surge, through transient, inlet recirculation and stable operation.



(a) (b)

Figure 6.6: Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 5% obtained with different decomposition methods a) EMD; b) SSA



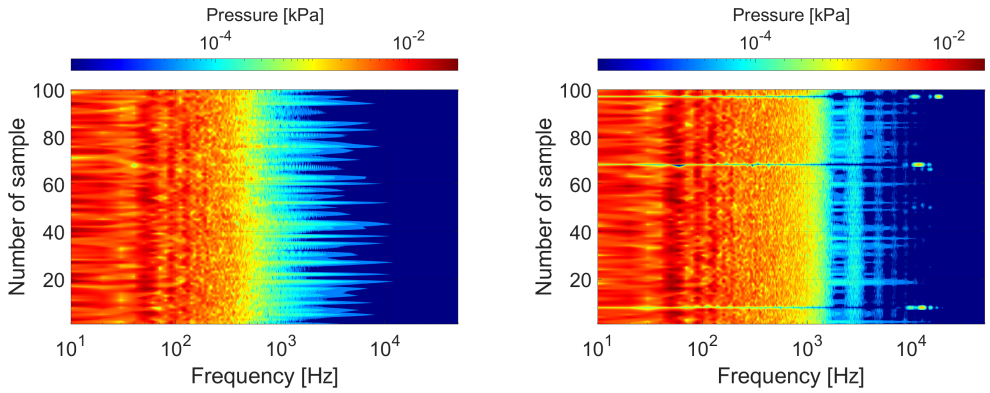
(a) (b)

Figure 6.7: Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 10% obtained with different decomposition methods a) EMD; b) SSA

For deep surge (Figure 6.6), the spectra for both components are very similar. They demonstrate important fluctuations between subsequent samples, which is reflected in the variability of the μ_G. Those fluctuations seem to be correlated for both methods, suggesting that they are related to the fluctuations of the pressure signal. In a number of input signals, the high amplitude part of the spectrum extends above the frequency of 100 Hz. For some of the input signals, the increased amplitude is observed for the frequency of around 90 Hz. It is 90% of the rotational speed, which may imply the temporal presence of some structures related to rotating stall.

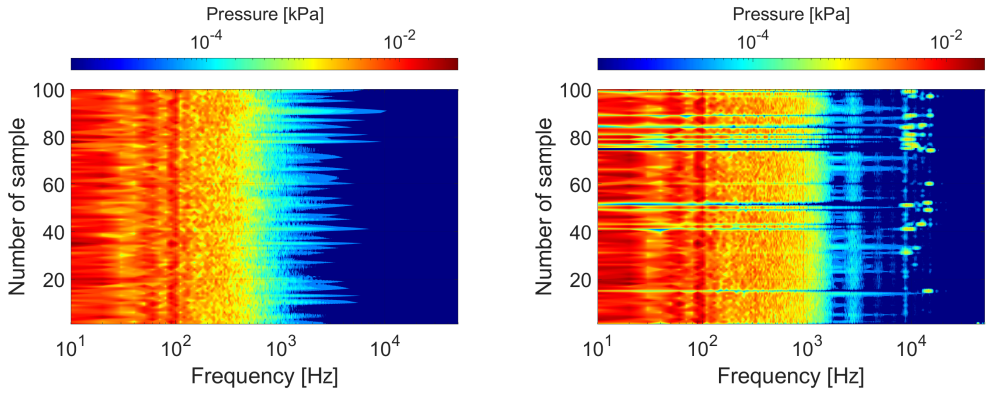
For the transient operation shown in Figure 6.7, the spectra are almost identical for both methods. There is great similarity in both the magnitude and location of the high-amplitude regions.

The spectrum for inlet recirculation (Figure 6.8) shows greater discrepancy between methods. In SSA, for certain input signals, the spectral content is very different than for the majority of the inputs. For such events, high amplitudes are observed for the frequency close to 10,000 Hz. The number of such observations is low, therefore they do not contribute much to the average and do not cause an important variation in the μ_G feature value.



(a) (b)

Figure 6.8: Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 20% obtained with different decomposition methods a) EMD; b) SSA



(a) (b)

Figure 6.9: Contour of a spectral content of μ_G feature for 100 input signal samples for TOA = 30% obtained with different decomposition methods a) EMD; b) SSA

For the stable conditions (Figure 6.9), the spectrum of the EMD-based feature is similar to that for the inlet recirculation zone, with relatively small changes in the overall frequency content and important variations in the amplitude for subsequent signal samples. In the case of SSA, even more samples than for the inlet recirculation zone demonstrate an entirely different spectral content than the averaged value. Such changes in the spectral content of the features

have a more important influence on the feature values, which is demonstrated in Figure 6.10.

The changes in SSA-based μ_G feature spectral content may indicate that RCs may be prone to changing their content when there is no dominating flow phenomena, as in the case of the pressure signal at the inlet in stable conditions. Figure 6.10 demonstrates the value of μ_G along with its spectral content for 100 input signals. The correlation between the change in the content and the drop in energy can be observed. The locations where μ_G does not hold low frequencies display lower levels of energy. It is reflected through the presence of sub-cluster in the feature space for stable and inlet recirculation conditions (see Figure 5.23).

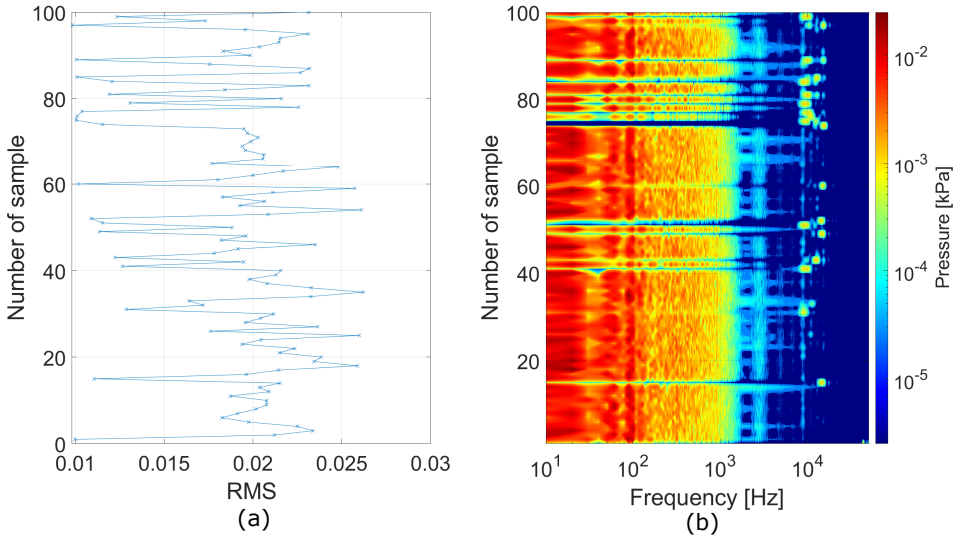


Figure 6.10: Representation of changes in the components for stable operation at TOA 30 % for 100 input subsequent signals; a) μ_G feature; b) spectral content of μ_G

6.2 Threshold-based classification with μ_L and μ_G

To investigate the performance of the feature space method based on EMD and SSA, it is validated with dynamic data and quasi-dynamic data. The dynamic

data are used to qualitatively investigate whether the mapping offered by quasi-dynamic data is representative of dynamic conditions. This will demonstrate if the pace of stability changes can have an influence on the model.

The limited number of dynamic data makes them unsuitable for quantitative validation. Therefore, the second part of validation is made with quasi-dynamic data. A large number of data points allows for obtaining meaningful measures of accuracy and a more detailed comparison of the features based on EMD and SSA.

6.2.1 Validation with dynamic data

For the detection system to operate well in various conditions, it should render identical results no matter the pace of transition between conditions. To investigate the performance of the feature space-based detection of instabilities for a quick transition, the dynamic data set was used. The dynamic data was processed identically to the quasi-dynamic data.

Figure 6.11 shows the SSA-based features obtained using dynamic data plotted onto the feature space along with the features obtained with quasi-dynamic. The points for dynamic data very closely follow the distribution of quasi-dynamic data features. The two clusters for stable conditions are present which are located very similarly to the quasi-dynamic clusters. The data for local instability follows closely the path established by quasi-dynamic features, the same goes for globally unstable conditions. The thresholds established using quasi-dynamic data prove to be valuable for dynamic data. With a quick transition, the locations of peaks of instabilities are accentuated with a higher density of the points in those locations, including differentiation of transient and surge conditions. It can be concluded that the approach based on the decomposition of the signal with SSA should render repeatable results, no matter the pace of changes in the system.

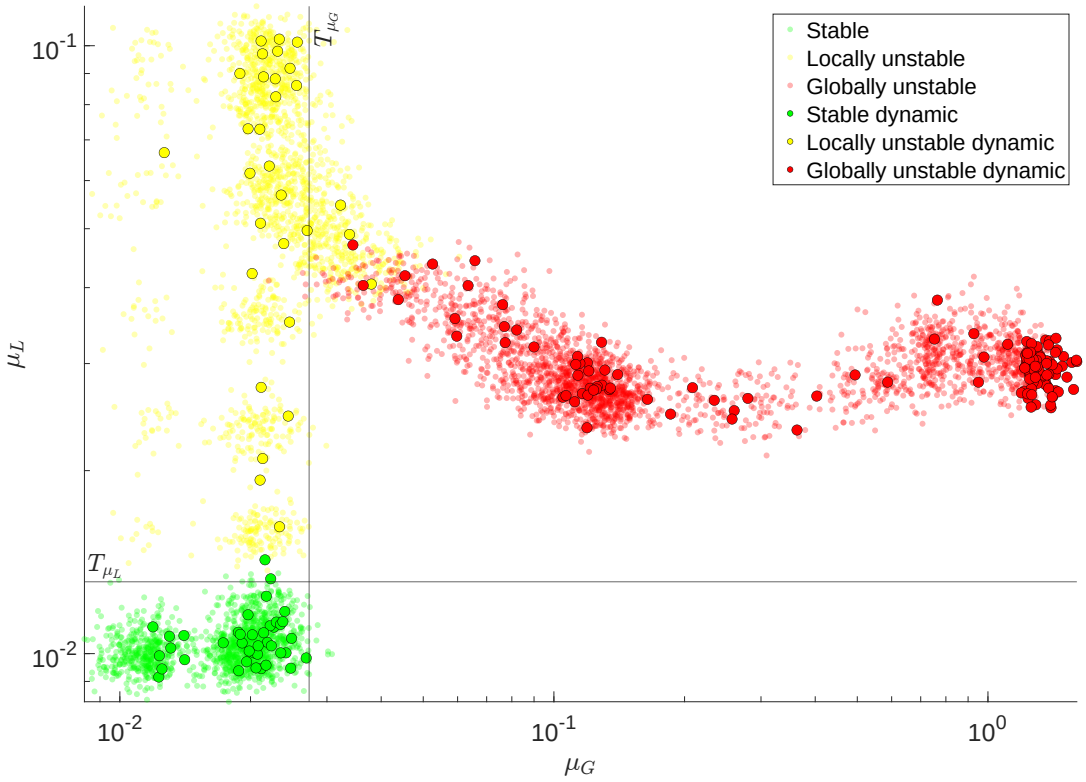


Figure 6.11: Feature space representation of compressor data using μ_L and μ_G features obtained from SSA decomposition; scatter points with black edge represent the dynamic data

Figure 6.12 demonstrates the EMD-based features obtained using dynamic data plotted onto the feature space along with the features obtained with quasi-dynamic. The general shape of the distribution is kept and the dispersion of the data is similar for both data collection protocols. The points for dynamic data do follow the thresholds established with quasi-dynamic data. The peaks of instabilities are also accentuated, with concentration for the peak of inlet recirculation, transient and surge conditions, but it is not as clear as for SSA. It can be noted that the points for dynamic data tend to be shifted to the right in the feature space, compared to the quasi-dynamic ones. The points for dynamic data do not fall over the range established by quasi-dynamic data, but the centres of the two distributions are shifted. The presence of an offset may indicate that the EMD-based features are affected by different dynamics of the

changes in throttling. Should the offset be greater, the model established with quasi-dynamic data would not be correct.

The dynamic data is used for qualitative validation. However, the number of data points obtained with this protocol is scarce and validation of method performance with such data may not be representative. Thus, the validation of the feature space approach for EMD and SSA features is performed using a test set of quasi-dynamic data.

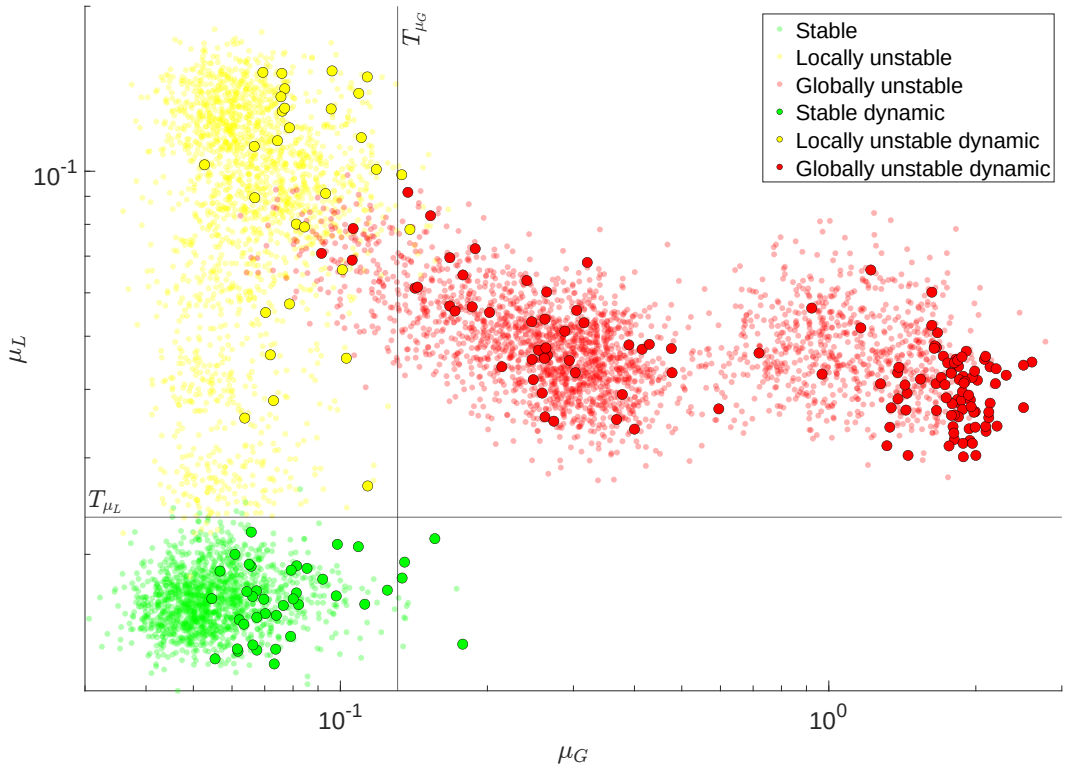


Figure 6.12: Feature space representation of compressor data using μ_L and μ_G features obtained from EMD decomposition; scatter points with black edge represent the dynamic data

6.2.2 Accuracy of classification

The accuracy for EMD and SSA is computed as a percentage of correct classification made to the total number of classified points. The threshold-based approach for classification is used and only the general stability conditions are

considered. To have a better insight into the misclassifications, a confusion matrix is used to show the type of misclassifications. Such confusion matrices for both methods are shown in Figure 6.13.

The accuracy of stable conditions detection is high for both EMD and SSA, reaching an identical 99.5% due to the way the threshold value is defined. The observed misclassifications are for the stable conditions being classified as locally unstable. The accuracy of globally unstable conditions detection is also similar, reaching almost 99% for SSA, and over 93% for EMD. All misclassifications are for the locally unstable class, meaning that a substantial number of points collected from globally unstable region has a value below a threshold set based on stable conditions. The biggest difference in accuracy is present for locally unstable conditions, where the accuracy for SSA drops to below 80%. A high number of points is classified in error as globally unstable. This is not observed for EMD, where almost 99% of the points are classified correctly and only below 1% is classified as globally unstable conditions.

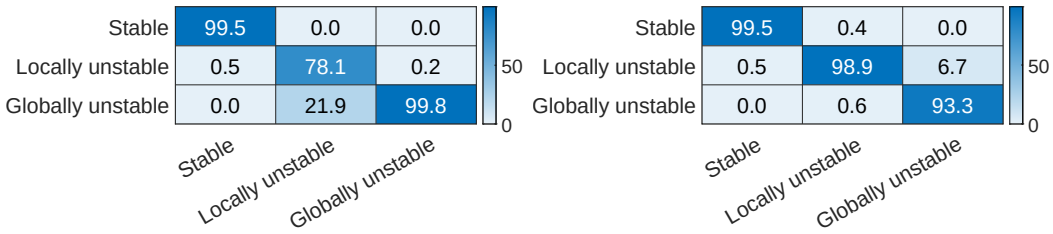


Figure 6.13: Confusion matrix between the conditions using features based on different decomposition methods a) SSA and b) EMD

To better understand the reasons for misclassifications, the exact indications for each TOA are investigated. Figures 6.14 and 6.15 show the percentage of data assigned to each class for SSA and EMD respectively. In the stable zone for both methods, all the misclassifications are coming from the same TOA at the edge of the stable range. Their number is identical, hence the identical value of misclassification percentage for both methods in Figure 6.13.

The behaviour in locally unstable region differs. For SSA, the number of misclassifications is stable and low at the beginning of the range and starts to increase when approaching the edge of the zone. It grows gradually, starting from TOA of around 21%. This gradual change in the number of misclassified points may suggest that the signature of global instability is temporarily present in the signal. The period of its presence increases with decreasing TOA value. Hence, a higher number of points for a given TOA is defined to be globally unstable while being in a locally unstable region. The presence of misclassifications of this type does not necessarily have to be considered an error, as the true conditions experienced by the compressor may be globally unstable. For the globally unstable region, the accuracy is very high and only misclassifications are present at the very edge of the region.

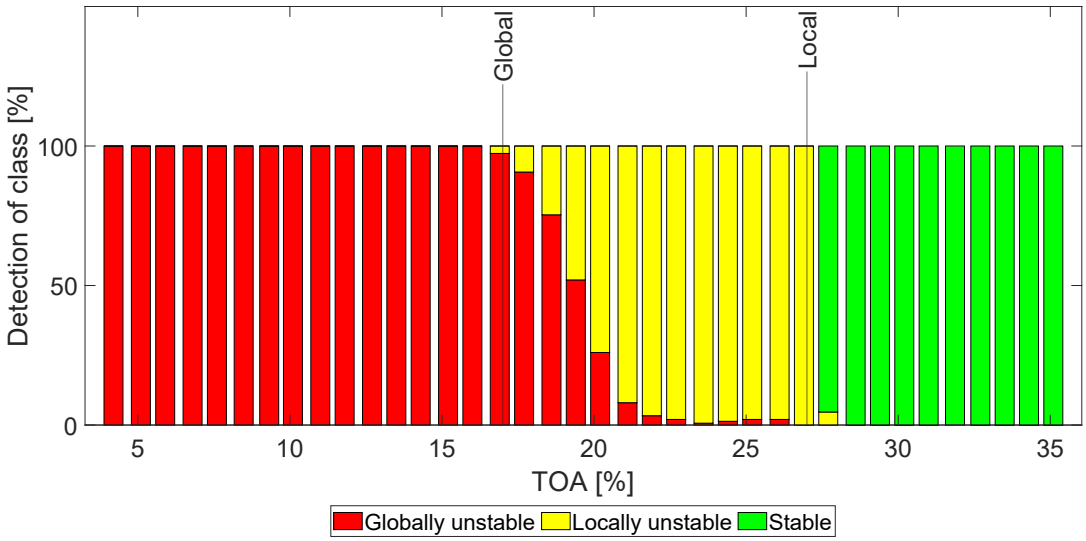


Figure 6.14: Classification of data for different TOAs based on feature space obtained using SSA-based features scattered in Figure 6.11

For EMD, the number of misclassifications for most of the locally unstable range is very close to zero. The increase in misclassifications appears for the last three TOAs close to the edge of the globally unstable region, but its number does not exceed 5% for each of the TOAs. Similarly as for the SSA, certain TOAs share the indication of locally and globally unstable conditions. The gradual

increase in the number of points classified as globally unstable with decreasing TOA is not as evident as in the case of SSA. It is more concentrated and shifted to lower TOA values. This shift may arise due to a higher dispersion of the stable conditions cluster, hence the shift in the thresholds T_{μ_G}. In the globally unstable range, the number of misclassifications is high at the beginning of the globally unstable range, coming from the development of the globally unstable conditions. For the rest of the range, the method is perfectly accurate.

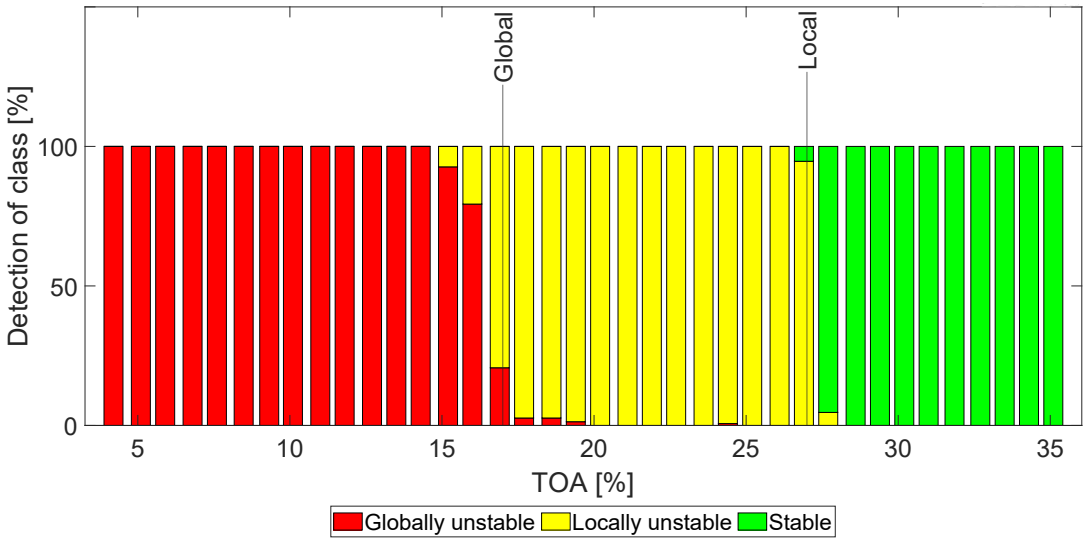


Figure 6.15: Classification of data for different TOAs based on feature space obtained using EMD-based features scattered in Figure 6.12

Both EMD and SSA provide physically interpretable components that generally perform well in differentiating the general operating conditions of a compressor. However, their accuracy drops in the zones where the transition from one instability to another takes place. This is quantified in Figure 6.16. The zone between locally unstable and globally unstable conditions is the location where the accuracy drops to as low as 30% for selected TOAs. This transition zone may be a challenge for the operation of a sensitive instabilities detection system, as it might lead to an unnecessarily early reaction of an anti-surge system. The transition in compressor operation from one condition to another is

unavoidable, hence it might be suitable to abstain from classification in certain zones of the feature space to avoid this region.

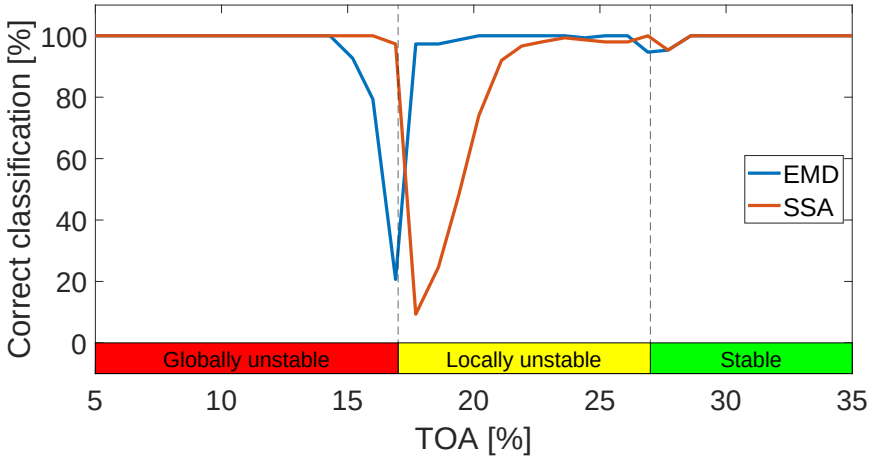


Figure 6.16: Class detection based on the feature space representation and thresholds for EMD and SSA

6.3 Practical consideration for an instabilities detection system

6.3.1 Selection of the components

One of the most important aspects of building a model for instabilities detection is the selection of the components to be used as features. In this thesis, the selection was backed by the analysis of the components behaviour in the whole operating range. This might not be possible in most cases, when the operation of the compressor in the unstable range is not feasible. The selection of components might be case specific, as the machines and acquisition systems may differ significantly. However, some general guidance can be provided based on the case study presented in this thesis.

When using EMD for extraction of the features of instabilities, the global instability feature can be defined as the one holding the subsynchronous frequencies of the signal. To compute a sum, the first IMF to be used has to be identified, but it can be done using data for stable conditions. The first IMF to

be used in the summation could be defined as the one holding frequencies below the blade passing frequency. The feature to be used should be a summation of the energy of this IMF and all subsequent ones. In this case study, the inlet location was the best for obtaining a precise indication, as only then the instability can be considered a global one. The local instability, being inlet recirculation, should be based on data from a sensor located close to the leading edge of an impeller blade. Only there and shortly downstream it is possible to reliably find the signature of this instability. However, the measurement from before impeller is much clearer and also far easier in practical terms. The signature of inlet recirculation is distributed between a number of IMFs, which demonstrate a similar detection performance. However, the definition of a central frequency of the expected IMF is not trivial and may depend on the geometry of the impeller and its rotational speed [43].

When using SSA, the selection of the components is dependent on the window length used for the decomposition. Should a short window be used, as recommended in this study, the first two components are of interest. The first one is expected to hold a signature of global instability, while the second should indicate local instability. The same selection of sensors location for each of the instabilities is recommended as for EMD. The local instability feature should be obtained from a sensor located before impeller and the global instability feature from the inlet.

The selection of components could possibly be facilitated with the aid of simulation techniques. It could be possible to use high-fidelity CFD simulations [16] or tailored numerical models [73] to predict the spectrum of the signal. The possibility of obtaining the signal for unstable operation without the need for endangering the compressor should allow greater accuracy in defining the components.

6.3.2 Advantages and disadvantages of the methods

The instability-sensitive features obtained with EMD and SSA are very similar in their physical nature. A difference can be observed in the dispersion of the data. It is bigger for EMD features, possibly due to the effects of mode mixing that are not fully eradicated even if a sum of IMFs is used. The lower dispersion translates into more concentrated clusters of conditions than those for EMD, when using the same input signal length. The pace of the decomposition is similar for both methods, with each of them looking promising in the context of online detection.

EMD shows higher overall accuracy when using a threshold-based approach. Fewer misclassifications are made in the local instability zone, compared to SSA. However, SSA produces fewer errors in the globally unstable range. The latter may be important, as the consequences of misclassifying a point that should belong to globally unstable conditions may be high. Misclassifying locally unstable conditions for global ones will on the other hand decrease the operating range of the machine, as the potential anti-surge system would act earlier.

To provide a conclusive statement, the conditions labelling of the data should be reviewed. Assignment of a single class to every TOA seems not right since the temporal changes in stability are possible withing the same TOA, especially in the zone where locally unstable conditions transition into globally unstable ones.

Assessing qualitatively the distribution of the points in the feature space, the stable cluster for EMD is more dispersed than that for SSA. Several data points are shifted to the right, influencing the placement of the threshold T_{μ_G}. It may indicate that EMD is more prone to producing outliers that would influence the outcome of the detection.

The validation for dynamic data shows that SSA-based components are more consistent with quasi-dynamic data than EMD-based components. With a small

sample of dynamic data used in this study, the detection accuracy is comparable between two methods. However, the offset may change for other machines, rendering the EMD-based approach unfit for exact detection of instabilities when a quick change in stability takes place. Both methods demonstrate a very similar performance on the investigated data set, therefore they both can be successful used for instabilities detection in centrifugal compressors. However, if an overall better method should be indicated, the SSA-based components provide a more precise indication of the stability of a compressor. The disadvantage of EMD is the mode mixing, that artificially increases the dispersion of the already noisy data from the compressor.

The timing of the methods is shown in Figure 6.17. Using the parameters proposed in this thesis, both EMD and SSA provide the decomposition time below the online processing limit. It can be translated to a system producing a continuous, online indication of conditions. The final approaches for EMD and SSA, using 15 IMFs and two RCs with $L = 50$, provide almost identical timing. It is highly probable that the time of decomposition could be furtherly decreased with different combinations of hardware and software.

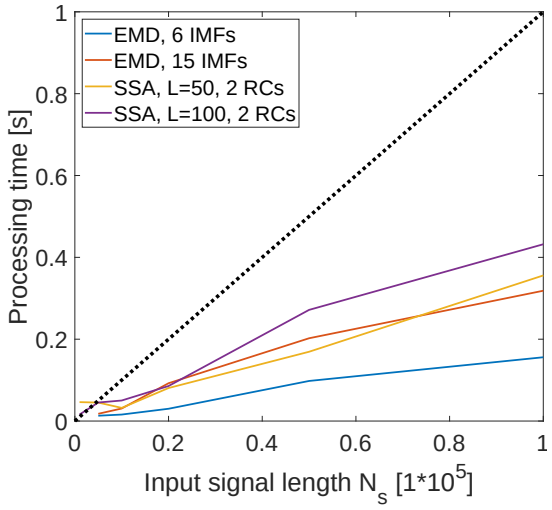


Figure 6.17: Timing of the methods for different input length, decomposition parameters and number of components; dashed black line is an online processing limit above which the method no longer can provide real-time indication

6.3.3 Extension to accommodate other instabilities

The methodology presented in this study focuses on extraction of local instability, being inlet recirculation and global instability, incorporating transient and deep surge conditions. Despite some resemblance of transient operation to rotating stall, the stall in its most common form [7] was not observed in the data. This brings a question about the extension of the approaches based on EMD and SSA to other machines, experiencing rotating stall. Some hypotheses regarding that situation can be made, based on the mechanisms of the decomposition and observations made for the transient operation. Should a signature of stall be consistent, it would be captured by some components of each decomposition.

Rotating stall is often manifested by frequencies between 40 to 100 % of the rotational speed of the machine, which are often lower than the frequency of inlet recirculation, but higher than surge frequency. Consequently, using EMD decomposition, rotating stall should be captured by the IMFs higher than those for inlet recirculation, but lower than that for surge. It should occupy the same frequency range as transient components demonstrated in this study. Thus, it might bring identical problems related to mode mixing and a value above threshold for the conditions past rotating stall.

When using SSA, the extraction of rotating stall component would be possible, provided that the window length is large enough. For a short window, the signature of stall would probably be captured in RC 1 along with other low-frequency pressure oscillations. The extraction of rotating stall based on a specific threshold might be challenging, as it is expected that stall-related frequencies will be close to deep surge frequency [7].

Nevertheless, the decomposition methods could still be useful for capturing rotating stall. The methodology proposed in this study could also easily incorporate additional stall feature to be used in the detection system.

6.3.4 Limiting the number of sensors for detection

In the presented study, the data from two sensors is used for defining a feature space used for classification. With the intention to limit the cost of the condition monitoring system, the question of using a single sensor for monitoring may be raised. The selection of the sensor location is dictated by the local instability, clearly detectable only using the sensor p_{s-imp1} before impeller. Consequently, to offer the detection using a single sensor, it should be evaluated if the global instability can be clearly captured by the p_{s-imp1} sensor.

Using the data from before impeller, it is possible to define features that hold the signature of both global and local instability. Unlike when using two sensors, it is not possible to obtain features that fully isolate the signature of local and global instability, as the increase of RMS takes place for both features in locally unstable conditions (see Figure 4.13 and 5.2). It renders the threshold-based approach inadequate. However, using the data for all operating condition for training of the model, the detection system using only the data from the sensor located before impeller should perform similarly to those obtained for inlet sensor. Figure 6.18 demonstrates the feature space representation for both EMD and SSA using only the data from the sensor p_{s-imp1} before impeller. The μ_L feature is the same as previously, while μ_L feature is based on RC 1 for SSA and sum of IMFs 8 to 15 for EMD.

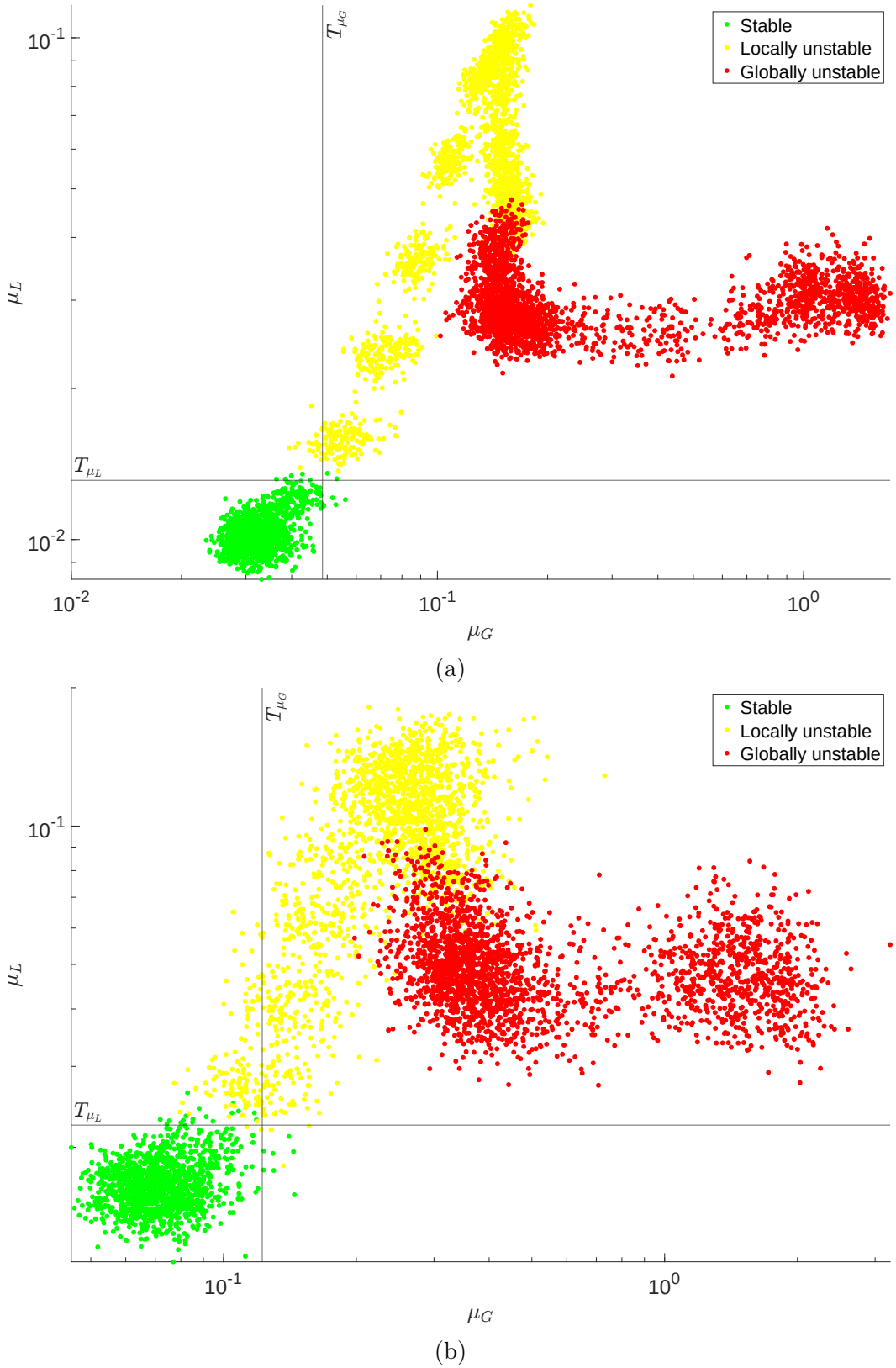


Figure 6.18: Feature space representation using only components obtained from sensor p_{s-imp1} before impeller a) SSA-based features b) EMD-based features

6.3.5 Challenges and space for improvements

Both methods investigated in this thesis demonstrate high accuracy in detection of general stability conditions based on thresholds derived from stable operating conditions. Using this approach, a few misclassifications can be observed in the TOAs between the locally unstable and the globally unstable range. The misclassifications seem to be related to the temporal presence of globally unstable traces in the locally unstable zone. In such context, the correct classification of conditions in that region may be challenging. Hence, the systematic approach allowing not to assign the class to points appearing in some locations in the feature space could be of help.

To increase the possible operating range of the compressor and allow for greater flexibility in setting the anti-surge system activation conditions, the differentiation of globally unstable conditions can be investigated. It is possible to differentiate the conditions based on the location of the points in the feature space, however the thresholds derived based on stable conditions are no longer useful. To allow for such differentiation, the data for unstable conditions has to be collected to initially map the feature space. This is associated with the risk of damaging the compressor. Consequently, the time of operation in unstable zone should be short, leading to generation of a limited number of points. Hence, the mapping of the feature space should be enabled with a low number of points from unstable conditions. Both of those aspects can be addressed with a probabilistic classification approach, which is demonstrated in Chapter 7.

7 Probabilistic classification of instabilities

This chapter demonstrates a probabilistic classification approach implemented to improve the accuracy of instabilities detection and address the misclassifications present for the threshold-based approach in the regions of transition between instabilities. Gaussian process classification (GPC) is used, as it produces a probabilistic output and enables flexible, data-driven shaping of the boundaries between classes. Four aspects are considered in this chapter. Firstly, it is demonstrated that the GPC model can be trained to accurately classify different unstable conditions in the feature space. Secondly, the concept of the rejection zone is explored, which defines the regions where the model should abstain from classification. It is shown that with a proper selection of the model parameters, the rejection zone based on the difference in the predicted probability of the two most probable classes can improve the accuracy of the classification model. Different scenarios are investigated with respect to the availability of training data. With the threat to compressor operation presented by the instabilities, the availability of training data for unstable conditions may be limited. The impact of limitations is demonstrated and discussed, showing that GPC is capable of providing accurate predictions using scarce training data. Lastly, it is demonstrated that with GPC, it is possible to use only data from a single sensor to provide a high accuracy classification, which could decrease the requirements for the potential instabilities detection system using the proposed methodology.

7.1 Gaussian process classification

Gaussian process classification (GPC) is a non-parametric Bayesian algorithm that models the probability of each class directly as a function of the input features [168]. It does not make any underlying assumptions about the functions used to model the dependencies. GPC is based on a Gaussian process regression (GPR) [203]. Given a classification task, a regression is solved first with an input feature vector and target class labels, often expressed as integer values. The regression solution is subsequently recomputed into classification, using for example logistic function [203]. Since the mathematical formulation of GPC is based on regression, a GPR process must be introduced.

GPR aims to find a function f that best relates the input variable x to the output variable t. In most cases, the measurements of both input and output variables have a certain level of uncertainty or noise present in the data, represented by ϵ [203]. A general regression problem is shown in Equation (7.1). When building a regression task for classification as in this case, the noise term can be omitted, as the labels are assumed to be correctly assigned.

$$t = f(x) + \epsilon \tag{7.1}$$

Observing the compressor system and processing the input through the decomposition methods, a data set $\boldsymbol{D} = (\boldsymbol{F}, \boldsymbol{c})$ can be obtained, where $\boldsymbol{F}$ collates all the features while $\boldsymbol{c}$ represents the class labels for each feature vector. The relation between input features and output class labels is modelled as a Gaussian process (Equation 7.2). A Gaussian process can be fully described by a mean function $m(\mu)$, and a covariance function $k(\mu, \mu')$, which can be defined as in Equations (7.3) and (7.4) respectively, where $E[\cdot]$ stands for an expected value.

$$f(\mu) \sim \mathcal{GP}(m(\mu), k(\mu, \mu')) \tag{7.2}$$

$$m(\mu) = E[f(\mu)] \tag{7.3}$$

$$k(\mu, \mu') = E[(f(\mu) - m(\mu))(f(\mu') - m(\mu'))] \tag{7.4}$$

A Gaussian process uses the distribution over functions, assigning the probability of a given function fitting the data. Having a new feature vector $\boldsymbol{\mu_*}$ for the conditions to be classified, the aim is to find the output function values t_*. Taking advantage of the Gaussian formulation, a joint distribution of the known functions values $\mathbf{t}$ and the unknown value t_* can be assumed (Equation (7.5)).

$$\begin{bmatrix} \mathbf{t} \\ t_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K_t} & \mathbf{K}_* \\ \mathbf{K}_*^{\mathbf{T}} & \boldsymbol{K_{**}} \end{bmatrix}\right) \tag{7.5}$$

It is a common approach in practical applications to assume a zero mean for the process, as it gets updated with the data [203]. The covariance matrix in Equation (7.5) is composed of four terms representing the relations between the training data $\mathbf{F_t}$ and the vector of new data points $\boldsymbol{\mu_*}$. $\mathbf{K_t} = k(\mathbf{F}, \mathbf{F})$ is the covariance between the training points, $\mathbf{K}_* = k(\mathbf{F}, \boldsymbol{\mu_*})$ represents the covariance of the training points and the new data points and $\boldsymbol{K_{**}} = k(\boldsymbol{\mu_*}\boldsymbol{\mu_*})$ represents the covariance for the new data vector.

The covariance can be computed directly, as in Equation (7.4) or approximated with a kernel function. The kernel function defines the covariance based solely on two inputs of a feature vector, marked μ and μ'. It importantly limits the computational effort compared to computing the covariance matrix for the data [168]. The choice of kernel function type for a task is a challenge and can have an important influence on the outcomes of the model. It has been a subject of extensive research [204], but remains largely task-specific.

Several basic kernels are used in practical applications [204]. Selected kernels relevant to the application in the problem of classification for the detection of instabilities are given in Table 7.1. It is possible to combine kernels with

Table 7.1: Selected kernels and their hyperparameters

Kernel name	Expression	Hyperparameters
Dot product (DP)	$k(\mu, \mu') = \sigma_0^2 + (\mu\mu')$	σ_0
Dot product squared (DP^2)	$k(\mu, \mu') = (\sigma_0^2 + (\mu\mu'))^p$	σ_0
Radial basis function (RBF)	$k(\mu, \mu') = \sigma_0^2 \exp\left(-\frac{d(\mu,\mu')^2}{2l^2}\right)$	σ_0, l
Rational quadratic (RQ)	$k(\mu, \mu') = \left(1 + \frac{d(\mu,\mu')^2}{2\alpha l^2}\right)^{-\alpha}$	α, l

each other through addition or multiplication [204], which produces a virtually infinite number of possible kernel functions.

Each of the basic kernels has its parameters, which are called hyperparameters. The hyperparameters of the kernel are optimised when the model is trained to optimally fit the training data. For DP kernel, the single hyperparameter is σ_0, which determines the offset in the linear model. The same hyperparameter σ_0 is present for the polynomial kernel. The p in the formula for polynomial kernel is not considered a hyperparameter, but it is set a priori as a positive integer. Setting $p = 2$, such a kernel becomes a dot product squared kernel (DP^2). For the RBF kernel, there are two hyperparameters. The variance σ defines the average distance of the function from its mean, while the length scale l influences the scale of function variability [204]. The RQ kernel is equivalent to adding several RBF kernels together. It has three hyperparameters, σ and l have the same interpretation as for RBF while α defines the relative weight of the large-scale and small-scale variations [204]. The hyperparameters of a kernel are grouped together in a vector $\boldsymbol{\theta}$ for optimisation. A comprehensive study of kernels in Gaussian processes can be found in [205].

Having defined the prior and using the data, it is possible to infer about the posterior. The posterior is given by Equation (7.6). It is normally distributed with mean $m(t_*)$ and covariance $cov(t_*)$, which can be computed as per Equations (7.7) and (7.8) respectively.

$$p(t_*|\mathbf{F}_*, \mathbf{F}, \mathbf{t}) = \mathcal{N}(m(t_*), cov(t_*)) \tag{7.6}$$

$$m(t_*) = \mathbf{K_{t_*}}^T \mathbf{K_t}^{-1} \mathbf{t} \tag{7.7}$$

$$cov(t_*) = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K_t}^{-1} \mathbf{K}_* \tag{7.8}$$

The covariance does not depend on the output values $\boldsymbol{t}$ but only on the covariance matrices, defined through the kernel function. Due to this, the prediction of the model strongly depends on the type of kernel chosen and its hyperparameters $\boldsymbol{\theta}$. To fit the model to the data, the hyperparameters are optimised. The goal is to maximise the log marginal likelihood L_M, given by Equation (7.9), with $|\mathbf{K_t}|$ being a determinant of $\mathbf{K_t}$. The two first terms of log marginal likelihood are model fit to the data and model complexity respectively [206], therefore the optimisation aims to find the trade-off between those two.

$$L_M(\theta) = log(p(\mathbf{t}|\mathbf{X})) = -\frac{1}{2} log|\mathbf{K_t}| - \frac{1}{2}\mathbf{t}^T \mathbf{K_t}^{-1}\mathbf{t} - \frac{N}{2} log(2\pi) \tag{7.9}$$

The GPR outcome is a function modelling the relation of the input $\boldsymbol{\mu}$ and the output $\boldsymbol{t}$. The value of $\boldsymbol{t}$ should provide the best fit for the training data however, it is not bounded and therefore it cannot be used directly for inferring the probability. To translate the regression problem into binary classification, the regression outcome $\boldsymbol{t}$ is transformed through an additional function. A commonly used function is a logistic function as in Equation (7.10), that allows mapping of the theoretical regression range $(-\infty, \infty)$ into a classification range $[0, 1]$ [203].

$$\sigma(t) = \frac{1}{1 + exp(-t)} \tag{7.10}$$

The aim of classification is to assign a correct class based on a vector of new observations $\boldsymbol{\mu_*}$. Using a probabilistic classification approach, the assigned class

is the one with the highest probability. For binary classification, it is enough to compute the probability for a single class $p(c_* = 1|\boldsymbol{\mu_*}, \boldsymbol{F}, \boldsymbol{t})$, as the other class probability $p(c_* = 0|\boldsymbol{\mu_*}, \boldsymbol{F}, \boldsymbol{t})$ can be deduced as $1-p(c_* = 1|\boldsymbol{\mu_*}, \boldsymbol{F}, \boldsymbol{t})$. To obtain the predictive distribution for a class $c = 1$, the integral from Equation (7.11) is evaluated. It is not analytically traceable, therefore it has to be approximated, for example using the Laplace method [203].

$$p(c_* = 1|\boldsymbol{\mu_*}, \boldsymbol{F}, \boldsymbol{t}) = \int \sigma(t_*)p(t_*|\boldsymbol{\mu_*}, \boldsymbol{F}, \boldsymbol{t})df_* \tag{7.11}$$

Binary classification can be translated into a multiclass one using a one-vs-all approach, where a binary classifier is trained independently for each label and the results are combined to provide final probability estimates [206]. The final class is assigned based on the highest probability predicted by the model that combines all the binary classifiers [207].

The GPC algorithm applied in this study is a Python implementation included in the scikit-learn library [207], based on the formulation proposed by Rasmussen [203]. The hyperparameters of the kernel function are optimised using a gradient-based L-BFGS-B optimiser [207]. The algorithm performs a number of one-vs-all classifications and the final model outcome is obtained by combining all trained models [168].

Using the probabilistic output of GPC, it is possible to use relations between class probability and define regions where no classification is made [203]. Such a region can be called a rejection zone or a no-classification zone. Assuming that a new feature vector $\boldsymbol{\mu_*}$ is obtained from the process, it is entered into the model and the probability for each of the classes is predicted. If the two most probable classes are c_i and c_j respectively, a rejection zone can be defined based on the difference in probability between the classes [182]. If such a difference is lower than a certain threshold (Equation (7.12)), $\boldsymbol{\mu_*}$ is not classified [178] and no decision on the stability of the compressor is made based on such an input.

$$p(c_i) - p(c_j) = p_{threshold} < t_{reject} \tag{7.12}$$

An example of GPC for one-dimensional data using different kernels is presented in Figure 7.1. The blue line demonstrates the probability for one of the classes, obtained through the translation of the regression result.

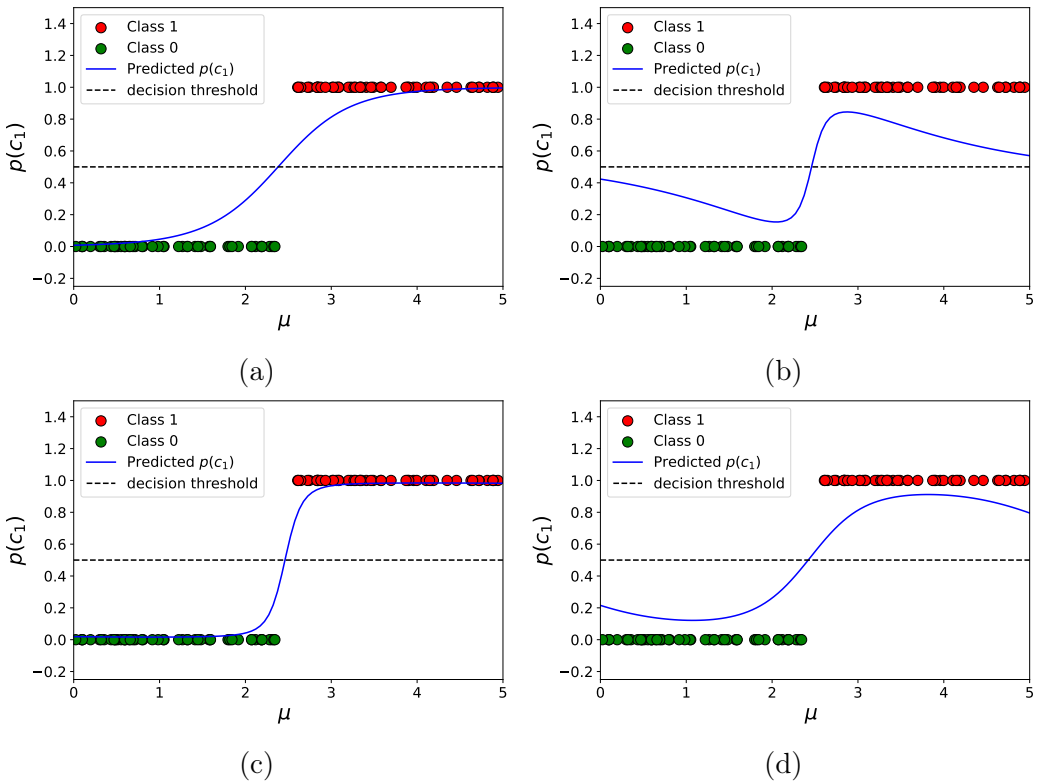


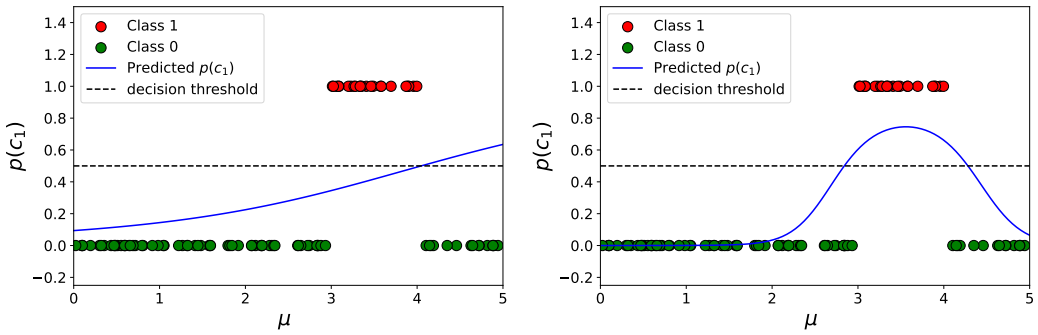
Figure 7.1: Classification of one-dimensional data with GPC using different kernels a) DP, kernel; b) RBF kernel c) DP^2 kernel d) RQ kernel

For the DP kernel (Figure 7.1a), the transition between classes is gradual and the threshold is crossed at the right location. When moving away from the interface between classes, the probability of the class asymptotically reaches the value of 1. The same can be observed for the DP^2 kernel (Figure 7.1c). The difference is that the change in probability level at the interface is more abrupt and the probability quicker reaches the value of 1. For the RBF kernel

(Figure 7.1b), there is an important change in the probability level at the interface between classes, while the probability decreases at some distance from the interface, where it approaches the decision threshold. The RQ kernel (Figure 7.1d) demonstrated a behaviour similar to DP at the interface between classes and to RBF further away from the interface. For RBF and RQ kernels, the probability of a class is the highest close to the interface and decreases away from it.

In a number of practical engineering applications, the RBF kernel is often used due to its flexibility in shaping class boundaries [179, 208]. The attempt to use this kernel for the problem of instabilities detection and classification was investigated by the author of this thesis in a separate study [209]. It was shown that the GPC based on the RBF kernel can provide accurate classification of the conditions. However, the probability level predicted by the model cannot be interpreted as the probability of instability. Due to the characteristics of the RBF kernel, the probability was decreasing for a high value of the feature space coordinates, where it should increase due to the higher certainty of the instability. Hence, it invalidated the concept of the rejection zone.

To take advantage of the interpretable feature space proposed in this study, the probability distribution should be interpretable in terms of instability probability. The higher the value of the feature, the more important the intensity of the instability, which should be reflected in the probability level. The DP kernel and the DP^2 kernel show the expected behaviour when a simple two-class problem is investigated (Figures 7.1 a) and c)). However, the DP kernel does not provide sufficient flexibility in modelling more complicated data, as demonstrated in Figure 7.2a. The dot product kernel does not fit the data well as it cannot adapt due to its structure. This flexibility can be provided by the kernel. Using the DP^2 kernel, the correct classification of the data is possible (Figure 7.2b) therefore, the DP^2 kernel is selected for further modelling.



(a) (b)

Figure 7.2: Comparison of probability prediction with GPC model for different kernels a) DP kernel; b) DP^2 kernel

7.2 Probabilistic classification for compressor instabilities detection

The GPC model was trained with data from SSA decomposition of the pressure signal. Data from sensor p_{s-imp1} before impeller was used to construct the local instability feature μ_L and data from p_{s-in} sensor at the inlet were used to obtain the global instability feature μ_G, as defined in Chapter 5. The training set consisted of 20 points per each TOA and four classes were defined in the data, following the labels provided by the authors of the experimental study [9]. The features from SSA decomposition were selected as they demonstrated less dispersion in the feature space and the data obtained with the quasi-dynamic protocol was a better representation of the dynamic data than in the case of EMD (Chapter 6). The features obtained with EMD were not analysed as they had a similar distribution in the feature space, therefore similar results of applying GPC were expected.

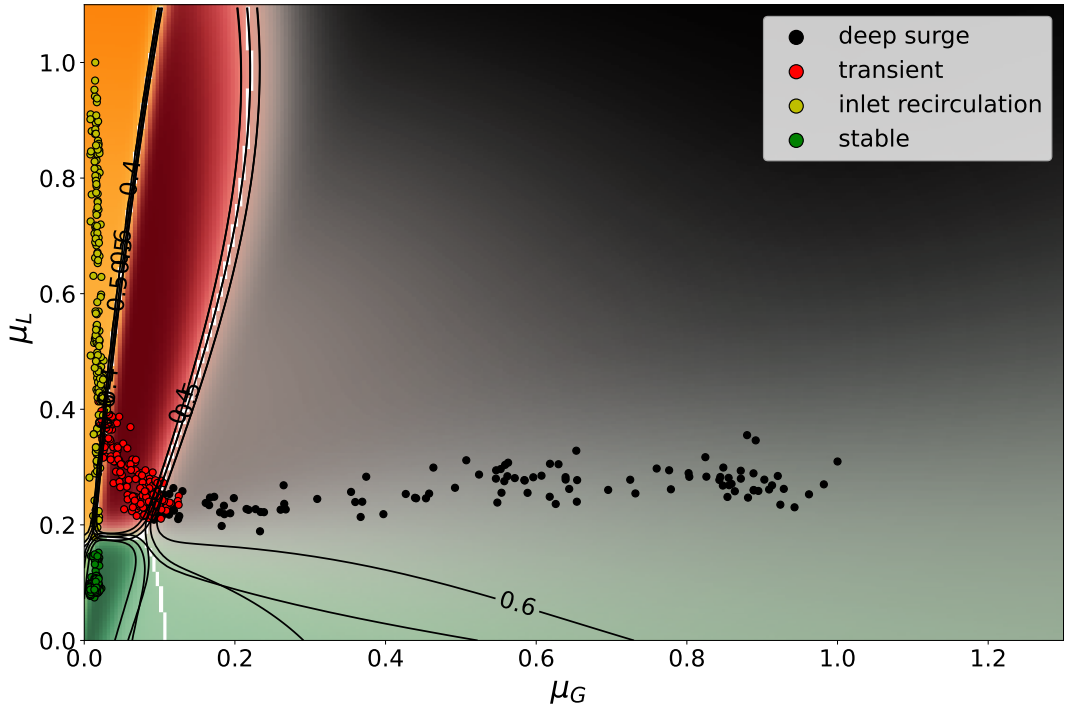


Figure 7.3: GPC results for the training data comprised of four classes of conditions. The colourmaps represent the probability for each of the classes. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points for different classes defined in the legend

Figure 7.3 shows the results obtained with GPC trained on quasi-dynamic data, scattered in the figure. The output of a GPC is the probability of each of the classes for all locations within the feature space. The probability level is represented with a separate colourmap for each of the conditions. The colourmaps are superimposed in the figure. The intensity of the colour and opacity increases with increasing probability level. To facilitate the interpretation of the probability distribution, the contours of probability for values $p \in \{0.4, 0.5, 0.6\}$ are drawn for each class. The probability for each of the classes is used to define class boundaries, which are plotted with white solid lines.

The class boundaries from the model fit the data well and provide a good distinction between different conditions. Stable, inlet recirculation and deep surge classes are shaped as expected. The transient class seems overly elongated

in shape with respect to the concentrated training data, stretching upwards to high values of μ_L. It is influenced by the DP^2 characteristics and the training data locations in the feature space which does not bound the transient class for higher values of μ_L. The contours of probability are located very close one to another for the majority of the feature space. The bigger distance between them can be observed at the boundary between stable and surge conditions, where the training points are not as closely located as for other class interfaces, which translates into a more gradual change in classes probability.

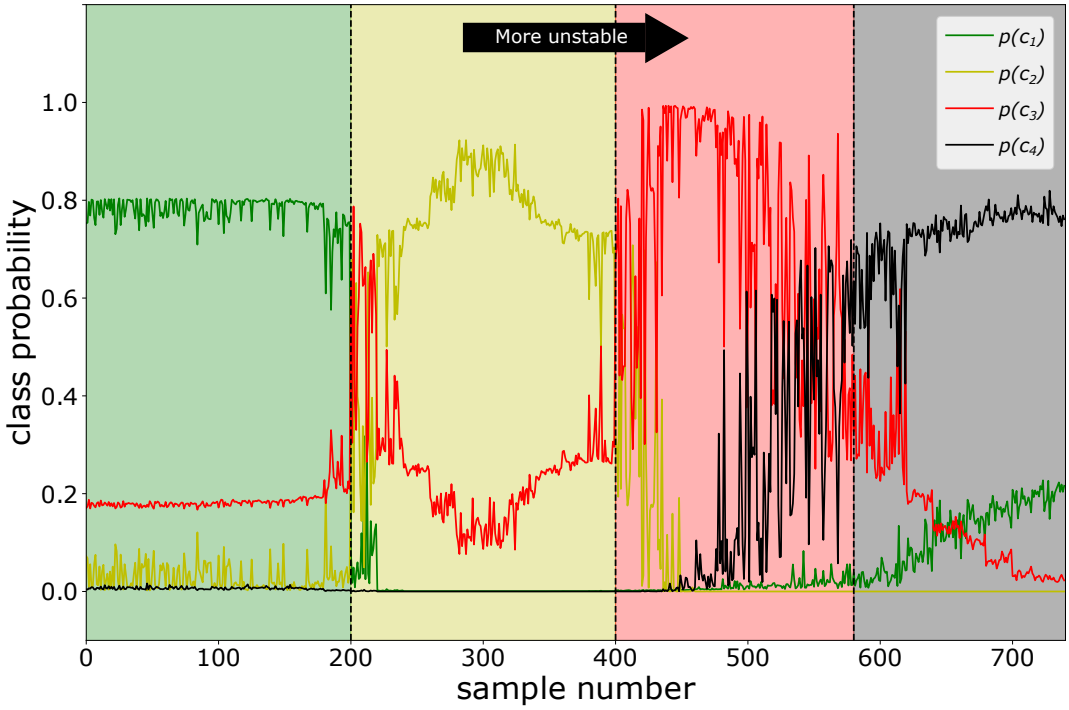


Figure 7.4: Changes in probability predicted by the model for the training points presented in Figure 7.3; vertical lines signify the boundaries of classes while colours demonstrate the regions defined as stable (green), inlet recirculation (yellow), transient (red) and deep surge (grey); the class probability is given for classes: c_1 - stable, c_2 - inlet recirculation, c_3 - transient and c_4 - deep surge

To investigate the changes in the probability of each class predicted by the model, its estimates are obtained for the 740 training points representing different classes of conditions. The probabilities for these points predicted by the

model are demonstrated in Figure 7.4. The points in the plot are ordered by the decreasing TOA value, hence the higher the sample number in the figure, the more unstable the conditions. Therefore, moving towards the right on the plot, the more unstable conditions are expected. The background shading demonstrate the labelled classes of the training data.

In the stable region, the probability of stable class is close to 0.8 except for the region close to transition, where it decreases locally to approximately 0.6. For neither of the points investigated, it reaches the value of 1. In a stable region, the probability of transient conditions is relatively constant at a value of 0.2. At the boundary, a number of observations demonstrate the highest probability of transient conditions, despite being sourced from the inlet recirculation class. It is caused by the elongated shape of the transient class predicted by the model and it reaching the points for the early stages of inlet recirculation (Figure 7.3). Increasing the sample number, the probability of the inlet recirculation class increases. It is the highest in the centre of the class, reaching a value above 0.9 and decreases when approaching the transient conditions. In the transient region, the transient class probability is high on average, but it is highly volatile for the region of transition between transient and deep surge conditions. With continuous changes in model probability, this is caused by the dispersion of the training data points in the feature space. From the middle of the transient region, the probability of deep surge starts to increase. It surpasses that of the transient class at the beginning of deep surge region. In the deep surge region, the highest probability is presented by the deep surge class, however, in the early stages of deep surge, the probability of the transient class shows a considerable value of 0.4. The probability of deep surge in the deep surge region does not exceed 0.8. The probability level does not decrease with an increasing sample number. This is a desired feature of the model, as the probability should not decrease for the points with high μ_G coordinate, which was the case when using an RBF kernel [209]. However, the probability does not seem to increase,

while such a trend can be observed for the stable class probability. Therefore, it might be expected that, despite the form of the kernel function, the probability computed by the model, especially in the regions where the training data is missing, does not behave in an interpretable manner.

7.3 Rejection zones

When closely inspecting the data, it can be noted that in some locations of the feature space, the points of different classes overlap. This is the case at the transition between different conditions, mostly between inlet recirculation and transient as well as transient and deep surge. Those regions are bound to result in misclassifications, regardless of the location of the boundaries between the classes. The problem of misclassification in the transition zones could be addressed by considering the probability values for each class and their relations.

Using the probabilistic output of GPC, it is possible to define regions where no classification is made [203]. Such regions can be called rejection zone [178] or no-classification zone. Two types of rejections can be considered. One related to the novelty of the data and it being outside of the modelled region and the other related to the ambiguity due to data overlap [182]. In this study, the ambiguity rejection is targeted. The rejection zone investigated in this study is based on a condition defined in Equation (7.12). The regions where such a condition is met can be interpreted as uncertain from the point of view of the model. Abstaining from classification there could improve overall accuracy and help to avoid errors [210]. To rely on rejection zones in classification, it has to be ensured that the probability contours reflect the data distribution in the feature space.

In the context of the compressor monitoring, if the point is in the no-classification zone, it is not used to infer about the stability and a new point is awaited. Since the points are arriving in a continuous manner, rejecting certain entries does not require any important alterations to the detection procedure.

However, it can decrease the responsiveness of the system as more data have to be collected before making a decision. Nevertheless, properly constructed rejection zones would result in rejecting points located at the class interfaces, therefore having a specific location in the feature space and not having an extreme impact on the compressor. The advantage of the rejection zone should be lowering the number of misclassifications in the regions of transition between conditions.

Figure 7.5 shows the extent of rejection zones for different levels of the threshold $t_{reject} \in \{0.1, 0.2, 0.3, 0.4\}$. They are implemented on the model described in Section 7.2. The extent of the rejection zone and the boundaries between the classes are demonstrated in each case.

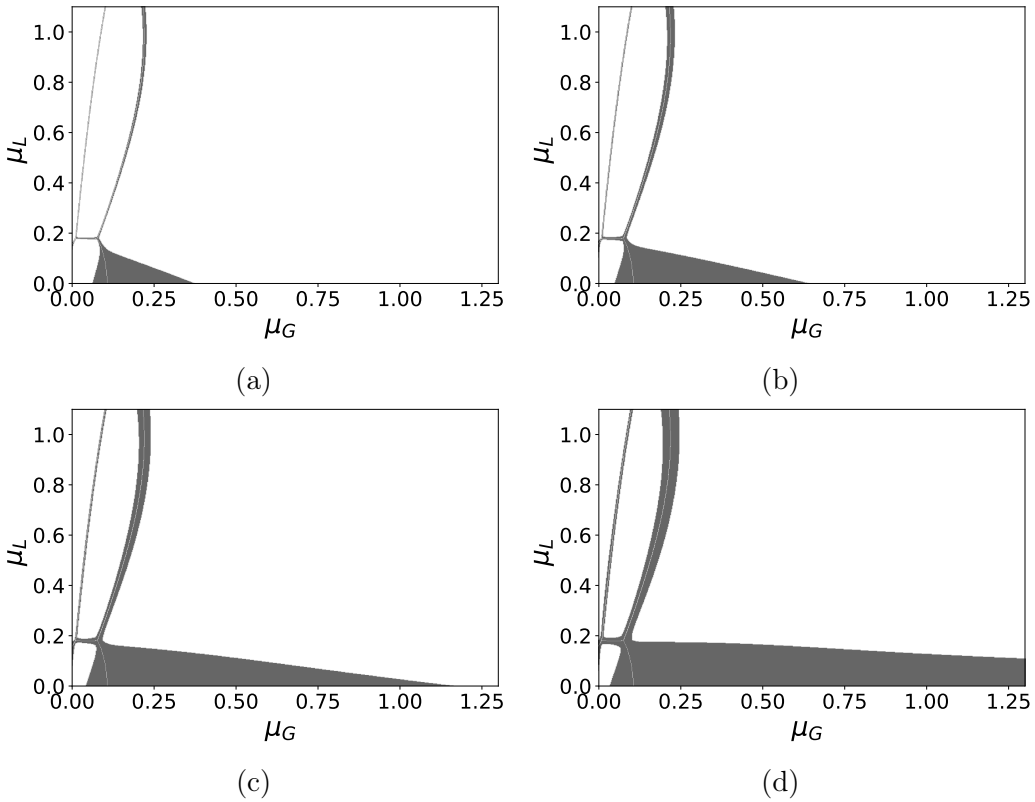


Figure 7.5: Demonstration of the rejection zones placement for a GPC model trained with 20 points for each TOA and different rejection threshold level t_{reject}, as defined by Equation (7.12) a) 0.1; b) 0.2; c) 0.3; d) 0.4; grey regions on the plot demonstrate the rejection zones, while white lines show the boundaries between classes

For the lowest value of threshold probability $t_{reject} = 0.1$, the rejection zone is present mainly in the proximity of the class boundaries as narrow regions along the boundaries, except for the interface between stable and deep surge conditions, where it forms a larger wedge. Increasing the level of t_{reject} threshold, the width of the region increases. The area of the rejection zone between stable and deep surge also increases, growing towards the high value of μ_G, where the probability of the stable class is considered high (high intensity of the green hue in Figure 7.3). Using the highest rejection threshold value importantly increases the extent of the rejection region at the bottom of the feature space, bringing the risk of rejecting some of the points for deep surge if they demonstrate a larger dispersion in μ_L feature value. At the same time, the increase in probability threshold does not result in an important growth of the rejection zone between other classes. Hence, the rejection zone based on threshold $t_{reject} = 0.3$ is selected for further analysis.

Figure 7.6 demonstrates the relation between the rejection zone and the training points. It can be seen that several points at the interface between classes fall into the rejection zone. The majority of such points are located between transient and deep surge conditions, as in that region the rejection zone is the largest. A much smaller number of points would be rejected between inlet recirculation and transient conditions and between stable and inlet recirculation. In each of the regions, the rejection zone is not large enough to cover the whole ambiguous region, where the points from different classes overlap. A further increase in threshold level did not significantly increase the width of the rejection zone between classes, hence it is not expected that for the trained GPC model it would be possible to define a rejection zone fully covering the ambiguity region without impacting the classification in other regions of the feature space.

The concept of the rejection zone is worth investigating as it has the potential to increase the accuracy of the classification. With the selected kernel type, only the ambiguity rejection is considered. Performing both ambiguity and

novelty rejection with a Gaussian process model may bring risks. GPC does not provide direct tools to differentiate the type of rejection and the shaping of the posterior probability is driven by both kernel selection and the training data. It may lead to posterior distribution that does not have a physical significance for the investigated system. The model incorporating a rejection zone should be thoroughly validated to avoid the issues of rejecting points in well-defined areas of the feature space [209].

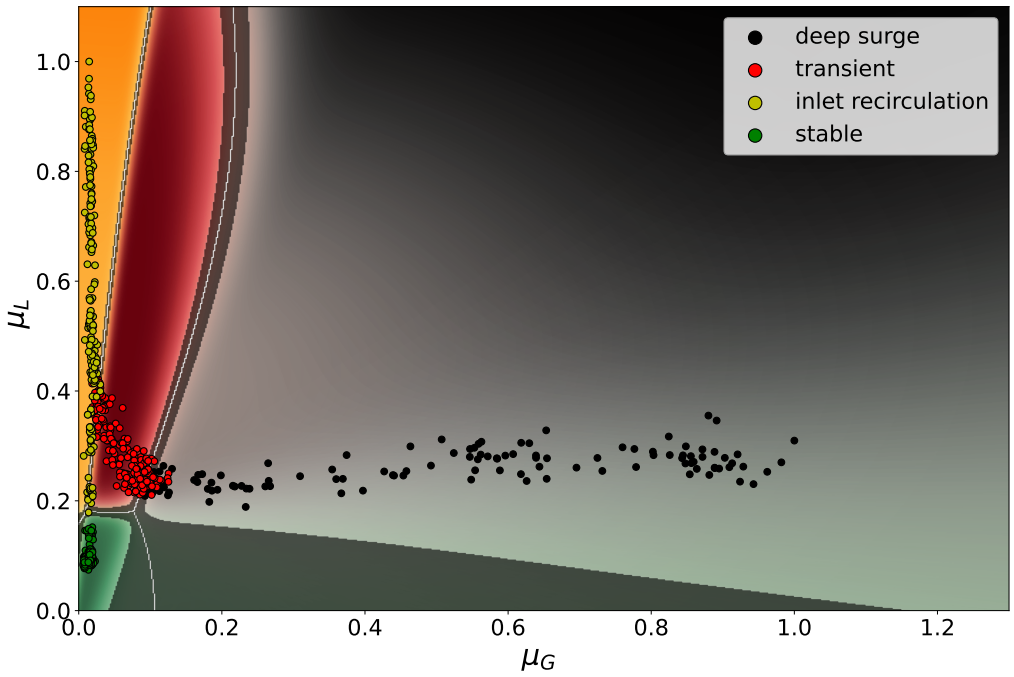


Figure 7.6: GPC model trained with 20 data points for each TOA; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold threshold $t_{reject} = 0.3$

7.4 Classification with limited training data availability

Operating the compressor in an unstable regime may bring serious risks to its integrity and operability. Some machines may not be capable of entering the

unstable region or can do so for short periods [18]. Hence, the added value of applying a classification model may come from providing accurate predictions using limited training data.

Two scenarios for limited data availability are investigated. The first is based on limiting the number of data points available for each condition. Such data could be obtained by running the compressor through the whole operating range in a short time. The limitation in compressor exposure time to unstable conditions may come from the mechanical parameters of the system [44]. Such limitations were demonstrated and discussed by Horodko [18], where the investigated industrial compressor operation in the unstable range was limited to two cycles of deep surge due to the risk of bearing system damage. It leads to a limited amount of data for the unstable operating range.

The other scenario may be related to using solely specific operating points, representing different conditions of a compressor. For a number of throttle positions, the instabilities can coexist and it may be difficult to clearly define the type of the instability present. Using the specific peaks of instabilities can help to capture a clear signature of instability. What is more, the data for the peaks of instabilities could possibly be obtained with numerical simulations of the compressor system. Facing the limitations of experimental procedures, it might be possible to obtain data for unstable regime using high-fidelity computational fluid dynamics (CFD) simulations [16]. It is not straightforward to ensure the full representativeness of the physical system with a model [211], especially in the context of data-driven decomposition techniques, where the signal composition at different time scales affects the decomposition outcomes [90]. Nevertheless, with the increasing fidelity of simulations and gains in computational power, it could be possible to complement the experimental data for experimentally safe regions with the model data for unsafe regions. CFD results are often obtained for specific operating points [16, 211] and changing the operating point requires a full rerun of the simulation. Therefore, the numerical

studies are often restricted to a few points, each representing different operating conditions. To investigate the scenario of using data for well-defined operating conditions, only the peaks of instabilities, as defined in Section 3.3.2 of Chapter 3, are used in the training of a GPC model.

7.4.1 Limited training data representing all conditions

To investigate the first scenario, the GPC model is trained using two points for each TOA value. Figure 7.7 demonstrates a probability distribution obtained with the model along with scattered training points.

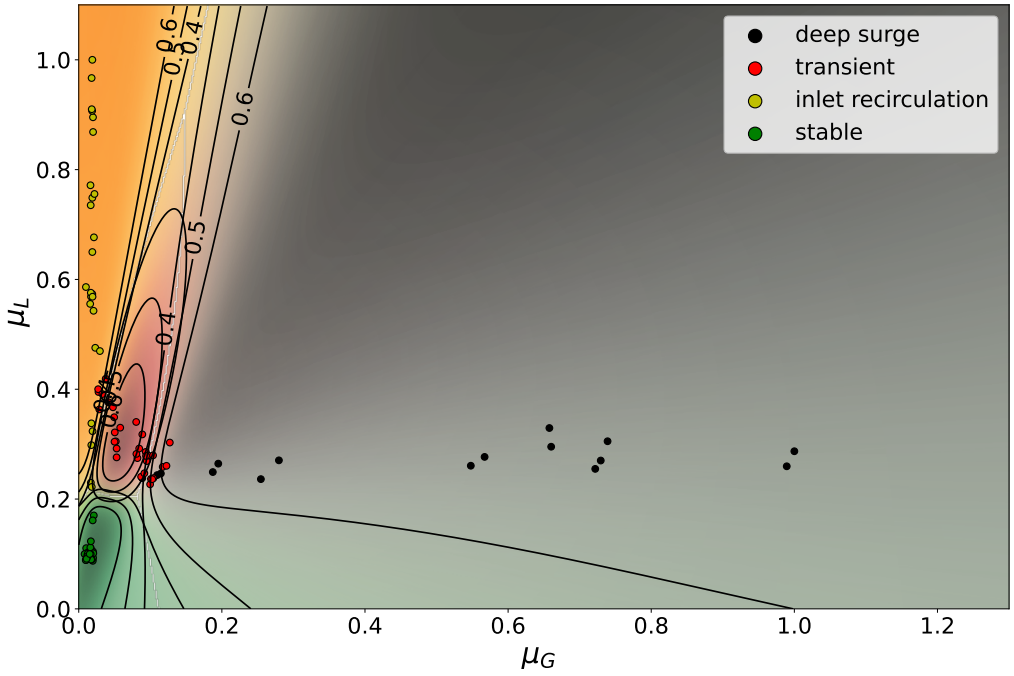


Figure 7.7: GPC results for the training data comprised of four classes of conditions. The colourmaps represent the probability for each of the classes. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points

The overall predicted shape of the classes is very similar to that obtained using a larger number of data points in training (Figure 7.3). The shape of the decision boundaries is similar in the proximity of the transitions between

conditions but differs away from the training points. With a lower number of training data, the probability contours are located further from one another. This can be understood as a lower certainty of the model at the interfaces between classes. It should lead to an increase in the size of the rejection zone if the same rejection threshold value t_{reject} is selected.

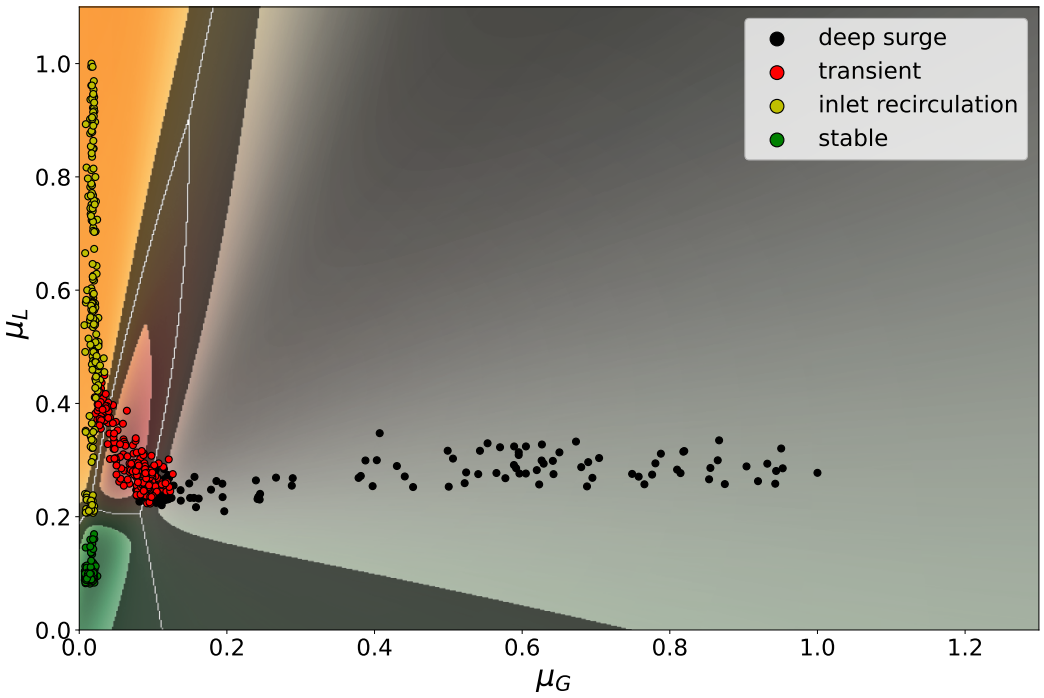


Figure 7.8: GPC model trained with data limited to two points per TOA; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold $t_{reject} = 0.3$

The classification of all data using the model built with limited training data is relatively accurate (Figure 7.8). The model shows some misclassifications for the very beginning of the inlet recirculation class and the transition zone between conditions. The transient class is smaller and less elongated, providing a tighter fit of the decision boundary around the training points. The rejection zone is much larger than for the base model, covering a substantial area of the transient class. It leads to the rejection of a higher number of points at the

boundaries between inlet recirculation and transient, as well as transient and deep surge. Furthermore, several inlet recirculation points land in the rejection zone, especially in the early stages of that instability.

7.4.2 Limited training data representing peaks of instabilities

Figure 7.9 demonstrates the outcomes of the model trained using only peaks of instabilities, as defined in Section 3.3.2. Each of the peaks is made up of 50 training points, which are scattered in the figure.

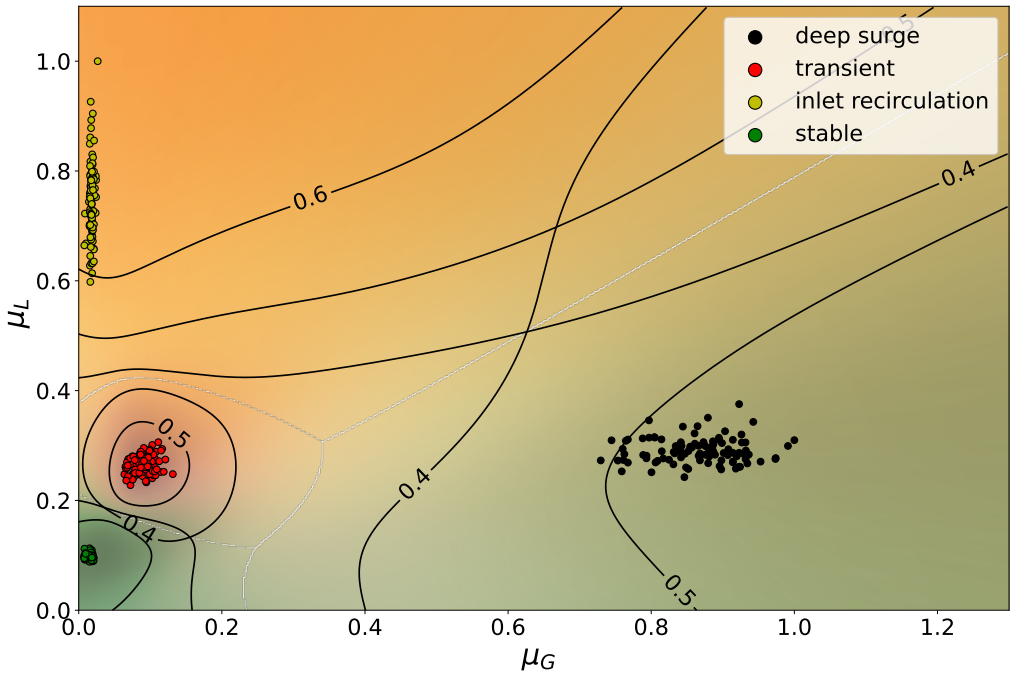


Figure 7.9: GPC results for the training data comprised of four classes of conditions. The colourmaps represent the probability for each of the classes. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points for different classes defined in the legend

The predicted shape of the classes is different from that obtained for the base case (Figure 7.3). The stable class is similar in shape and area however, the highest probability contour ($p = 0.6$) is not present for this class. It is also

not obtained for transient and deep surge classes. The inlet recirculation class is much larger and its boundary with deep surge class is shifted to the right. Using only peaks of instabilities, the transient class has a different shape. It is overall more compact but extends to the very limit of the feature space on the left. The surge class is located similarly as for the base case, but the high probability region is present for higher values of μ_G.

Taking into account the classification of the data for all operating conditions with this model (Figure 7.10), it would render inaccurate classification in the early stages of inlet recirculation. The points there would be considered to belong to the transient class. Additionally, a number of points labelled as deep surge class would be classified as transient conditions due to the shift in the location of the boundary.

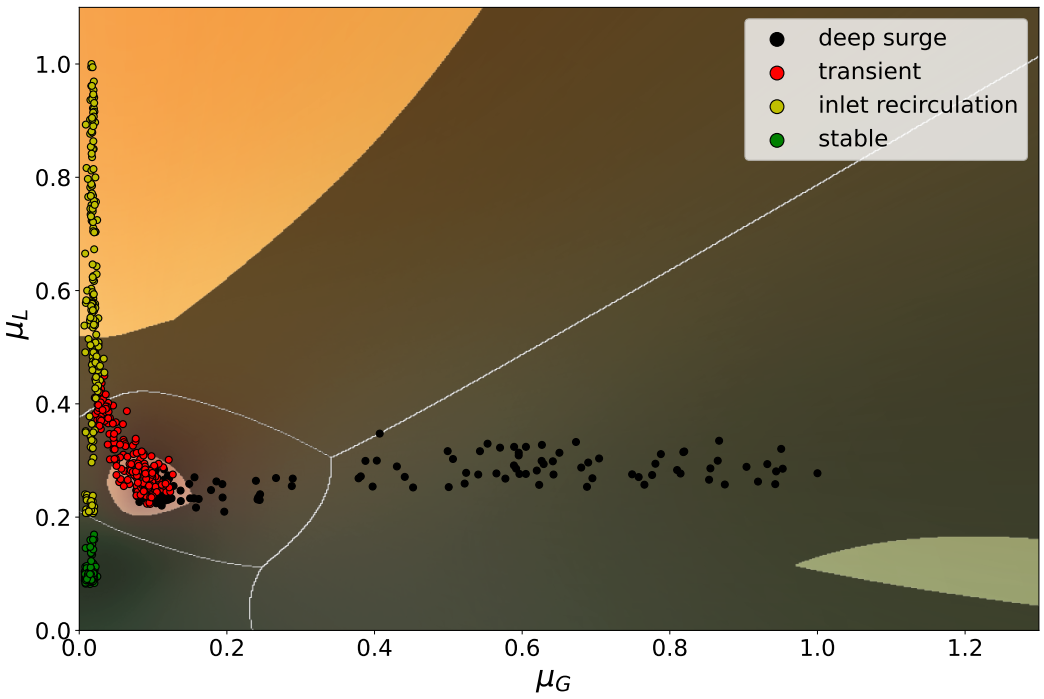


Figure 7.10: GPC model trained with data limited to peaks of instabilities; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold $t_{reject} = 0.3$

The rejection zone obtained for this model is significantly larger than for other models. It covers almost the entire investigated feature space, rejecting all of the stable points and almost all deep surge points. It indicates low differences between the probability for different classes predicted by the model. It can be explained by the fact that the probability of each class, even in the centres of training data, is low. Having only peaks of instabilities as training data, the GPC with selected kernel and defined threshold level for the rejection zone would not provide an accurate classification.

Overall, the GPC provides the possibility of building an effective classification model with a limited number of training data. It might be considered an important advantage in the field of compressors where the available data can be limited. For the model to be representative of the observed system, the data representing the whole transition path should be included in the training to obtain a meaningful and accurate model. When having only data for peaks of instabilities, it could be attempted to generate synthetic data along the expected path in the feature space, or a different classification method could be employed.

7.5 Classification with single sensor data

In Chapter 6 it was shown that the change in the operating conditions in the investigated compressor, similar to that observed with two sensors, could be demonstrated using only the sensor p_{s-imp1} before impeller. In that case, it was not possible to provide a vertical threshold for the detection of conditions as the signature of different instabilities was not fully decoupled with either of the investigated decomposition methods (see Figure 6.18).

Figure 7.11 demonstrates the GPC-based classification of compressor operating conditions using features based on RC 1 and RC 2 from SSA decomposition of the data obtained from p_{s-imp1} before impeller.

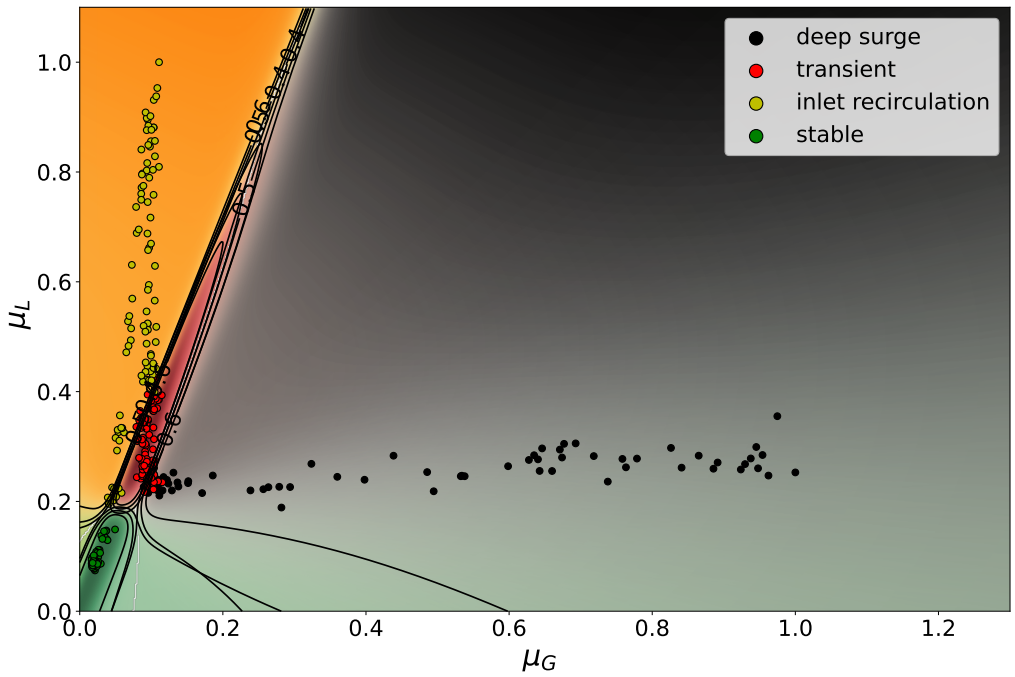


Figure 7.11: GPC results for the model trained on data from the sensor before impeller. White solid lines show the boundaries between classes; Black lines represent probability contours for probability levels $p \in \{0.4, 0.5, 0.6\}$; scattered points are training data points for different classes defined in the legend

In this case, the points for stable and inlet recirculation are located diagonally, and not on a straight vertical line as in the case of the feature space based on features from two sensors. It implies that the two features obtained from the sensor before impeller are correlated in that region and the global instability feature also experiences an increase in the initial stage of inlet recirculation. However, the points representing different classes occupy distinct locations in the feature space and can be used for classification.

The GPC model using DP^2 kernel function and features based on a signal from a single sensor was capable of providing good differentiation of operating conditions. The clusters for stable, inlet recirculation, and deep surge have similar shapes to those obtained with two sensors (see Figure 7.3). The transient class is narrower and shorter. The boundary between inlet recirculation and

transient conditions is more inclined. The contours of probability are located closely, as in the case of the model using data from two sensors.

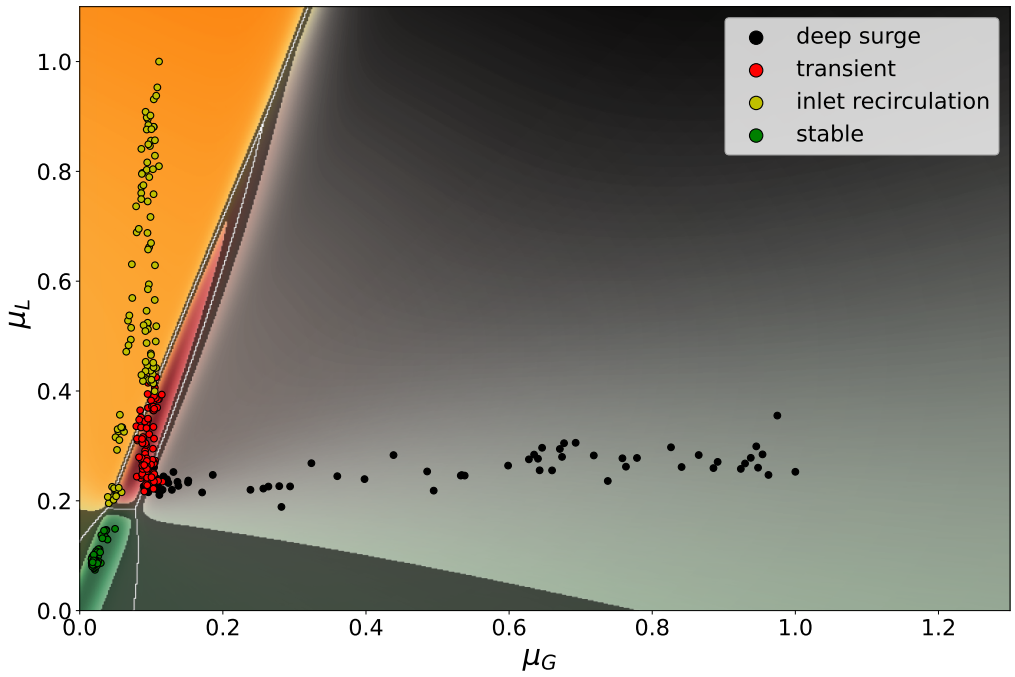


Figure 7.12: GPC model trained with data from a single sensor before impeller; the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; white lines demonstrate boundaries between classes; dark shaded regions demonstrate the rejection zone for a rejection threshold $t_{reject} = 0.3$

The location of the rejection zone is shown in Figure 7.12. Due to a quick transition of the probability predictions at class interfaces, the rejection zone is narrow close to the boundaries and incorporates only a small number of points from inlet recirculation, transient, and deep surge conditions. Using a single sensor, there is only a small difference in the placement of the points for stable and deep surge conditions, resulting in a very similar shape and size of the rejection zone near the boundary separating those conditions. By investigating the placement of the boundaries with respect to the scattered training points, there are no obvious misclassifications and possible loss of accuracy may be related to misclassification of the points close to the boundaries between conditions. Con-

sequently, it could be possible to obtain a meaningful and accurate indication of instabilities using features built on data from a single sensor located before impeller coupled with a classification algorithm such as GPC.

7.6 Accuracy of classification

The performance of GPC trained with different input data is compared using the confusion matrix. Such a matrix provides the classification accuracy for each of the classes, as well as the type of error made by the model, that is, where the misclassified points were assigned [212]. It allows understanding of the type of error made, as confusing adjacent classes would be less severe than misclassifying deep surge conditions as stable or vice versa. The confusion matrix is computed using 100 randomly selected validation data points for each TOA, with a total of 3700. The selected models are considered:

- the model trained with the data for each TOA (*Model 1*),
- the model trained with limited data for each TOA (*Model 2*)
- the model trained with the data from a single sensor obtained for each TOA (*Model 3*)

For each of those models, a rejection zone is introduced, with a rejection threshold $t_{reject} = 0.3$. The models incorporating a rejection zone are labelled with R at the end of the name, making *Model 1* with rejection to be *Model 1R*. It is assumed that the points landing in the rejection zone are not considered in the analysis and no matter if classified correctly or not, they have no impact on the accuracy.

Figure 7.13 shows the confusion matrix for *Model 1* and *Model 1R*. For Model 1, the overall accuracy is high. The stable class accuracy is 100%. The inlet recirculation and deep surge accuracies are close to 95%, while the transient class

demonstrates an accuracy below 80%. The points from transient conditions are misclassified to both adjacent classes, with a higher percentage to deep surge.

By introducing the rejection zone, the accuracy of classification for all unstable conditions increases. The largest increase is observed for transient conditions, the accuracy of which approaches 90%. The accuracy of inlet recirculation and surge also increases to 97 and 99% respectively. The majority of misclassifications are then present for surge class identified to be transient. It is possible that increasing the rejection threshold would help to further increase the accuracy, however, it might result in rejecting well-defined points, as demonstrated in Figure 7.5.

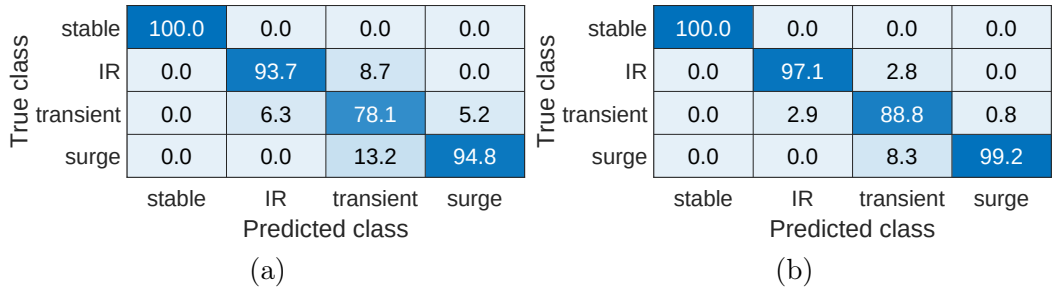


Figure 7.13: Confusion matrix for investigation of classification accuracy for the models based on all data for two sensors a) *Model 1*; b) *Model 1R* incorporating the rejection zone

Figure 7.14 shows the confusion matrix for *Model 2* and *Model 2R*. For *Model 2*, the overall accuracy is lower than for *Model 1*. The stable class accuracy is 93% and a number of points from inlet recirculation are classified as stable. The accuracy for deep surge is also 93%, while inlet recirculation and transient accuracies are close to 85%. Misclassifications are made between adjacent classes.

By introducing the rejection zone, the accuracy of classification for all conditions increases significantly. The largest increase is observed for stable and deep surge, which attain 100% accuracy. The accuracies of inlet recirculation

and transient also increase to 93 and 97% respectively. In this model, the rejection zone has a higher impact on the overall accuracy. It may be attributed to the larger extent of such zone for the same probability threshold $t_{reject} = 0.3$, which comes from more slowly varying probability at the class interfaces (Figure 7.7).

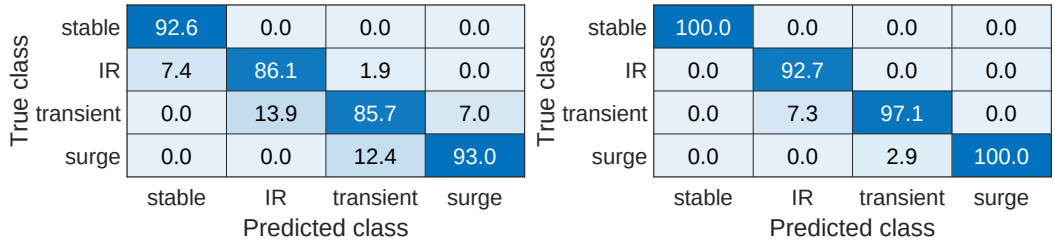


Figure 7.14: Confusion matrix for investigation of classification accuracy of the models based on limited data for two sensors a) *Model 2*; b) *Model 2R* incorporating the rejection zone

Figure 7.15 shows the confusion matrix for *Model 3* and *Model 3R*. For *Model 3*, the overall accuracy is slightly lower than for *Model 1*. The stable class accuracy is almost 100%, inlet recirculation and transient it is close to 85% and for deep sure it is 90%. The misclassifications in this case are made not only between adjacent classes but also between inlet recirculation and surge, which may result in a very early reaction of an anti-surge system.

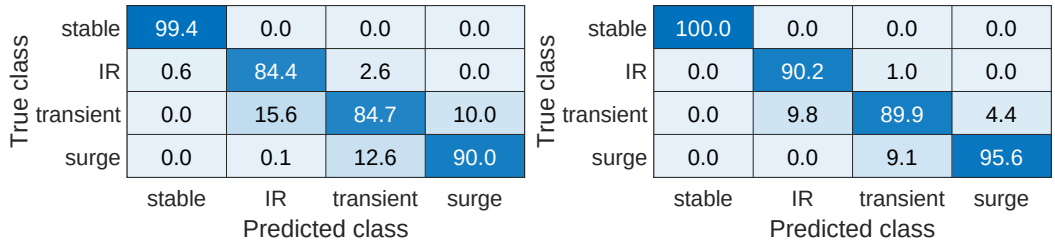


Figure 7.15: Confusion matrix for investigation of classification accuracy for the models based on all data from a single sensors a) *Model 3*; b) *Model 3R* incorporating the rejection zone

By introducing the rejection zone, the accuracy of classification for all conditions increases, but not as importantly as for the model with a limited number of training points. With the rejection zone, the accuracy of stable conditions reaches 100%, while all other accuracies increase by around 5%. Several misclassifications are still visible between adjacent classes, which demonstrates that the rejection zone does not cover the whole region of transition from one condition to another. Nevertheless, the introduction of a rejection zone can help to increase the accuracy of model classification.

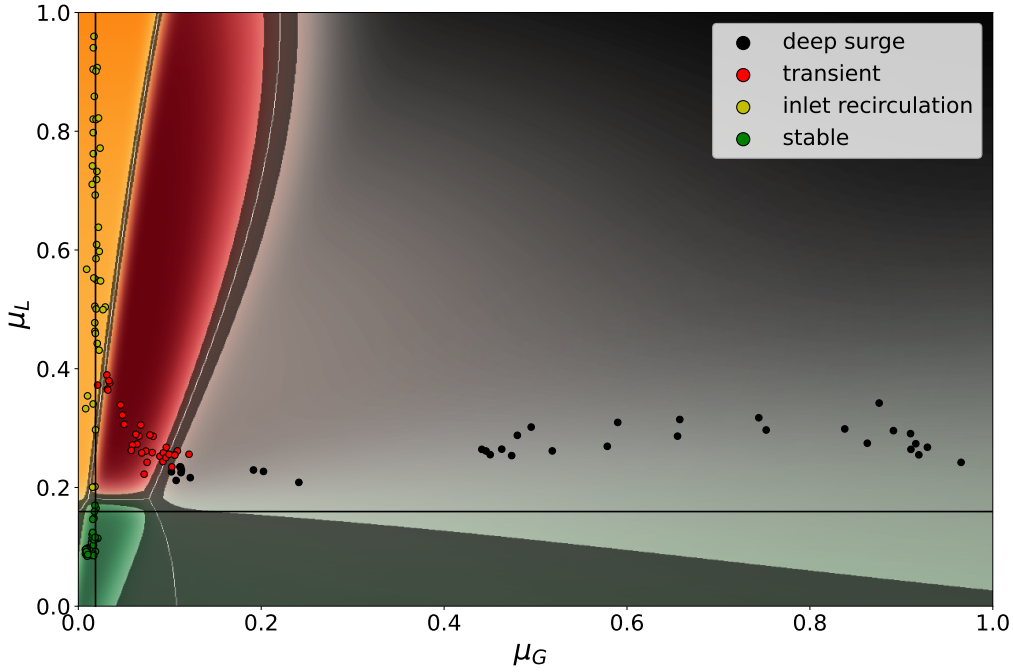


Figure 7.16: Comparison of the threshold-based model boundaries (solid black lines) with GPC-defined boundaries obtained with *Model 1R* (white solid lines); the colourmaps represent the probability for each of the classes; the scattered points represent the test data for all conditions; the shaded region demonstrates a rejection zone for rejection threshold $t_{reject} = 0.3$

Figure 7.16 demonstrates the difference between interpretable boundaries and GPC-based boundaries obtained from*Model 1R*. The comparison of boundaries for all classes is not possible, as the threshold-based approach was used only to define 3 classes while GPC was employed to differentiate 4 classes. How-

ever, the boundary between inlet recirculation and transient corresponds to the boundary between locally and globally unstable conditions.

The GPC-based boundaries between stable and inlet recirculation and inlet recirculation and transient are moved to higher μ_L and μ_G values respectively, compared to the threshold-based ones. The only exception is the region between stable and inlet recirculation, where part of the GPC-derived boundary is present for lower values of μ_G. This might suggest that the boundaries drawn with the threshold-based approach are too conservative in reference to the data. Hence, it could be considered to choose a higher percentile of the distribution or the maximum of the stable class as a threshold value. However, to do so, it should be ensured that any possible outliers present in stable region are removed.

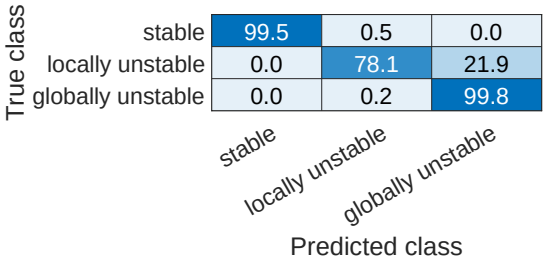


Figure 7.17: Confusion matrix between the general conditions using features based on SSA decomposition

To quantify the advantage of GPC over the threshold approach, the accuracy of both methods can be compared. The accuracy for transient and deep surge in GPC model is considered jointly for comparison. The confusion matrix for the threshold approach is recalled in Figure 7.17 and is compared with *Model 1R* for GPC, demonstrated in Figure 7.13. Overall, the introduction of GPC led to a more accurate classification of conditions. It is especially significant for the inlet recirculation region, where the accuracy of the interpretable approach was limited to 80%. It could possibly be increased by altering the placement of the threshold, as can be deduced from Figure 7.16 based on the placement of the inlet recirculation points in relation to the threshold.

The rejection zone, which is not present in the threshold-based approach, also helps to increase accuracy. With the data points present in Figure 7.16 it can be seen that to obtain high accuracy in unstable conditions differentiation, the rejection zone must be introduced as it is not possible to provide a single accurate boundary. This may be associated with the data labelling approach, which assigns a single label to all data collected for a single TOA. Such a labelling strategy should be reconsidered. Ideally, a label should be obtained individually for each point, which could change the data distribution in the feature space. Then, it is possible that the region of conditions overlap would decrease and the thresholds derived from stable data would provide even better classification accuracy. However, drawing a strong conclusion on this aspect requires further research.

7.7 Summary, discussion and conclusions

In this chapter, a Gaussian process classification (GPC) was implemented as a probabilistic classification method for the detection of instabilities. GPC was trained on a number of data sets to investigate its performance for different scenarios, relevant to the compressor instabilities detection. The aim was to obtain an accurate classification for all operating conditions and to validate the possibility of introducing rejection zones based on the probability predicted by the model.

Rejection zones could help to increase the accuracy of classification and avoid errors at class interfaces, where data for different conditions overlap in the feature space. Rejection zones were defined based on the difference of probability between classes, assuming that a low difference in probability would indicate the regions with high ambiguity. The points located in such zones would not be classified and used to infer the stability of the system. Using the rejection zone can decrease the method responsiveness, but it should only affect the regions between classes, where the operating conditions are relatively safe. The

rejection zone approach does not increase the safety of the compressor, but it instead allows one to avoid too early detection of the unstable conditions by pushing the decision boundary to a location where the instability indication is sufficiently certain.

A very important aspect of the Gaussian process classification is the selection of the kernel function [205]. It affects the behaviour of the probability distribution predicted by the model and may impose certain constraints on it [203]. In this study, a dot product squared (DP^2) kernel was used. It provided an acceptable level of flexibility in shaping the class boundaries while having the expected characteristics away from the interface between classes, as demonstrated with the synthetic one-dimensional data.

The base model for classification was trained with 20 data points for each TOA. The boundaries predicted by the model fit the data well and resulted in an accurate classification of conditions. The shaping of the boundaries resulted from the selection of the kernel function and the data. The selection of DP^2 kernel led to a few misclassifications in one region and an overly elongated shape of the transient class (Figure 7.3). It is possible that different selection of kernel function could lead to improved classification. However, when aiming for the placement of the rejection zones reflecting the uncertainty in the training data, the kernel should behave similarly to DP^2, otherwise the rejection zones may appear at the locations where the conditions are well-defined in the feature space with interpretable dimensions. It was demonstrated in a separate study, where RBF kernel was used to classify the same data [209]. Building a classification model with a rejection zone introduces the need of evaluating the posterior probability distribution, not only the accuracy of classification, making this procedure more manual.

The advantage of GPC is its flexibility in shaping the boundaries and providing accurate classification based on a limited number of data points [203]. Within this study, two different scenarios of limited training data availability

were investigated. Firstly, it was demonstrated that using only 2 points for each TOA, it is possible to obtain the classification of similar accuracy as that of the model trained with a much larger number of data points for each condition. The second scenario assumed using a larger number of training points, but representing only four selected TOA values related to peaks of instabilities. In this case, the probability distribution predicted by the model was not representative of the whole operating range, and the probability predicted by the model was low. Hence, it can be concluded that the training data should possibly represent the entire range of operating conditions rather than only the peaks of instabilities. Changing the number of training data points can also affect the probability levels predicted by the model, influencing the size and location of the rejection zone. Using the same level of probability difference, the rejection zone for *Model 1R* (Figure 7.8) was present only close to the boundaries, while for the model using only peaks (Figure 7.10) it covered almost the entire investigated area of the feature space.

Using GPC as a classification algorithm, it was demonstrated that all conditions can be accurately differentiated using data from a single sensor before impeller. By investigating two different components of SSA decomposition, it was possible to obtain feature space representation similar to that with two sensors. The same kernel function was used in GPC, but different training data led to different shaping of the probability distribution, mainly for the transient class. Using a single sensor, the rejection zone also increased the accuracy of the model. Hence, it may be stated that using the data from a single sensor, it should be possible to obtain an accurate indication of the conditions. However, a prior mapping of all operating conditions is needed, as the thresholds based on stable conditions do not provide correct classification.

GPC model is dependent on the selection of the kernel function, but the training data drive the optimisation of its hyperparameters. Hence, the outcomes of the model may be altered with the selection of training data. The

classification and computation of accuracy in this study were based on labels provided by the author of the experimental study [9]. It was demonstrated in Chapter 6 that the labelling methodology could be improved. To properly train the classification model, one should be certain of the training labels. The model prediction may importantly change with changing training data. It could be possible to craft the training data to obtain desired characteristics of the model, for example by removing training points near the boundaries to increase the uncertainty in that region and alter the size of the rejection zone. It is also possible that certain selection of training data brings different results than expected due to the interaction of data and kernel. Hence, the model training strategy should be thoroughly validated to ensure that it meets the requirements of the system, including the placement of the rejection zone.

Aside from being a potent classification algorithm, GPC has the important advantage of being extendable to an infinite number of dimensions [203]. The approach proposed in this study can be easily extended to a larger number of features, should a larger number of unstable conditions be differentiated or a method capable of differentiating transient and surge conditions be proposed. GPC could also be used for the merging of different feature spaces. Such a need may arise as a separate feature space could be obtained for each rotational speed. Then, the rotational speed could be considered another dimension of the multi-dimensional feature space. GPC could be used to define the boundaries of different conditions, including their variations due to the rotational speed. Obtaining the experimental data for such a design space with data may be challenging due to the need for a large number of tests necessary to obtain the training data. Therefore, the method capable of providing accurate indications with limited data is of value. When considering rotational speed as a feature, the selection of a kernel function should be thoroughly investigated to ensure that it can effectively represent the data distributions in each dimension.

Overall, GPC is a powerful and flexible classification method capable of improving the accuracy of instabilities detection. It offers the possibility of introducing a rejection zone to address the problem of classes overlap in the transition zones. GPC requires all of the classes to be represented in the training data and is dependent on the selection of the kernel function. The kernel selection may be case-dependent and it should be adjusted to the data but also the expected probability distribution behaviour, especially if the model is not only used for classification but also incorporates the rejection zone. The probabilistic approach can be an important component of a data-driven instabilities detection system, helping increase its accuracy compared to the threshold approach. However, the added value of a probabilistic enhancement should be critically evaluated against the cost and risk of obtaining data for model training and validation.

8 Closing remarks

This chapter summarises the main outcomes of the work presented in this thesis. It collects the most important results, reiterates the attained objectives and highlights the contributions to knowledge. The conclusions based on the presented research are drawn as well as limitations and risks associated with the proposed approach are discussed. Potential further research directions are proposed to address the challenges identified within the outcomes of this work.

8.1 Conclusions

Aerodynamic instabilities are harmful phenomena affecting the performance of centrifugal compressors and endangering their operation. The detection of instabilities is possible using pressure signals from the compressor, however, it is a challenging task. The signals are often complex, highly dynamic and contain multiple components of different origin as well as a substantial level of noise. An effective method of instabilities detection can be based on data-driven decomposition techniques. These techniques have the ability to isolate and enhance the signature of aerodynamic instabilities in the signal.

There is a limited number of contributions that have utilised data-driven decomposition methods for the detection of instabilities. The existing works rarely focus on the stochastic character of the pressure signal nor discuss the pace of detection. Both of those aspects are very important for a performant detection method. The research often focuses on specific, selected instabilities. Consequently, a comprehensive detection of different instabilities with a single method has not been proposed. This thesis intends to fill these gaps by proposing a data-driven decomposition methodology for online, robust and comprehensive instabilities detection in centrifugal compressors.

The methodology developed in this thesis allows obtaining the indication of compressor stability based on a pressure signal collected from the inside of the compressor. It has two main steps, feature extraction and decision-making. The feature extraction relies on extracting the signature of instabilities with data-driven decomposition methods. The features are constructed by taking the RMS of the selected decomposition components. Such features are related to the signal power. They were shown to have distinct values for different operating conditions, hence they can be used for the detection and differentiation of instabilities. A detailed description of the methodology is given in Chapter 2.

Two signal decomposition methods are investigated for feature extraction, empirical mode decomposition (EMD) and singular spectrum analysis (SSA). They both were shown to be capable of instabilities detection in centrifugal compressors, but the scope of research was limited. EMD and SSA represent two distinct types of methods, relying on very different approaches to decomposition. Therefore, comparing their performance on the same data set can provide valuable insights into the feasibility of feature extraction and offer guidance on the optimal methods for this specific application.

The proposed methodology can incorporate multiple features, which are combined in the feature space to be used jointly in decision-making. The decision-making step is based on supervised classification. Two classification approaches are investigated.

The first takes advantage of interpretable thresholds, defined solely based on stable data. Using this approach can reduce the risk involved in collecting data from an unstable operating range, as it only uses data from stable operation to train the model. It is shown that interpretable thresholds can provide an accurate classification of general operating conditions: stable, locally unstable and globally unstable, as defined in Section 3.3. For a threshold-based approach to be feasible, each instability has to be detected by a single feature. Each such feature has to fully isolate the signature of a specific instability. Within the threshold-based approach, some misclassifications were observed close to the boundaries between conditions, which were attributed to noise and variability of the pressure signal.

The other classification approach is based on a probabilistic classification algorithm. Gaussian process classification (GPC) method is leveraged, providing a probabilistic output in the form of the probability of each class in each location in the feature space. GPC is investigated in the context of a possible enhancement in classification accuracy due to more flexible shaping of the boundaries. The probabilistic output is also used to introduce a rejection zone,

allowing not to classify points in ambiguous regions of the feature space. For the probabilistic approach, it is necessary to obtain experimental data for all conditions and all of the classes have to be appropriately represented.

The methodology is validated using experimental data from a low-pressure centrifugal compressor. Pressure data collected with two different experimental protocols are used. In the first protocol called quasi-dynamic, the data is obtained for a set throttle valve position and a measurement is taken for each throttle setting. In the other protocol called dynamic, the data is collected for dynamically varying throttle position. The two protocols represent the different pace of changes in operating conditions. The details on data acquisition are given in Chapter 3. The data from the quasi-dynamic protocol is used to build the model and provide quantitative validation, while the data from the dynamic protocol is used for qualitative validation. By employing the data collected with two different protocols, it can be evaluated if the features extracted with either method are sensitive to the pace of stability changes. Such sensitivity may affect their detection performance for different scenarios.

It is proven that the developed methodology can be used for quick, robust and comprehensive detection of aerodynamic instabilities in centrifugal compressors. Using the threshold-based approach, it is possible to differentiate general conditions, being *stable*, *locally unstable* and *globally unstable*. The *globally unstable* conditions can be furtherly divided into *transient* and *deep surge*. Differentiating the instabilities from *globally unstable*[*] conditions is more difficult and requires obtaining not only stable data to be quick and accurate. The probabilistic approach can increase the accuracy of detection through more flexible shaping of the boundaries and the rejection zone, but it is sensitive to training data and the selection of parameters. Nevertheless, it can provide improvement in accuracy over the threshold-based approach. The next sections

[*]The *italic* notation is used here to highlight the types of conditions investigated; it is omitted further in this chapter

collate the most important observations and conclusions regarding the results of the decomposition methods, their comparison and the implementation of the probabilistic approach for conditions classification.

8.1.1 EMD-based instabilities detection

EMD was applied to a pressure signal from a centrifugal compressor in order to obtain intrinsic mode functions (IMFs) that may be used to create features sensitive to instabilities. The performance of the features depends on several parameters, including both the location of the sensor used to collect the data as well as the EMD decomposition parameters, as shown in Section 4.2.

Character of the components

The IMFs are ordered by their central frequency and cover certain frequency ranges. They are extracted in an iterative procedure, starting from the IMFs covering the highest frequency. The selection of the components to be used for obtaining features is dependent on the input signal and the parameters of the decomposition, as the content of IMFs is driven by the input data.

Influence of the decomposition parameters

The input signal length N_s affects the dispersion of the data, while the number of sifting iterations SN influences the spectral content of the IMFs, making them more narrow-banded with an increasing number of siftings. It was shown in Section 4.2.1 that the decomposition with $SN = 8$ is sufficient to obtain instabilities-sensitive features while ensuring that the decomposition time is low enough for online detection. For optimal accuracy, the input signal length N_s should be longer than the period of oscillations related to Helmholtz frequency f_H of the system in order to accurately capture the signature of surge.

Mode mixing and selection of the components

EMD is prone to mode mixing, which may be enhanced by the variable spectral content of the input signal. The information related to specific instability can travel from one IMF to another for subsequent input signals, as shown in Section 4.4.1. Consequently, the mean value of the component clearly demonstrates a reaction to instability, however, the dispersion of the component values is high. This is observed mostly for global instability, the signature of which is contained in the highest extracted IMFs. Hence, to obtain components of low dispersion for robust identification of globally unstable conditions, it is proposed to use a sum of high IMFs to construct a global instability feature. Such a feature can incorporate the entire signature of global instability and proves to be less dispersed than features based on single IMFs, as shown in Section 4.4.2.

Local and global instability features

For the investigated machine, the local instability was captured best by IMF 6 from the sensor before impeller but was also present in adjacent IMFs. The global instability was best captured by a sum of IMFs from 8 to 15 from inlet sensor, however, a similar performance was offered by the components sourced from the outlet sensor. A signal representing 0.1 seconds of machine operation was sufficient to construct the features capable of differentiating general operating conditions, being: stable, locally unstable and globally unstable.

Differentiation of transient and deep surge

Globally unstable conditions could be furtherly divided into two distinct instabilities: transient and deep surge. EMD did not allow for precise differentiation of each instability based on separate features. The signature of neither of them was fully isolated by any of the components, making threshold-based detection not feasible, as shown in Section 4.4. To address this issue, it is proposed to use a global instability feature for differentiating transient and deep surge

conditions. The feature is related to the physics of instabilities, containing a low-frequency range of the signal covering both, transient and deep surge. Consequently, the substantial change in the value of the feature at deep surge onsets can be connected with the flow structures related to instabilities.

Using the global feature for differentiating stable, transient and deep surge conditions, it is no longer possible to use a single threshold based on stable data. Consequently, it is necessary to obtain the data from transient or deep surge to define the boundaries. Then, the advantage of using only stable data to build the model is lost. However, if the compressor is allowed to operate in transient conditions, then the risk associated with running the machine at such conditions should not be excessive. Differentiating transient and deep surge can help to increase the operating range of the compressor by allowing the machine to operate in the transient region. Defining exact instabilities within the globally unstable range enables more customization of the detection system, making such a data-driven detection approach more adaptable and accurate.

8.1.2 SSA-based instabilities detection

Using SSA for the decomposition of the pressure data from a compressor, it was possible to obtain a number of reconstructed components (RCs) that could be used to construct instabilities-sensitive features. The selection of the components and performance of the features is strongly dependent on the character of the input signal and the parameters of the decomposition.

Character of the components

The RCs are ordered by their variance content, with RC 1 holding the most variance among the components. The spectral content of each RC depends on the input signal and the window length L. It is not possible to generalize the relation of the RC number and its frequency spectrum, as it was possible for

IMFs. A single RC can hold various frequency ranges and the content of RC can change with a changing structure of the signal.

Influence of the decomposition parameters

The selection of the components to be used for obtaining features is dependent on the input signal length N_s and the decomposition parameters, especially the window length L. The size of the window length defines the number of components the signal is split into. It was shown in Section 5.2.2 that by extending the window length, the signature of local instability can be split between a number of components, making its detection less clear. The same is observed for global instability, the signature of which can also be shared by a number of components.

Local and global instability features

The over-decomposition of the signal can be compensated with the grouping of the components. However, increasing the number of components for reconstruction results in a longer processing time. Incorporating a grouping stage increases the need for expert knowledge when defining features. Hence, using the window adapted to capture inlet recirculation is recommended. With $L = 50$ and the input signal length corresponding to 0.1 seconds, it is possible to differentiate local and global instabilities present in the system, as shown in Sections 5.3.1 and 5.3.2. The global instability was captured in RC 1 obtained from the inlet sensor signal, while the local instability signature was present in RC 2 from the sensor before impeller.

Differentiation of transient and deep surge

The exact differentiation of transient and deep surge signatures with separate features was not possible. The SSA did not allow to fully isolate the signature of either of those two instabilities, even when the window length was greatly increased, as demonstrated in Section 5.4.1. It can be attributed to the close

relation of transient and deep surge conditions and the joint presence of the signature of each of them in the signal. To enable quick and robust identification of transient and deep surge, it is possible to take advantage of the global instability feature, as demonstrated in Section 5.4.2. It was shown that it can be used for the differentiation of transient and deep surge conditions with good accuracy. With the onset of deep surge, there is a significant change in the value of the global instability feature. This change can serve as a clear indication of the conditions.

Using a single feature for the differentiation of stable, transient and deep surge conditions results in the need to obtain feature values for at least one of the unstable conditions. Consequently, the advantage of using only stable data to build a model is lost. However, such an approach offers a more precise definition of the conditions. It can be useful for extending the operating range of the compressor, should the presence of transient conditions be allowable for the machine.

8.1.3 Comparison of EMD and SSA for instabilities detection

To understand the performance of the decomposition methods, local and global instability features obtained using EMD and SSA were compared. The emphasis was put on quantifying the detection potential and identifying the detection errors as well as the differences in features obtained with both methods. The stable, locally unstable and globally unstable conditions were differentiated using thresholds based solely on stable data.

Local and global instability features

Both EMD and SSA allow obtaining an accurate indication of locally unstable and globally unstable conditions of the compressor with the threshold approach. The local instability features μ_L obtained using either of the methods display

an almost identical behaviour across the entire operating range. They both encompass a frequency range that is characteristic of the inlet recirculation observed within the machine under investigation. The global instability features from both methods are also very similar for all conditions. It regards both their values as well as their spectrum, containing mostly sub-synchronous frequencies. Using the decomposition parameters as defined in this study, it is possible to obtain an online indication of general operating conditions with each of the methods as shown in Section 6.2.

Differentiation of transient and deep surge

The global instability feature can be used for the differentiation of transient and deep surge for each of the methods. The differentiation is based on the important change of the feature value between transient and deep surge. However, then the threshold based on stable data can no longer be used. The differentiation of transient and deep surge can be considered a higher level of detection accuracy. When detecting only the general conditions, the operation of the compressor would be interrupted when entering the globally unstable range. By differentiating transient and deep surge from the globally unstable operation, the interruption can be shifted to the deep surge region. Consequently, the operating range of the machine can be increased, provided that operation in transient range is safe for the compressor. The effect of the transient instability on a compressor would be dependent on the type of machine. It is possible that for some compressors, the operation in the transient range would be acceptable. Consequently, differentiation of transient and deep surge would allow the extension of the operating range.

Features dispersion

Comparing the features based on EMD and SSA, a difference in their dispersion can be observed, as shown in Sections 6.1.1 and 6.1.2. The SSA-based features

are more concentrated around the mean than EMD-based ones for the same input length. This is noticeable for both, local and global instability features. It can possibly be attributed to mode mixing present in EMD. The information contained in a single RC obtained with SSA spreads between different IMFs obtained with EMD, resulting in the dispersion of the features based on those IMFs. The lower dispersion for SSA-based features transfers into more compact clusters in the feature space. Both methods have thresholds that are defined using stable conditions. However, because EMD-based features have a larger dispersion, the thresholds in the EMD-based approach are shifted to higher values. The placement of thresholds influences the classification of the conditions.

Accuracy of the threshold-based classification

The accuracy of classification is computed using quasi-dynamic data and shown in Section 6.2.2. The indications obtained with the EMD and SSA-based methods are compared with the labels assigned by the author of the experimental study [9]. The overall accuracy of classification is higher using EMD-based features. It is mainly due to higher accuracy in the locally unstable range. However, the accuracy for the globally unstable range is lower than in the case of SSA. Considering the nature of instabilities in centrifugal compressors, the importance of misclassification may be different, depending on its type. Misclassifying a globally unstable operation for a locally unstable one is highly dangerous, as it may allow the compressor to operate at deep surge conditions. Misclassifying locally unstable conditions as globally unstable ones can cause unnecessary interruptions in the compressor, but it does not pose any risk to the machine.

When considering classification accuracy for specific TOAs, as demonstrated in Section 6.2.2 the number of misclassifications for SSA steadily increases with decreasing TOA. It could be related to a temporal presence of globally unstable conditions for the TOAs labelled as locally unstable. It is to some extent

observed for EMD as well, but not as clearly due to different placement of the threshold.

To have greater confidence about the conditions present in the compressor, the approach to data labelling should be reconsidered. Originally, a single label was assigned for each TOA. For the TOAs close to the boundary between different conditions, it is possible that different unstable structures are temporarily present in the compressor. Hence, a label should be provided for each short portion of the input signal, rather than in bulk for each TOA. Then, it would be possible to quantify the performance of data-driven instabilities detection methods with higher confidence.

Representativeness for dynamic change of stability

To investigate the discrepancies between features obtained with two experimental protocols, the features from each of them are scattered onto the feature space (Section 6.2.2). Comparing their relative location, the SSA-based features demonstrate far less difference between protocols than EMD-based features. The SSA-based features for dynamic protocol match very well the representation build with quasi-dynamic data. In the case of EMD, an offset is observed in the global instability feature value. Extracting the components from the highest to lowest frequencies, EMD seems to be more sensitive to the character of the input signal. The offset of points for different data acquisition protocols may indicate that EMD-based features are unfit for instabilities detection with the proposed approach. The classification model trained with quasi-dynamic data, would not be representative for a quick transition in stability, emulated with dynamic data.

Risks associated with adaptive base of the decomposition

The points in the feature space for SSA show some sub-clusters for stable conditions. Those sub-clusters have the same value of μ_L, but different value of

μ_G. This is related to the adaptability of SSA, where for some input signals RC 1 represents an entirely different frequency range, as shown in Section 6.1.2. The shift is only observed for the inlet sensor, probably due to a lack of strong, dominating phenomenon in the flow for those conditions. Such behaviour is not observed for the unstable range, where the signal is more dominated by the signature of instabilities. The presence of sub-clusters does not invalidate the proposed approach. However, when implementing the methodology, it should be validated if such a phenomenon takes place and how much it can impact the detection. It is possible that for a different machine, the shift would be more significant, altering the placement of the threshold for the global instability feature.

Limitations of the linear threshold approach

Using the linear thresholds for the obtained features distribution, it is not possible to avoid misclassifications. It is especially true for the global instability threshold T_{μ_G}, where the points for locally unstable and globally unstable range overlap. The misclassifications can be caused by the incorrect labelling ot the data. Nevertheless, due to the presence of noise in the signal and small distance between subsequent TOAs, some of the points would still overlap in the feature space. To address the problem of misclassifications in the regions of transition, a probabilistic approach is proposed.

8.1.4 Probabilistic classification enhancement

The probabilistic classification aims to address the problem of misclassifications at the transition between conditions and quantify how a performant classifier, offering highly flexible, data-driven boundaries between classes can help to ameliorate the accuracy of instabilities detection. It is based on Gaussian process classification (GPC), which provides the estimated class probability for each class in all locations of the feature space. The probabilistic output is used

for the classification of the operating conditions as well as the formulation of rejection zones, defined based on the relation of probability for different classes.

Kernel function selection

One of the crucial aspects of the GPC model is the selection of a kernel function. The kernel defines the relation between points in the feature space. The choice of the kernel affects the probability distribution, including the flexibility of its shaping and behaviour away from training data. In this study, a multiplication of two dot product kernels is used. It provided sufficient flexibility to model the compressor data while generating an interpretable posterior probability distribution, where the probability of instability increases with an increase of the feature value.

Rejection zone effects

Using the posterior probability distribution, it is possible to define rejection zones. They are present in the locations where the difference in probability of the two most probable classes is lower than a specified threshold. Section 7.6 demonstrates that such zones help to increase the accuracy of classification by rejecting the points present at the interfaces between conditions. When a point is rejected, it is not considered by the detection system and a new point is awaited. The location and size of the rejection zones are dependent on the selection of the threshold value, type of kernel and the training data set. The rejection zones may offer an advantage for classification, but their placement should be thoroughly reviewed.

Training data requirements

The strength of a Gaussian process as a classifier lies, among others, in the capacity of accurate classification using limited training data. To investigate this in the context of a centrifugal compressor, two limited-data scenarios are proposed, as shown in Section 7.4. The first one uses data representing all TOAs

with a low number of points, while the other is based solely on data for selected TOAs, representing the peaks of instabilities. It is shown that the first scenario offers much better classification, even using as little as 2 points for each TOA. Considering the need to acquire data for unstable conditions, this can help to limit the time the compressor has to operate in such conditions for generating an appropriate training data set.

Using single sensor data for detection

The probabilistic approach is also used to investigate the accuracy of conditions classification using the data from a single sensor located before impeller. The features were obtained using SSA decomposition. Similarly to the setup using two sensors, RC 1 was a base for the global instability feature while RC 2 was used for the local instability feature. It was demonstrated that a high level of accuracy can be obtained when using data from a single sensor. This indicates that data-driven decomposition methods can effectively decouple the signature of different conditions from the same signal. It can also be concluded that the location before impeller may be important for obtaining a comprehensive indication of machine stability, as it captures the signature of both local and global instability.

The GPC is a powerful approach that can provide high accuracy of classification using a single sensor and limited training data. It is a great advantage for an instabilities detection system as it can both limit the cost of such a system as well as greatly limit the risk of machine damage due to shortening the required data acquisition in the unstable range.

8.2 Limitations of the study

Although data-driven decomposition methods have shown promising results for detecting instabilities, there are various limitations that need to be ad-

dressed. These limitations pertain to both the methodology itself and the scope of method validation presented in the case study.

Selection of the components

The selection of the components is a challenging procedure that requires expert knowledge and can vary between compressors. In this study, the components were selected by investigating their values in the unstable range. Knowing only the stable operating range of a new compressor system would make the selection of components significantly more challenging, potentially leading to increased uncertainty. It is especially true for SSA, where the components order is driven by the variance content. Proposing an automated selection of the components based on the stable range would provide a significant improvement to the proposed methodology.

Adaptable base of the decomposition

The data-driven decomposition methods used in this study do not have a fixed decomposition base. Therefore, the decomposition varies with the input signal character. Data-driven methods may offer more accurate detection and extract better features compared to fixed-based decompositions, however, they are sensitive to the selection of the parameters and the length of the input signal. The information of interest may move between the components, which was observed for both EMD and SSA. It makes the monitoring of the phenomena subjected to higher uncertainty and again, complicated the task of components selection.

Isolation of instabilities signature

It was not possible to obtain features isolating solely transient or deep surge signature. Hence, it was not possible to differentiate those two instabilities with a threshold-based approach neither with EMD nor with SSA. Using the global instability feature for differentiation of stable, transient and deep surge conditions requires obtaining data for unstable compressor operation. It brings

the risk of compressor damage when obtaining the training data. It is possible that by using a different decomposition method, the signature of transient and deep surge could be decoupled. However, it appears that focusing on the power of selected frequencies, the separation of transient and deep surge may not be possible.

Types of observed instabilities

The machine used in the case study was a low-pressure compressor. It did not experience rotating stall, which is commonly reported in centrifugal compressors. The presence of rotating stall could affect the components selection and requirements for the decomposition parameters. Stall is often manifested in sub-synchronous frequencies, covered by the global instability feature defined in this study. To isolate the signature of rotating stall, it may be necessary to use a longer window length in the case of SSA or to approach the construction of the global feature differently if using EMD. Depending on the type of rotating stall, a different sensor location may need to be used. In principle, the proposed methodology should soundly accommodate any number of additional instabilities, however, the performance of the components and the approach to the selection of the sensor location may alter as it may no longer be possible to differentiate all conditions using data from a single sensor.

Range of rotational speeds

The data used in this study was limited to a single rotational speed. Hence, the proposed methodology was not validated for different rotational speed, not for dynamically varying rotational speed. It should be confirmed that a specific component or set of components is capable of capturing the signature of unstable conditions across the whole range of rotational speeds. Having the feature space representation for a number of fixed rotational speeds, they would need to be merged to allow for the instabilities detection at various rotational speeds. The

merging could be performed with GPC, however, the performance of such a model would have to be thoroughly validated.

8.3 Further work

Several different paths could be undertaken to complement the research presented in this thesis, confirm the presented findings and address the identified challenges.

Validation for other compressors

The methodology proposed in this thesis should be validated on other machines, possibly of different types and sizes. It could provide a better understanding of the approach to components selection and decomposition parameters as well as prove the universality of the methodology. The studies should be performed for a number of rotational speeds and the suggested approach to merging data should be explored. It is also important to account for instabilities detection during speed transients, that may induce the onset of instabilities.

Feature extraction alteration

Further work could focus on the feature extraction block of the methodology. The very similar behaviour of the features obtained with entirely different methods might suggest that similar results could be obtained also with other decomposition methods, such as intrinsic time-scale decomposition or variational mode decomposition. Different decomposition could possibly provide features with lower dispersion or generate the components capable of differentiating transient and deep surge using a threshold built on stable data. The data-driven approaches with variable decomposition base should be critically compared with fixed-base decomposition methods such as synchrosqueezing transform to quantify the advantage of the adaptability offered by the data-driven methods.

It is also possible to alter the way the decomposition components are processed into features. The RMS value proposed in the methodology is related to signal power, which should change with the onset of instabilities. However, due to the character of the components, it does not allow the decoupling of the signature of instabilities that overlap in the frequency spectrum. A promising metric may be based on entropy, which evaluates the complexity of the signal. Due to their different nature, the entropy-based features may be more sensitive to some conditions than power-related features.

Automated components selection

The automated selection of the components could be investigated in further research. Such selection could be for example based on their frequency content. However, it can be challenging when working with SSA and similar methods, as there is no clear connection between the component number and its frequency spectrum. It could also be possible to aid the selection of the components with some modelling techniques, such as CFD to obtain the signal for unstable conditions without risking the damage to the compressor. In general, to ensure the universality of the automatic components selection, the proposed method should be validated on multiple machines. To automate the selection of components, the crucial step is gathering a comprehensive dataset representing varying compressor designs and rotational speeds. This will ensure generalisation of the components behaviour, which should allow for defining an automated selection procedure.

Enhancements to decision-making

Different rules for setting the threshold for an interpretable approach could be explored. It is possible to vary the rules regarding decision-making, for example, requiring two points in a row to be above the threshold for making a decision. When considering the probabilistic approach, it may be beneficial to

investigate more sophisticated kernel functions that provide greater flexibility in data modelling. Such a need may arise when attempting to accommodate the data for varying rotational speeds.

Building a comprehensive data set

The focus could also be put on experimental data labelling automation. It requires appropriate instrumentation of the test rig, and should therefore be considered at an early stage of its design. The instabilities can have a different impact on the compressor structure, hence the labelling should aim to reflect the instability effect on the machine. It could be possible with the measurements of the structure vibration or strain on crucial compressor components. Sufficient resolution of the labelling should be used to ensure that the variation of stability for the same setting of the throttle can be captured. Such a data set would be very useful for the validation of different data-driven instabilities detection approaches.

Bibliography

[1] M. Casey and C. Robinson, *Radial Flow Turbocompressors*. Cambridge University Press, 2021.

[2] B. Ribi, "Lecture Series 1996-01 FLOW IN RADIAL TURBOMACHINES," tech. rep., von Karman Institute for Fluid Dynamics, 1996.

[3] Y. Lin, T. Fan, and X. Zheng, "Roles of recirculating bubble on the performance of centrifugal compressors," *Aerospace Science and Technology*, vol. 118, 2021.

[4] Unknown, "Photograph of a airfoil in a wind tunnel, showing separated flow over the top surface.," 1915.

[5] E. Lennemann and J. H. Howard, "Unsteady flow phenomena in rotating centrifugal impeller passages," *Journal of Engineering for Gas Turbines and Power*, vol. 92, no. 1, pp. 65–71, 1970.

[6] E. M. Greitzer, "The Stability of Pumping Systems—The 1980 Freeman Scholar Lecture," *Journal of Fluids Engineering*, vol. 103, no. 2, pp. 193–242, 1981.

[7] R. Van den Braembussche, *Design and Analysis of Centrifugal Compressors*. John Wiley and Sons, 2019.

[8] X. He and X. Zheng, "Flow instability evolution in high pressure ratio centrifugal compressor with vaned diffuser," *Exp. Therm. Fluid Sci.*, vol. 98, no. May, pp. 719–730, 2018.

[9] G. Liskiewicz, *Numerical model of the flow phenomena preceding surge in the centrifugal blower and assessment of its applicability in designing anti-surge devices.* PhD thesis, University of Strathclyde, 2014.

[10] K. H. Lüdtke, *Process Centrifugal Compressors.* Springer, 2004.

[11] H. Bloch, *A practical guide to compressor technology.* John Wiley and Sons, 2006.

[12] R. C. Pampreen, *Compressor surge and stall.* Concepts ETI inc., 1993.

[13] R. Mailach, I. Lehmann, and K. Vogeler, "Rotating instabilities in an axial compressor originating from the fluctuating blade tip vortex," *J. Turbomach.*, vol. 123, no. 3, pp. 453–460, 2001.

[14] H. Wang, Y. Wu, Y. Wang, and S. Deng, "Evolution of the flow instabilities in an axial compressor rotor with large tip clearance: An experimental and urans study," *Aerospace Science and Technology*, vol. 96, p. 105557, 2020.

[15] B. Semlitsch and M. Mihăescu, "Flow phenomena leading to surge in a centrifugal compressor," *Energy*, vol. 103, pp. 572–587, 2016.

[16] E. Sundström, B. Semlitsch, and M. Mihăescu, "Generation Mechanisms of Rotating Stall and Surge in Centrifugal Compressors," *Flow, Turbulence and Combustion*, vol. 100, no. 3, pp. 705–719, 2018.

[17] R. H. Aungier, *Centrifugal Compressors: A Strategy for Aerodynamic Design and Analysis.* ASME Press, 2000.

[18] L. Horodko, "Zastosowanie czasowo-częstotliwościowej analizy sygnałów do badania niestatecznej pracy sprężarki promieniowej," *Zeszyty Naukowe. Rozprawy Naukowe/Politechnika Łódzka*, no. 353, pp. 4–114, 2006.

[19] Z. Niu, Z. Sun, B. Wang, and X. Zheng, "Effects of Nonaxisymmetric Volute on Rotating Stall in the Vaneless Diffuser of Centrifugal Compressors," *J. Eng. Gas Turbines Power*, vol. 144, no. 5, pp. 1–10, 2022.

[20] M. Cicciotti, *Adaptive Monitoring of Health-State and Performance of Industrial Centrifugal Compressors*. PhD thesis, Imperial College London, 2015.

[21] K. Brun and M. Nored, "Guideline for Field Testing of Gas Turbine and Centrifugal Compressor Performance," *Machinery*, no. August, 2006.

[22] E. M. Greitzer, "Surge and Rotating Stall in Axial Flow Compressors - 1, 2.," *Am. Soc. Mech. Eng.*, no. 75 -GT-9, pp. 190–198, 1975.

[23] C. Riegler, M. Bauer, and J. Kurzke, "Some aspects of modelling compressor behavior in Gas turbineperformance calculations," *J. Turbomach.*, vol. 123, no. 2, pp. 372–378, 2001.

[24] R. C. White and R. Kurz, "Surge avoidance for compressor systems," *35th Turbomach. Symp.*, pp. 123–134, 1988.

[25] G. J. Skoch, "Experimental investigation of centrifugal compressor stabilization techniques," *J. Turbomach.*, vol. 125, no. 4, pp. 704–713, 2003.

[26] R. Kurz, R. Zwerver, S. Simons, A. Alvarado, K. Brun, *et al.*, "Dynamic simulation and testing to assess rundown speed of a compressor," in *Asia Turbomachinery & Pump Symposium. 2018 Proceedings.*, Turbomachinery Laboratory, Texas A&M Engineering Experiment Station, 2018.

[27] N. A. Cumpsty, "Compressor aerodynamics," 1989.

[28] C. Schreiber, *Inlet Recirculation in Radial Compressors*. PhD thesis, University of Cambridge, 2017.

[29] J. März, C. Hah, and W. Neise, "An experimental and numerical investigation into the mechanisms of rotating instability," *Journal of Turbomachinery*, vol. 124, no. 3, pp. 367–374, 2002.

[30] N. Kämmer and M. Rautenberg, "A distinction between different types of stall in a centrifugal compressor stage," *Journal of Engineering for Gas Turbines and Power*, vol. 108, no. 1, pp. 83–92, 1986.

[31] R. Dehner, P. Sriganesh, A. Selamet, and K. Miazgowicz, "Generation mechanism of broadband whoosh noise in an automotive turbocharger centrifugal compressor," *Proceedings of the ASME Turbo Expo*, vol. 8, pp. 1–12, 2020.

[32] J. März, C. Hah, and W. Neise, "Closure to "discussion of 'an experimental and numerical investigation into the mechanisms of rotating instability'"(2002, asme j. turbomach., 124)," *J. Turbomach.*, vol. 124, no. 3, pp. 375–375, 2002.

[33] P. X. Harley, S. W. Spence, J. Early, D. Filsinger, and M. Dietrich, "Inlet recirculation in automotive turbocharger centrifugal compressors," *Institution of Mechanical Engineers - 11th International Conference on Turbochargers and Turbocharging*, pp. 89–100, 2014.

[34] N. Poujol, P. Duquesne, and I. Trébinjac, "Experimental and Numerical Analysis of Impeller Recirculation: Inlet Guide Vanes Stagger Angle Effects and Rotating Disturbances Interaction," *14th European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, ETC 2021*, pp. 0–13, 2021.

[35] J. F. Gülich, *Centrifugal Pumps*. Springer Berlin Heidelberg, 2008.

[36] B. Eck, *Ventilatoren*. Springer, 1972.

[37] S. Mizuki, T. Hattori, I. Ariga, and I. Watanabe, "Reversed Flow Phenomena Within Centrifugal Compressor Channels At Lower Flow Rate.," *American Society of Mechanical Engineers (Paper)*, no. 76 -GT-86, 1976.

[38] Y. Ribaud, "Experimental Aerodynamic Analysis Relative To Three High Pressure Ratio Centrifugal Compressors.," *American Society of Mechanical Engineers (Paper)*, pp. 1–8, 1987.

[39] X. Qiu, D. Japikse, and M. Anderson, "A meanline model for impeller flow recirculation," *Proceedings of the ASME Turbo Expo*, vol. 6, no. PART B, 2008.

[40] M. Schleer, S. J. Song, and R. S. Abhari, "Clearance effects on the onset of instability in a centrifugal compressor," *Journal of Turbomachinery*, vol. 130, no. 3, pp. 1–11, 2008.

[41] R. J. McKee, S. P. Siebenaler, and D. M. Deffenbaugh, "Increased Flexibility of Turbo-Compressors In Natural Gas Transmission Through Direct Surge Control," tech. rep., Southwest Research Institute, 2005.

[42] P. Harley, S. Spence, D. Filsinger, M. Dietrich, and J. Early, "Meanline modeling of inlet recirculation in automotive turbocharger centrifugal compressors," *Journal of Turbomachinery*, vol. 137, no. 1, pp. 1–9, 2015.

[43] Z. Sun, W. Zou, and X. Zheng, "Instability detection of centrifugal compressors by means of acoustic measurements," *Aerospace Science and Technology*, vol. 82-83, pp. 628–635, 2018.

[44] U. Haupt, M. Rautenberg, and A. N. Abdel-Hamid, "Blade excitation by broad-band pressure fluctuations in a centrifugal compressor," *Journal of Turbomachinery*, vol. 110, no. 1, pp. 129–137, 1988.

[45] G. Liskiewicz and L. Horodko, "Time-frequency analysis of the Surge Onset in the Centrifugal Blower," *Open Engineering*, vol. 5, no. 1, pp. 299–306, 2015.

[46] J. Andersen, F. Lindström, and F. Westin, "Surge definitions for radial compressors in automotive turbochargers," *SAE International Journal of Engines*, vol. 1, no. 1, pp. 218–231, 2009.

[47] A. Logan, D. G. Cava, and G. Liśkiewicz, "Singular spectrum analysis as a tool for early detection of centrifugal compressor flow instability," *Measurement: Journal of the International Measurement Confederation*, vol. 173, 2021.

[48] I. J. Day, "Stall, surge, and 75 years of research," *J. Turbomach.*, vol. 138, no. 1, pp. 1–16, 2016.

[49] P. Frigne and R. Van Den Braembussche, "Distinction Between Different Types of Impeller and Diffuser Rotating Stall in a Centrifugal Compressor With Vaneless Diffuser," *Journal of Engineering for Gas Turbines and Power*, vol. 106, no. 2, pp. 468–474, 1984.

[50] F. Grapow, *Compressor Vaneless Diffuser Rotating Stall: Instability Development and Control*. PhD thesis, Lodz University of Technology, 2021.

[51] S. Mizuki, Y. Kawashima, and I. Ariga, "Investigation concerning rotating stall and surge phenomena within centrifugal compressor channels," in *Turbo Expo: Power for Land, Sea, and Air*, vol. 79719, p. V01AT01A009, American Society of Mechanical Engineers, 1978.

[52] P. Frigne and R. Van den Braembussche, "Distinction between different types of impeller and diffuser rotating stall in a centrifugal compressor with vaneless diffuser," in *American Society of Mechanical Engineers,*

International Gas Turbine Conference and Exhibit, 28 th, Phoenix, AZ, 1983.

[53] R. Hunziker and G. Gyarmathy, "The operational stability of a centrifugal compressor and its dependence on the characteristics of the subcomponents," *Journal of Turbomachinery*, vol. 116, no. 2, pp. 250–259, 1994.

[54] P. Jenny and Y. Bidaut, "Experimental determination of mechanical stress induced by rotating stall in unshrouded impellers of centrifugal compressors," *Journal of Turbomachinery*, vol. 139, no. 3, 2017.

[55] F. Grapow, K. Olasek, G. Liśkiewicz, R. Magiera, and W. Kryłłowicz, "Experimental study of vaneless diffuser rotating stall development and cell-merging phenomena," *Journal of Turbomachinery*, vol. 143, no. 5, 2021.

[56] A. Bianchini, D. Biliotti, M. Giachi, E. Belardini, L. Tapinassi, L. Ferrari, and G. Ferrara, "Some guidelines for the experimental characterization of vaneless diffuser rotating stall in stages of industrial centrifugal compressors," *Proceedings of the ASME Turbo Expo*, vol. 2D, pp. 1–12, 2014.

[57] L. Horodko and W. Kryllowicz, "Investigation of the rotating stall in a centrifugal compressor," *Am. Soc. Mech. Eng. Fluids Eng. Div. FED*, vol. 257, no. 1 B, pp. 1547–1551, 2002.

[58] A. Bianchini, G. Andreini, L. Ferrari, D. T. Rubino, and G. Ferrara, "Development of a criterion for a robust identification of diffuser rotating stall onset in industrial centrifugal compressors," *Journal of Engineering for Gas Turbines and Power*, vol. 141, no. 2, pp. 1–7, 2019.

[59] Z. Spakovszky and C. Roduner, "Spike and modal stall inception in an advanced turbocharger centrifugal compressor," *Journal of Turbomachinery*, vol. 131, no. 3, p. 031012, 2009.

[60] L. Horodko, "Identification of rotating pressure waves in a centrifugal compressor diffuser by means of the wavelet cross-correlation," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 4, no. 2, pp. 373–382, 2006.

[61] F. Botero, V. Hasmatuchi, S. Roth, and M. Farhat, "Non-intrusive detection of rotating stall in pump-turbines," *Mechanical Systems and Signal Processing*, vol. 48, no. 1-2, pp. 162–173, 2014.

[62] C. A. Niccolini Marmont Du Haut Champ, P. Silvestri, M. L. Ferrari, and A. F. Massardo, "Incipient surge detection in large volume energy systems based on wigner-ville distribution evaluated on vibration signals," *Journal of Engineering for Gas Turbines and Power*, vol. 143, no. 7, pp. 1–14, 2021.

[63] F. Reggio, M. L. Ferrari, P. Silvestri, and A. F. Massardo, "Vibrational analysis for surge precursor definition in gas turbines," *Meccanica*, vol. 54, no. 8, pp. 1257–1278, 2019.

[64] J. Wang, J. Chen, G. Dong, and H. Hua, "Wavelet Features and Hidden Markov Model-Based Aerodynamic Instability Detection for Compressors," *Journal of Turbomachinery*, vol. 141, no. 11, pp. 1–9, 2019.

[65] C. A. Niccolini Marmont Du Haut Champ, P. Silvestri, M. L. Ferrari, and A. F. Massardo, "Signal Processing Techniques to Detect Centrifugal Compressors Instabilities in Large Volume Power Plants," *Journal of Engineering for Gas Turbines and Power*, vol. 142, no. 12, 2020.

[66] Y. Bousquet, N. Binder, G. Dufour, X. Carbonneau, I. Trebinjac, and M. Roumeas, "Numerical investigation of Kelvin-Helmholtz instability in a centrifugal compressor operating near stall," *J. Turbomach.*, vol. 138, no. 7, 2016.

[67] J. T. Gravdahl, O. Egeland, and S. O. Vatland, "Drive torque actuation in active surge control of centrifugal compressors," *Automatica*, vol. 38, no. 11, pp. 1881–1893, 2002.

[68] N. Uddin and J. T. Gravdahl, "Piston-actuated active surge control of centrifugal compressor including integral action," in *2011 11th international conference on control, automation and systems*, pp. 991–996, IEEE, 2011.

[69] G. Arnulfi, F. Blanchini, P. Giannattasio, D. Micheli, and P. Pinamonti, "Extensive study on the control of centrifugal compressor surge," *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 220, no. 3, pp. 289–304, 2006.

[70] B. D. Jager, "Rotating stall and surge control: A survey," *October*, no. December, pp. 1857–1862, 1995.

[71] D. A. Fink, N. A. Cumpsty, and E. M. Greitzer, "Surge dynamics in a free-spool centrifugal compressor system," *Proc. ASME Turbo Expo*, vol. 1, 1991.

[72] K. E. Hansen, P. Jorgensen, and P. S. Larsen, "Experimental and Theoretical Study of Surge in a Small Centrifugal Compressor," *Bull. Am. Math. Soc.*, vol. 103, no. 393, 1981.

[73] K. Powers, I. Kennedy, J. Archer, P. Eynon, J. Horsley, C. Brace, C. Copeland, and P. Milewski, "A new first-principles model to predict mild and deep surge for a centrifugal compressor," *Energy*, vol. 244, p. 123050, 2022.

[74] E. Sundström, B. Kerres, S. Sanz, and M. Mihăescu, "On the assessment of centrifugal compressor performance parameters by theoretical and computational models," in *Turbo Expo: Power for Land, Sea, and Air*, vol. 50800, p. V02CT44A029, American Society of Mechanical Engineers, 2017.

[75] T. Engelberth, D. Krawczyk, and A. Verl, "Model-based method for condition monitoring and diagnosis of compressors," *Procedia Cirp*, vol. 72, pp. 1321–1326, 2018.

[76] K. Kabalyk, M. Jasek, G. Liskiewicz, and L. Horodko, "Experimental analysis of the influence of outlet network volume and inlet guide vane positioning on surge behavior in a single-stage low-speed centrifugal compressor," *Proc. Inst. Mech. Eng. Part A J. Power Energy*, vol. 232, no. 4, pp. 350–363, 2018.

[77] B. Kerres, M. Mihaescu, M. Gancedo, and E. Gutmark, "Optimal Pressure Based Detection of Compressor Instabilities Using the Hurst Exponent," *SAE Int. J. Engines*, vol. 10, no. 4, pp. 1917–1926, 2017.

[78] X. Wu, Y. Liu, R. Liu, and L. Zhao, "Surge detection methods using empirical mode decomposition and continuous wavelet transform for a centrifugal compressor," *Journal of Mechanical Science and Technology*, vol. 30, no. 4, pp. 1533–1536, 2016.

[79] N. Wagner, S. Konig, G. Winkes, S. Sub, *et al.*, "On-site determination of the control line for integrally geared compressors for avoidance of impeller fatigue failures," in *Middle East Turbomachinery Symposia. 2011 Proceedings.*, Turbomachinery Laboratory, Texas A&M Engineering Experiment Station, 2011.

[80] K. Botros and J. Henderson, "Developments in centrifugal compressor surge control: A technology assessment," in *Turbo Expo: Power for Land, Sea, and Air*, vol. 78897, p. V002T06A001, American Society of Mechanical Engineers, 1993.

[81] E. Munari, M. Morini, M. Pinelli, K. Brun, S. Simons, and R. Kurz, "A new index to evaluate the potential damage of a surge event: the surge

severity coefficient," *Journal of Engineering for Gas Turbines and Power*, vol. 141, no. 3, 2019.

[82] N. Poujol, I. Trébinjac, and P. Duquesne, "Effects of inlet guide vanes on the performance and stability of an aeronautical centrifugal compressor," *Proceedings of the ASME Turbo Expo*, vol. 2E-2020, pp. 1–12, 2020.

[83] R. Bullock, W. Wilcox, and J. Moses, "SURGING IN CONTINUOUS-FLOW," *NACA Report*, vol. 861, 1946.

[84] N. M. Mcdougalli, N. A. Cumpsty, and T. P. Hynes, "Stall Inception in Axial Compressors," *Int. Gas Turbine Aeroengine Congr. Exhib.*, 1989.

[85] Y. Liu, J. Li, J. Du, F. Li, and H. Zhang, "Application of Fast Wavelet Analysis on Early Stall Warning in Axial Compressors," *Journal of Thermal Science*, vol. 28, no. 5, pp. 837–849, 2019.

[86] A. Young, I. Day, and G. Pullan, "Stall Warning by Blade Pressure Signature Analysis," *J. Turbomach.*, vol. 135, no. 1, pp. 1–10, 2012.

[87] M. Dhingra, Y. Neumeier, J. V. Prasad, A. Breeze-Stringfellow, H. W. Shin, and P. N. Szucs, "A stochastic model for a compressor stability measure," *J. Eng. Gas Turbines Power*, vol. 129, no. 3, pp. 730–737, 2007.

[88] H. Zhang, C. Yang, W. Wang, C. Yang, and Y. Li, "Experimental Investigation of the Pre-Stall and Stall Evolution in a Centrifugal Compressor With a Volute," *Journal of Turbomachinery*, vol. 144, no. 8, pp. 1–13, 2022.

[89] B. Kerres, *On Stability and Surge in Turbocharger Compressors*. PhD thesis, KTH Royal Institute of Technology, 2017.

[90] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[91] N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of time series structure: SSA and related techniques*. CRC press, 2001.

[92] J. Pečinka, A. Jílek, and P. Kmoch, "Small jet engine centrifugal compressor stability margin assessment," *Proceedings of the ASME Turbo Expo*, vol. 2D-2017, pp. 1–8, 2017.

[93] A. X. Liu and X. Q. Zheng, "Methods of surge point judgment for compressor experiments," *Experimental Thermal and Fluid Science*, vol. 51, pp. 204–213, 2013.

[94] G. Liśkiewicz, "Efficient and reliable surge prevention algorithm for centrifugal compressor," *Aircraft Engineering and Aerospace Technology*, 2019.

[95] L. Romani, L. Bosi, A. Baroni, L. Toni, D. Biliotti, G. Ferrara, and A. Bianchini, "Detection of vaneless diffuser rotating stall by means of dynamic pressure sensors and acoustic measurements," *E3S Web of Conferences*, vol. 312, p. 11007, 2021.

[96] N. Courtiade and X. Ottavy, "Experimental study of surge precursors in a high-speed multistage compressor," *Journal of Turbomachinery*, vol. 135, no. 3, pp. 1–9, 2013.

[97] G. Liśkiewicz, L. Horodko, M. Stickland, and W. Kryłłowicz, "Identification of phenomena preceding blower surge by means of pressure spectral

maps," *Experimental Thermal and Fluid Science*, vol. 54, pp. 267–278, 2014.

[98] S. Scholl, "Fourier, gabor, morlet or wigner: comparison of time-frequency transforms," *arXiv preprint arXiv:2101.06707*, 2021.

[99] W. C. Oakes, P. B. Lawless, J. R. Fagan, and S. Fleeter, "High-speed centrifugal compressor surge initiation characterization," *Journal of Propulsion and Power*, vol. 18, no. 5, pp. 1012–1018, 2002.

[100] L. Cohen, *The uncertainty principle for the short-time Fourier transform and wavelet transform.* Springer, 2001.

[101] S. Liao and J. Chen, "Time-frequency analysis of compressor rotating stall by means of wavelet transform," in *Turbo Expo: Power for Land, Sea, and Air. Vol. 78729.*, American Society of Mechanical Engineers, 1996.

[102] C. Brown, S. Sawyer, and W. Oakes, "Wavelet based analysis of rotating stall and surge in a high speed centrifugal compressor," in *38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, p. 4080, 2002.

[103] P. Silvestri, S. Marelli, and M. Capobianco, "Incipient surge analysis in time and frequency domain for centrifugal compressors," *Journal of Engineering for Gas Turbines and Power*, vol. 143, no. 10, pp. 1–13, 2021.

[104] S. Zhou, J. Jin, and Y. Wei, "Research on online diagnosis method of fuel cell centrifugal air compressor surge fault," *Energies*, vol. 14, no. 11, 2021.

[105] A. Y. Goharrizi and N. Sepehri, "Internal leakage detection in hydraulic actuators using empirical mode decomposition and hilbert spectrum," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 2, pp. 368–378, 2012.

[106] W. K. Ngui, M. S. Leong, L. M. Hee, and A. M. Abdelrhman, "Wavelet analysis: mother wavelet selection methods," *Applied mechanics and materials*, vol. 393, pp. 953–958, 2013.

[107] L. Debnath and F. A. Shah, *Wavelet transforms and their applications*. Springer, 2002.

[108] V. Jurdic, P. Joseph, and J. Antoni, "Investigation of rotor wake turbulence through cyclostationary spectral analysis," *AIAA Journal*, vol. 47, no. 9, pp. 2022–2030, 2009.

[109] M. Stajuda, D. Garcia Cava, and G. Liśkiewicz, "Cyclostationary approach for instabilities detection and condition monitoring of centrifugal compressor," in *Turbo Expo: Power for Land, Sea, and Air*, vol. 84102, p. V02ET41A032, American Society of Mechanical Engineers, 2020.

[110] J. Antoni, "Cyclostationarity by examples," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 987–1036, 2009.

[111] E. Munari, G. D'Elia, M. Morini, E. Mucchi, M. Pinelli, and P. R. Spina, "Experimental Investigation of Vibrational and Acoustic Phenomena for Detecting the Stall and Surge of a Multistage Compressor," *Journal of Engineering for Gas Turbines and Power*, vol. 140, no. 9, pp. 1–9, 2018.

[112] S. Debert, M. Pachebat, V. Valeau, and Y. Gervais, "Ensemble-empirical-mode-decomposition method for instantaneous spatial-multi-scale decomposition of wall-pressure fluctuations under a turbulent flow," *Experiments in fluids*, vol. 50, pp. 339–350, 2011.

[113] A. Delgado-Bonal and A. Marshak, "Approximate entropy and sample entropy: A comprehensive tutorial," *Entropy*, vol. 21, 2019.

[114] Z. Huo, M. Martínez-García, Y. Zhang, R. Yan, and L. Shu, "Entropy measures in machine fault diagnosis: Insights and applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2607–2620, 2020.

[115] M. Zhang, J. Zhang, A. Hou, A. Xia, W. Tuo, and Y. Lv, "Aerodynamic system instability identification with sample entropy algorithm based on feature extraction," *Propuls. Power Res.*, pp. 1–15, 2022.

[116] F. Lou and N. L. Key, "Compressor stall warning using nonlinear feature extraction algorithms," *Journal of Engineering for Gas Turbines and Power*, vol. 142, no. 12, 2020.

[117] Y. Hong, P. Lin, T. Chen, and C. Wang, "Analysis of the Stall Inception on Axial Compressors Using Permutation Entropy," *Proceedings - 2020 Chinese Automation Congress, CAC 2020*, pp. 7479–7483, 2020.

[118] Y. Liu, D. Ding, K. Ma, and K. Gao, "Descriptions of entropy with fractal dynamics and their applications to the flow pressure of centrifugal compressor," *Entropy*, vol. 21, no. 3, p. 266, 2019.

[119] J. Antoni, "A critical overview of the "filterbank-feature-decision" methodology in machine condition monitoring," *Acoustics Australia*, vol. 49, no. 2, pp. 177–184, 2021.

[120] T. Alexandrov, S. Bianconcini, E. B. Dagum, P. Maass, and T. S. McElroy, "A review of some modern approaches to the problem of trend extraction," *Econometric Reviews*, vol. 31, no. 6, pp. 593–624, 2012.

[121] Y. Liu, G. Yang, M. Li, and H. Yin, "Variational mode decomposition denoising combined the detrended fluctuation analysis," *Signal Processing*, vol. 125, pp. 349–364, 2016.

[122] M. Civera and C. Surace, "A comparative analysis of signal decomposition techniques for structural health monitoring on an experimental benchmark," *Sensors*, vol. 21, no. 5, p. 1825, 2021.

[123] M. G. Frei and I. Osorio, "Intrinsic time-scale decomposition: time–frequency–energy analysis and real-time filtering of non-stationary signals," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2078, pp. 321–342, 2007.

[124] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE transactions on signal processing*, vol. 62, no. 3, pp. 531–544, 2013.

[125] J. S. Smith, "The local mean decomposition and its application to eeg perception data," *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.

[126] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[127] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," *A practical approach to microarray data analysis*, pp. 91–109, 2003.

[128] P. J. Schmid, "Dynamic mode decomposition and its variants," *Annual Review of Fluid Mechanics*, vol. 54, pp. 225–254, 2022.

[129] F. Cong, J. Chen, G. Dong, and F. Zhao, "Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis," *Mechanical systems and signal processing*, vol. 34, no. 1-2, pp. 218–230, 2013.

[130] Z. Dang, Y. Lv, Y. Li, and G. Wei, "Improved dynamic mode decomposition and its application to fault diagnosis of rolling bearing," *Sensors*, vol. 18, no. 6, p. 1972, 2018.

[131] D. Garcia, R. Palazzetti, I. Trendafilova, C. Fiorini, and A. Zucchelli, "Vibration-based delamination diagnosis and modelling for composite laminate plates," *Composite Structures*, vol. 130, pp. 155–162, 2015.

[132] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, no. 1-2, pp. 108–126, 2013.

[133] Z. Liu and L. Zhang, "A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings," *Measurement*, vol. 149, p. 107002, 2020.

[134] T. Eriksen and N. Rehman, "Data-driven nonstationary signal decomposition approaches: a comparative analysis," *Scientific Reports*, vol. 13, no. 1, p. 1798, 2023.

[135] R. Li and D. He, "Rotational machine health monitoring and fault detection using EMD-based acoustic emission feature quantification," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 4, pp. 990–1001, 2012.

[136] M. Žvokelj, S. Zupan, and I. Prebil, "Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemble Empirical Mode Decomposition method combined with Principal Component Analysis," *Mechanical Systems and Signal Processing*, vol. 24, no. 4, pp. 1049–1067, 2010.

[137] X. Wu and X. Chen, "Internal leakage detection for inlet guide vane system at gas turbine compressor with ensemble empirical mode decomposition," *Measurement: Journal of the International Measurement Confederation*, vol. 134, pp. 781–787, 2019.

[138] M. Barbosh, P. Singh, and A. Sadhu, "Empirical mode decomposition and its variants: A review with applications in structural health monitoring," *Smart Materials and Structures*, vol. 29, no. 9, p. 093001, 2020.

[139] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE signal processing letters*, vol. 11, no. 2, pp. 112–114, 2004.

[140] U. B. de Souza, J. P. L. Escola, and L. da Cunha Brito, "A survey on hilbert-huang transform: Evolution, challenges and solutions," *Digital Signal Processing*, vol. 120, p. 103292, 2022.

[141] N. E. Huang, *Hilbert Huang Transform and its applications*. World Scientific, 2014.

[142] Y.-j. Xue, J.-x. Cao, H.-k. Du, G.-l. Zhang, and Y. Yao, "Does mode mixing matter in emd-based highlight volume methods for hydrocarbon detection? experimental evidence," *Journal of Applied Geophysics*, vol. 132, pp. 193–210, 2016.

[143] M. Bueno-López, E. Giraldo, M. Molinas, and O. B. Fosso, "The mode mixing problem and its influence in the neural activity reconstruction.," *IAENG International Journal of Computer Science*, 2019.

[144] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[145] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Advances in adaptive data analysis*, vol. 2, no. 02, pp. 135–156, 2010.

[146] G. Bin, J. Gao, X. Li, and B. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711, 2012.

[147] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.

[148] Y. Liu, K. Ma, H. He, and K. Gao, "Obtaining information about operation of centrifugal compressor from pressure by combining eemd and imfe," *Entropy*, vol. 22, no. 4, p. 424, 2020.

[149] C. Li, Y. Lei, and R. Fu, "Aerodynamic instability detection in compressor based on Hilbert-Huang transform," *Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011*, vol. 4, no. 1, pp. 355–358, 2011.

[150] S. Yue, Y. Wang, L. Wei, Z. Zhang, and H. Wang, "The joint empirical mode decomposition-local mean decomposition method and its application to time series of compressor stall process," *Aerospace Science and Technology*, vol. 105, p. 105969, 2020.

[151] W. Wang, H. Zhang, C. Yang, and C. Hu, "Experimental analysis of flow instability detection in a centrifugal compressor using variational mode decomposition," *The Aeronautical Journal*, vol. 126, no. 1306, pp. 2058–2082, 2022.

[152] F. Foucher and P. Ravier, "Determination of turbulence properties by using empirical mode decomposition on periodic and random perturbed flows," *Experiments in fluids*, vol. 49, pp. 379–390, 2010.

[153] S. Sanei and H. Hassani, *Singular spectrum analysis of biomedical signals*. CRC press, 2015.

[154] B. Muruganatham, M. Sanjith, B. Krishnakumar, and S. S. Murty, "Roller element bearing fault diagnosis using singular spectrum analysis," *Mechanical systems and signal processing*, vol. 35, no. 1-2, pp. 150–166, 2013.

[155] B. Bhowmik, M. Krishnan, B. Hazra, and V. Pakrashi, "Real-time unified single-and multi-channel structural damage detection using recursive singular spectrum analysis," *Structural Health Monitoring*, vol. 18, no. 2, pp. 563–589, 2019.

[156] D. Garcia, M. Stickland, and G. Liśkiewicz, "Dynamical system analysis of unstable flow phenomena in centrifugal blower," *Open Engineering*, vol. 5, no. 1, 2015.

[157] Y. V. Kozhukhov, A. A. Lebedev, N. M. Chai, and A. M. Yablokov, "Automatic centrifugal compressor pre-surge detection," *AIP Conference Proceedings*, vol. 2285, no. November, 2020.

[158] S. M. Shaharudin, N. Ahmad, and F. Yusof, "Effect of window length with singular spectrum analysis in extracting the trend signal on rainfall data," in *AIP conference proceedings*, vol. 1643, pp. 321–326, American Institute of Physics, 2015.

[159] R. Wang, H.-G. Ma, G.-Q. Liu, and D.-G. Zuo, "Selection of window length for singular spectrum analysis," *Journal of the Franklin Institute*, vol. 352, no. 4, pp. 1541–1560, 2015.

[160] Y. Komatsubara and S. Mizuki, "Dynamical system analysis of unsteady phenomena in centrifugal compressor," *Journal of Thermal Science*, vol. 6, pp. 14–20, 1997.

[161] R. Dehner, N. Figurella, A. Selamet, P. Keller, M. Becker, K. Tallio, K. Miazgowicz, and R. Wade, "Instabilities at the low-flow range of a turbocharger compressor," *SAE International Journal of Engines*, vol. 6, no. 2, pp. 1356–1367, 2013.

[162] X. Xue and T. Wang, "Stall recognition for centrifugal compressors during speed transients," *Applied Thermal Engineering*, vol. 153, pp. 104–112, 2019.

[163] Y. Liu, J. Li, J. Du, H. Zhang, and C. Nie, "Stall Warning Strategy Based on Fast Wavelet Analysis in a Multistage Axial Flow Compressor," *Journal of Engineering for Gas Turbines and Power*, vol. 144, no. 4, pp. 1–6, 2022.

[164] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical systems and signal processing*, vol. 35, no. 1-2, pp. 108–126, 2013.

[165] I. Guyon and A. Elisseeff, "An introduction to feature extraction," *Feature extraction: foundations and applications*, 2006.

[166] C. Zhang, A. A. Mousavi, S. F. Masri, G. Gholipour, K. Yan, and X. Li, "Vibration feature extraction using signal processing techniques for structural health monitoring: A review," *Mechanical Systems and Signal Processing*, vol. 177, p. 109175, 2022.

[167] C. C. Aggarwal, *Data classification*. Springer, 2015.

[168] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4. Springer, 2006.

[169] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[170] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pp. 986–996, Springer, 2003.

[171] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical systems and signal processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[172] K. Hamad, L. Obaid, A. B. Nassif, S. Abu Dabous, R. Al-Ruzouq, and W. Zeiada, "Comprehensive evaluation of multiple machine learning classifiers for predicting freeway incident duration," *Innovative Infrastructure Solutions*, vol. 8, no. 6, p. 177, 2023.

[173] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020.

[174] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[175] H. A. Abu Alfeilat, A. B. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, and V. S. Prasath, "Effects of distance measure choice on k-nearest neighbor classifier performance: a review," *Big data*, vol. 7, no. 4, pp. 221–248, 2019.

[176] J. A. Hartigan, M. A. Wong, *et al.*, "A k-means clustering algorithm," *Applied statistics*, vol. 28, no. 1, pp. 100–108, 1979.

[177] S. B. Imandoust, M. Bolandraftar, *et al.*, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background," *International journal of engineering research and applications*, vol. 3, no. 5, pp. 605–610, 2013.

[178] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Transactions on information theory*, vol. 16, no. 1, pp. 41–46, 1970.

[179] A. Epp, J. C. Hahn, and D. U. Sauer, "Optimization strategy for coupled battery system design models using gaussian process regression and classification," *Journal of Energy Storage*, vol. 52, p. 104998, 2022.

[180] Y. Li, S. Liu, and L. Shu, "Wind turbine fault diagnosis based on gaussian process classifiers applied to operational data," *Renewable Energy*, vol. 134, pp. 357–366, 2019.

[181] Z. Liu, M. Ardabilian, A. Zine, and M. Ichchou, "Crack damage identification of a thick composite sandwich structure based on gaussian processes classification," *Composite Structures*, vol. 255, p. 112825, 2021.

[182] K. Hendrickx, L. Perini, D. Van der Plas, W. Meert, and J. Davis, "Machine learning with a reject option: A survey," *arXiv preprint arXiv:2107.11277*, 2021.

[183] J. Snell and R. Zemel, "Bayesian few-shot classification with one-vs-each pólya-gamma augmented gaussian processes," *arXiv preprint arXiv:2007.10417*, 2020.

[184] K. Loparo, "Case western reserve university bearing data center," *Bearings Vibration Data Sets, Case Western Reserve University*, pp. 22–28, 2012.

[185] N. Buffaz and I. Trébinjac, "Aerodynamic instabilities in transonic centrifugal compressor," *Mechanics & Industry*, vol. 15, no. 3, pp. 191–196, 2014.

[186] L. Yongbo, S. Shubin, L. Zhiliang, and L. Xihui, "Review of local mean decomposition and its application in fault diagnosis of rotating machinery," *Journal of Systems Engineering and Electronics*, vol. 30, no. 4, pp. 799–814, 2019.

[187] P. Prandoni and M. Vetterli, *Signal processing for communications*. EPFL press, 2008.

[188] G. Wang, X. Y. Chen, F. L. Qiao, Z. Wu, and N. E. Huang, "On intrinsic mode function," *Advances in Adaptive Data Analysis*, vol. 2, no. 3, pp. 277–293, 2010.

[189] M. Stajuda, G. Liskiewicz, and D. Garcia, "Flow Instabilities Detection in Centrifugal Blower Using Empirical Mode Decomposition," *Proceedings of the Global Power & Propulsion Society*, 2019.

[190] Z. Wu and N. E. Huang, "On the filtering properties of the empirical mode decomposition," *Advances in Adaptive Data Analysis*, vol. 2, no. 04, pp. 397–414, 2010.

[191] X. Fan and M. J. Zuo, "Machine fault feature extraction based on intrinsic mode functions," *Measurement Science and Technology*, vol. 19, no. 4, 2008.

[192] M. Feldman, "Time-varying vibration decomposition and analysis based on the hilbert transform," *Journal of Sound and Vibration*, vol. 295, no. 3-5, pp. 518–530, 2006.

[193] R. Magiera, "Kompleksowe badania eksperymentalne i numeryczne układów wlotowych dmuchawy promieniowej wyposażonej w nastawną kierownicę wstępną," tech. rep., Lodz University of Technology, 2010.

[194] J. Galindo, J. R. Serrano, C. Guardiola, and C. Cervelló, "Surge limit definition in a specific test bench for the characterization of automotive turbochargers," *Experimental Thermal and Fluid Science*, vol. 30, no. 5, pp. 449–462, 2006.

[195] M. Stajuda, D. G. Cava, and G. Liśkiewicz, "Aerodynamic instabilities detection via empirical mode decomposition in centrifugal compressors," *Measurement*, vol. 199, p. 111496, 2022.

[196] R. Rato, M. D. Ortigueira, and A. Batista, "On the hht, its problems, and some solutions," *Mechanical systems and signal processing*, vol. 22, no. 6, pp. 1374–1394, 2008.

[197] R. Fonseca-Pinto, J. Ducla-Soares, F. Araújo, P. Aguiar, and A. Andrade, "On the influence of time-series length in emd to extract frequency content: Simulations and models in biomedical signals," *Medical engineering & physics*, vol. 31, no. 6, pp. 713–719, 2009.

[198] Z.-S. Chen, S. H. Rhee, and G.-L. Liu, "Empirical mode decomposition based on fourier transform and band-pass filter," *International Journal of Naval Architecture and Ocean Engineering*, vol. 11, no. 2, pp. 939–951, 2019.

[199] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[200] K. Das and S. N. Pradhan, "An efficient hardware realization of emd for real-time signal processing applications," *Int. J. Circuit Theory Appl.*, vol. 48, no. 12, pp. 2202–2218, 2020.

[201] M. Stajuda, D. García Cava, and G. Liśkiewicz, "Comparison of empirical mode decomposition and singular spectrum analysis for quick and robust detection of aerodynamic instabilities in centrifugal compressors," *Sensors*, vol. 22, no. 5, p. 2063, 2022.

[202] W. Byun, Y. Ku, J.-H. Kim, and H. Kim, "Fast auditory evoked potential extraction with real-time singular spectrum analysis," *Electronics Letters*, vol. 53, no. 16, pp. 1094–1096, 2017.

[203] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006.

[204] D. Duvenaud, *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge, 2014.

[205] D. Duvenaud, "The kernel cookbook: Advice on covariance functions," *URL https://www. cs. toronto. edu/˜ duvenaud/cookbook*, 2014.

[206] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[207] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[208] H. Valladares and A. Tovar, "Multi-objective bayesian optimization supported by gaussian process classifiers and conditional probabilities," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 86236, p. V03BT03A039, American Society of Mechanical Engineers, 2022.

[209] M. Stajuda, D. G. Cava, and G. Liskiewicz, "Statistically enhanced classification of centrifugal compressor operating condition," in *The 2022 Leuven Conference on Noise and Vibration Engineering*, 2022.

[210] E. M. Vernon, N. Masuyama, and Y. Nojima, "Error-reject tradeoff analysis on two-stage classifier design with a reject option," in *2022 World Automation Congress (WAC)*, pp. 312–317, IEEE, 2022.

[211] M. Kulak, F. Grapow, and G. Liśkiewicz, "Numerical analysis of centrifugal compressor operating in near-surge conditions," in *Journal of Physics: Conference Series*, vol. 1101, p. 012017, IOP Publishing, 2018.

[212] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.