



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

DZIEDZINA NAUKI INŻYNIERYJNO-TECHNICZNE

DYSCYPLINA AUTOMATYKA, ELEKTRONIKA I ELEKTROTECHNIKA

ROZPRAWA DOKTORSKA

*Metoda zautomatyzowanego wykrywania uszkodzeń
dużych maszyn wirnikowych z wykorzystaniem danych
ze stanów przejściowych*

Autor: mgr inż. Mateusz Tadeusz Zabaryłło

Promotor rozprawy: Prof. dr hab. inż. Tomasz Barszcz
Promotor pomocniczy: Dr hab. inż. Mariusz Banaszkiewicz

Praca wykonana: Akademia Górniczo-Hutnicza im. Stanisława Staszica w
Krakowie, Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii
Biomedycznej

Kraków, 2022



AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIELD OF SCIENCE ENGINEERING AND TECHNOLOGY

SCIENTIFIC DISCIPLINE AUTOMATION, ELECTRONIC AND ELECTRICAL
ENGINEERING

DOCTORAL THESIS

*Automated fault detection of large turbomachinery
using Machine Learning on transient data*

Author: mgr inż. Mateusz Tadeusz Zabaryłło

First supervisor: Prof. dr hab. inż. Tomasz Barszcz
Assisting supervisor: Dr hab. inż. Mariusz Banaszekiewicz

Completed in: AGH University of Science and Technology in Krakow,
Faculty of Electrical Engineering, Automatics, Computer Science and
Biomedical Engineering

Krakow, 2022

*To my wife Michalina, for her absolute dedication to our family. She always worked tirelessly,
especially when her "boy" went away to play with his machines.*

*To my daughters, Zuzia and Hania, who, despite their age, often exceeded their father with
understanding.*

To my Parents who always believed in me.

List of abbreviations

<i>0-pk</i>	-	zero-peak
<i>1X</i>	-	synchronous response of the system
<i>2X</i>	-	second harmonic of the system
<i>AI</i>	-	Artificial Intelligence
<i>ANN</i>	-	Artificial Neural Networks
<i>c</i>	-	damping of the rotor-to-bearing system
<i>CMS</i>	-	condition monitoring system
<i>CS</i>	-	Cubic Spline
<i>D_S</i>	-	overall vibration amplitude
<i>DAIU</i>	-	Digital-to-Analog Interface Unit
<i>DE</i>	-	Differential Evolution algorithm
<i>DE</i>	-	driver end
<i>EEMD</i>	-	Ensemble Empirical Mode Decomposition
<i>FDI</i>	-	Fault Detection and Identification
<i>FEM</i>	-	Finite Element Method
<i>FSNL</i>	-	Full Speed No Load
<i>GA</i>	-	Genetic Algorithm
<i>k</i>	-	stiffness of the rotor-to-bearing system
<i>KURT</i>	-	Kurtosis from the whole transient;
<i>m</i>	-	mass of the system
<i>MAX_{Oo}OpEn</i>	-	maximum distance above the OpEn upper value;
<i>MD3</i>	-	Multidimensional Data Driven Decomposition
<i>MIN_{Oo}OpEn</i>	-	maximum distance below the OpEn lower value.
<i>ML</i>	-	Machine Learning
<i>ML</i>	-	machine learning
<i>MPa</i>	-	Mega Pascal
<i>MW</i>	-	Mega Watt
<i>NDE</i>	-	non-driven end
<i>OpEn</i>	-	Operating Envelope
<i>OpEn BEAR</i>	-	Operating Envelope Baseline Ellipsis Acceptance region matrix (3D case)

<i>OpEn BULL</i>	-	Operating Envelope Baseline Upper and Lower Limits matrix (2D case)
<i>OpEn_dist_mtrx</i>	-	Operating Envelope distance matrix
<i>OpEn_Lower</i>	-	Upper Acceptable parameter value of the Operating Envelope (2D case)
<i>OpEn_Upper</i>	-	Upper Acceptable parameter value of the Operating Envelope (2D case)
<i>RMS</i>	-	Root Mean Squared
<i>RMSE</i>	-	Root Mean Square Error from the whole transient;
<i>SVD</i>	-	Singular Value Decomposition
<i>T</i>	-	time of revolution
<i>t_1</i>	-	time trigger occur
<i>t_2</i>	-	time between trigger and the highest waveform point
<i>u</i>	-	unbalance mass
<i>WPA</i>	-	wavelet packet analysis
<i>WPT</i>	-	wavelet packet transform
<i>X, Y</i>	-	Cartesian coordinates
<i>θ_S</i>	-	phase angle
<i>ω</i>	-	rotational speed

Content

Streszczenie	9
Abstract	10
1. Introduction.....	11
1.1. Large turbomachinery in industry	11
1.2. Fundamentals of rotating machinery measurements.....	17
1.2.1. Accelerometers	17
1.2.2. Velocity sensors.....	19
1.2.3. Displacement sensors.....	20
1.2.4. Amplitude and phase lag measurement	24
1.3. Selected aspects of rotordynamics	26
1.3.1. Transient rotor response	26
1.3.2. Unbalance response of the system undergoing transient states.....	29
1.4. Problem formulation	31
1.4.1. The motivation	31
1.4.2. Goal of the thesis.....	34
1.4.3. Scope of the thesis	35
2. Measurement data of large turbomachinery.....	37
2.1. Design and operation of large steam turbines.....	37
2.2. Portable measurement systems.....	40
3. Transient data preparation and preprocessing.....	43
3.1. Portable equipment used for data collection	43
3.2. Measurement system configuration	44
3.3. Transient data point sampling.....	50
3.4. Structure of measurement data.....	52
3.5. Data interpolation	53
3.5.1. Cubic spline.....	57
3.5.2. Cubic Spline spacing	58
4. Fault detection method.....	60
4.1. Anomaly detection method	60
4.2. OpEn method	62
4.3. Baseline measurements – selection and its consequences	65

4.4. 2D baseline	66
4.5. Upper and lower OpEn values	68
4.6. OpEn 2D severity parameter definition	68
4.7. 3D baseline	72
4.8. OpEn 3D severity parameter definition	74
5. Fault Identification method.....	78
5.1. Parameter identification method.....	78
5.2. Malfunction identification methods.....	79
5.3. Multidimensional Data Driven Decomposition (MD3).....	81
5.4. Identification parameters.....	85
5.5. Differential Evolution algorithms as a part of Genetic Algorithms family	86
6. Architecture of the automated FDI system	89
6.1. System architecture.....	89
6.2. Algorithms for the detection and identification methods	92
6.2.1. Data import module	92
6.2.2. Detection method for the OpEn 2D and 2D case	93
6.2.3. MD3 method algorithm.....	96
6.2.4. Key Python functions.....	97
7. Validation of the anomaly detection method	116
7.1. Baseline measurements for the validation method.....	116
7.2. Single channel: 1X amplitude validation (OpEn 2D case).....	119
7.3. 1X amplitude and phase validation (OpEn 3D case)	124
8. Validation of the identification method.....	135
8.1. Validation of model data	135
8.2. Validation of MD3 method on real turbine.....	139
9. Conclusions.....	143
9.1. Final remarks	143
9.2. Author's contribution	144
9.3. Further research.....	146
10. References.....	147
List of Figures.....	151
List of tables	154

Streszczenie

Wykrywanie i identyfikacja uszkodzeń dużych maszyn wirnikowych jest jednym z najważniejszych zagadnień diagnostyki maszyn energetycznych dużej mocy. Czynnikiem stwarzającym duże przeszkody w analizie tak dużych obiektów badawczych jest złożoność ich odpowiedzi drganiowej dla linii wałów, składającej się z kilku części. Istnieje szereg procedur i metod pozwalających wykrywać i identyfikować anomalie podczas pracy w stanach ustalonych maszyn energetycznych. Ważniejsza pod względem diagnostyki uszkodzeń jest jednak analiza stanów przejściowych tych maszyn. Zasadniczą wadą obecnego podejścia jest konieczność angażowania ekspertów z dużym doświadczeniem, co jest bardzo kosztowne i pracochłonne.

Celem pracy było zaproponowanie kompletnego systemu automatycznej diagnostyki uszkodzeń dużych maszyn wirnikowych na bazie ich odpowiedzi w stanach przejściowych – w szczególności podczas zmiany prędkości obrotowej.

Problem badawczy, to jest detekcja i identyfikacja uszkodzeń podczas uruchomienia lub odstawienia turbosespołu dużej mocy został rozwiązany poprzez analizę szeregu badań diagnostycznych wykonanych przez autora na obiektach zainstalowanych w elektrowniach zawodowych oraz symulację konkretnych niesprawności na stanowisku laboratoryjnym. Bazy danych z badaniami diagnostycznymi zawierały zarówno odpowiedzi turbosespołów wolnych od uszkodzeń, jak i z konkretnymi uszkodzeniami zweryfikowanymi w trakcie badań diagnostycznych na obiektach. Autor w poniższej pracy używał danych pochodzących z przenośnego urządzenia diagnostycznego. Autor opracował i zaproponował dwie metody: Operating Envelope – OpEn (do automatycznej detekcji uszkodzeń podczas stanów przejściowych) oraz Multidimensional Data Driven Decomposition – MD3 (do automatycznej identyfikacji niesprawności). Ustalenie danych referencyjnych jak i przygotowanie danych z aktualnego stanu przejściowego jest oparte na interpolacji Cubic Spline (w celu ujednoczenia interwałów, dla których będzie przeprowadzana analiza wszystkich danych). W metodzie detekcji uszkodzeń (OpEn) autor wykorzystał koncepcję zbliżoną do obwiedni sygnału (Spectrum Envelope) w celu określenia regionu akceptacji poprawności odpowiedzi turbosespołu. Do automatycznej identyfikacji parametrów dekomponowanych funkcji, został wykorzystany algorytm Differential Evolution (DE), który wywodzi się z rodziny algorytmów genetycznych Genetic Algorithms (GA). Pozostałymi narzędziami składającymi się na cały zaproponowany system są: dla metody detekcji: dwu- i trójwymiarowy rejon akceptacji dla każdego z czujników poszczególnych części turbosespołu, a dla metody identyfikacji: zbiór trzech scenariuszy z odpowiednio zmodyfikowanymi funkcjami dekompozycji wraz z miarą ich dopasowania. Wszystkie metody zostały przebadane na danych pochodzących z symulowanego środowiska na stanowisku laboratoryjnym oraz na danych z obiektów rzeczywistych.

Słowa kluczowe: maszyny energetyczne, wykrywanie uszkodzeń, dynamika wirników, przetwarzanie sygnałów, dekompozycja sygnału, algorytmy genetyczne, Differential Evolution.

Abstract

Detection and identification of malfunctions in large rotating machines are among the most critical subjects in the diagnostics of utility power generation machinery. The factor that creates significant obstacles in analyzing such large research objects is the complexity of the vibration response for the entire shaft line train, composed of several parts. There are several procedures and methods to detect and identify anomalies during the steady-state operation of turbomachinery. More important in fault diagnosis is the analysis of transient states of these machines. Key disadvantage of these methods is involvement of human experts with strong experience.

The aim of the research was to propose a complete system of automatic fault diagnosis of large rotating machines based on their responses in transient states - particularly during changes in rotational speed.

The research problem, i.e., the detection and identification of failures during the commissioning or shutdown of a high-power turbine set, was solved by analyzing a number of diagnostic tests performed by the author on facilities installed in utility power plants and simulating specific malfunctions at a test rig. The databases with diagnostic tests contained both the responses of fault-free turbine sets and those with specific damages confirmed during diagnostic tests on objects. The author used data from a portable diagnostic device. The author developed and proposed two methods: the Operating Envelope - OpEn method (for automatic fault detection during transient states) and the Multidimensional Data Driven Decomposition - MD3 method (for automatic fault identification). Determining the reference data and preparing the data from the current transient state is based on the Cubic Spline interpolation (to standardize the intervals for which all data will be analyzed). In the failure detection method (OpEn), the author used a concept similar to the signal envelope (Spectrum Envelope) to determine the region of acceptance of the correctness of the turbine set response. In his analysis, the Differential Evolution (DE) algorithm was used to automatically identify the parameters of the decomposed functions derived from the Genetic Algorithms (GA) family of genetic algorithms. The remaining tools that make up the entire proposed system are, for the detection method: two- and three-dimensional acceptance regions for each of the sensors of individual parts of the turbine set, and for the identification method: a set of three scenarios with appropriately modified decomposition functions along with a measure of their matching. All the methods were tested on data from a simulated environment on a laboratory stand and data from real turbo generators.

Keywords: power generation machinery, fault detection, rotor dynamics, signal processing, signal decomposition, genetic algorithms, Differential Evolution.

1. Introduction

1.1. Large turbomachinery in industry

Large sets of turbogenerators are the main contributors to the world's electric generation. Although in last years renewable energy sources are rapidly increasing, large utility power plants with large turbogenerators will play the overwhelming role in the industry. Such a scenario, according to forecasts, will remain for at least a few decades. Turbines coupled with generators are the primary machines in every large nuclear and fossil fuel utility power plant worldwide, and their ability to operate is critical for the power generation process. Such necessary units are called "critical machines" because their unavailability can degrade the operation of the national electric grid system, which can be harmful to the national industry. Therefore, their malfunctions should be detected to avoid catastrophic failures and unplanned shutdowns.

To show the perspective, utility power plants are large-size facilities. The turboset plays a central role as it is a unit that converts mechanical energy into electrical energy. A turbine hall can be as long as 0.5 km or even 1 km in extreme cases. Figure 1 shows two types of turbine hall arrangement. Top of the figure depicts an old type of machine hall. This arrangement is characterized by many smaller units, typically eight to twelve, located parallel to each other. The power output of a single unit usually is up to 230MW. Figure 1, bottom, shows a new arrangement of the power generating units. It is characterized by a small number of turbine sets, usually one or two, located one after the other. These are high-power units, as the unit's output power is often close to 1GW, which is four to five times greater than old types of turbine sets.



Figure 1. Power plant machine hall: top – old arrangement; bottom – new arrangement.

Figure 2 illustrates an example of a 13K215 type turbine hall (top of Figure 2) and the turbine part cross-section (bottom of Figure 2). These units are the most popular turbine type in Poland's power

generation industry. These machines were mostly manufactured in the 1980s. They have been successively repaired and modernized to meet the operational requirements and increase their availability.

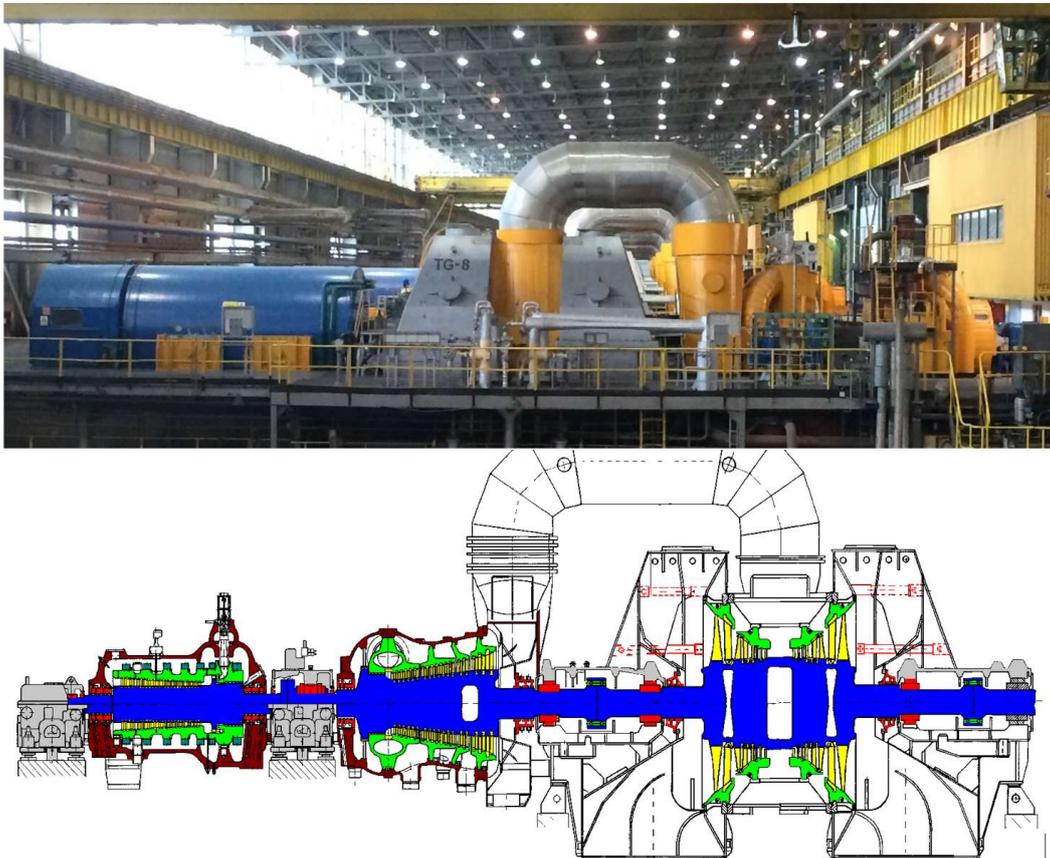


Figure 2. Turbine hall of 13K215 units installed in a power plant (top) and cross-section of the turbine type (bottom).

Figure 3 shows two opposite steps of the process of replacement of one of the key machine components - the Intermediate Pressure (IP) rotor. The top part of the figure shows the old rotor after it has been disassembled. The bottom part shows the new rotor installation in the IP cylinder body. Such regular modernizations allow extending the life of machines for many years. However, it must be supported by appropriate supervision of the device's safety. Together with increased electricity demand, new challenges arise both for the machines that are expected to be more reliable and for monitoring systems to diagnose malfunction without expert knowledge in a reliable way. These machines are expected to operate for long periods without the necessity of being shut down. From start-up to coast down, intervals between transient states can be measured in months or, in some cases, even longer. They are operated in varying conditions such as load change from 40% up to the nominal load, different steam temperatures, and pressures. This type of operation can introduce a large amount of stress, which eventually can lead to fatigue and, in extreme cases, to a failure.

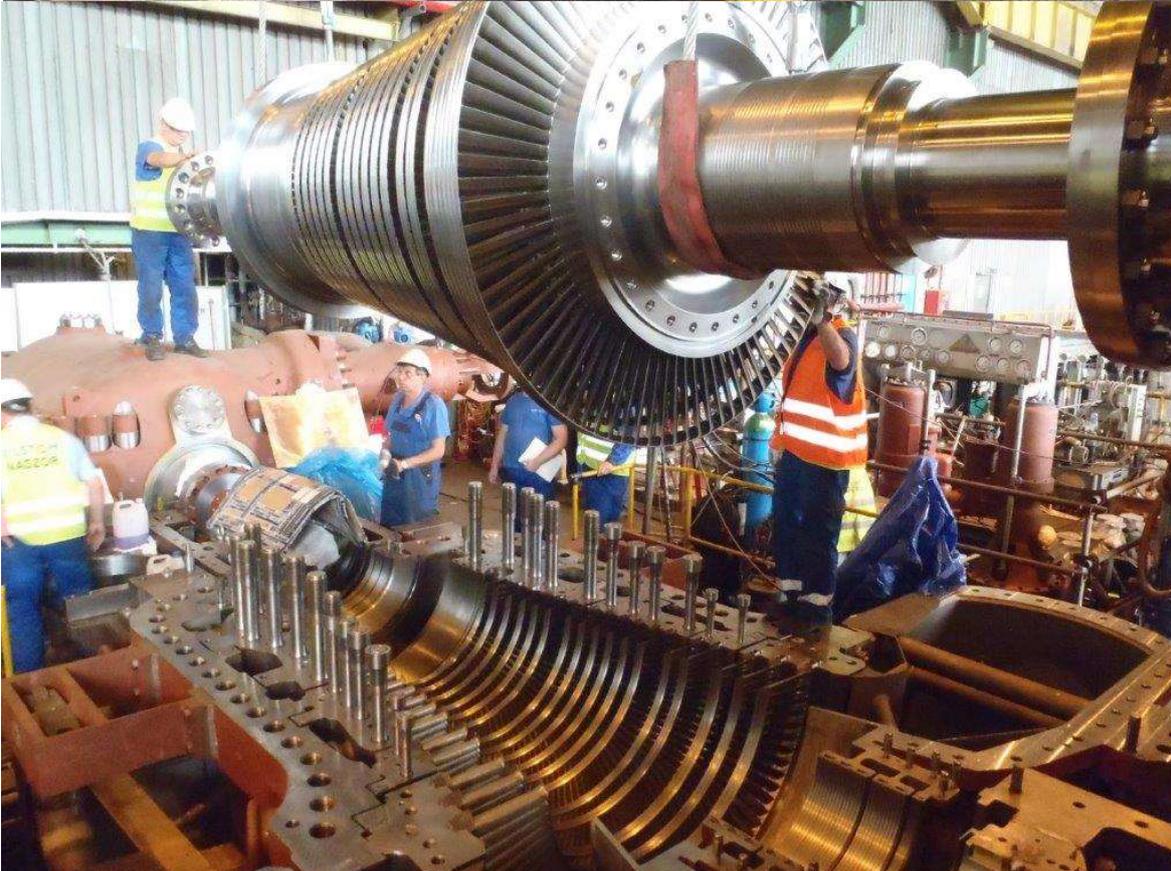


Figure 3. Intermediate Pressure (IP) turbine rotor: top - disassembly of the old one; bottom - new one assembly.

To put this into perspective, the Figure 4 presents the smallest and the largest part of the turbine drivetrain. The smallest of the turbogenerator rotors is the high-pressure rotor (HP). It is about four meters long and weighs about five tons. Figure 4 (a) shows the on-site assembly of an HP cylinder with a rotor. On the other hand, the largest part of the turbine is the rotor of the low pressure (LP) cylinder. It measures about 10 meters in length and weighs above 100 tons. Figure 4 (b) and (c) presents the LP part (cylinder and rotor) assembly process.

These types of equipment have to be robust, and their assembly must be extremely exact. For such large elements to run smoothly, they must be assembled with very high precision. Usually, the alignment of the particular shaft axis has to fall into a ± 0.015 mm tolerance. The process of alignment of each shaft is complicated in terms of the logistics of cumbersome objects and constraints that have to be met. In addition, it can last for 72h for a single coupling. These operations make the overhaul activities long-lasting and tedious for both the contractor and the machine's owner. Turbine sets are equipped with vibration monitoring and supervision systems to avoid unplanned outages and unwanted repairs.



(a)



(b)



(c)

Figure 4. Steam turbine parts: (a) – HP cylinder assembly; (b) – 200MW class LP rotor assembly; (c) – two LP rotors of +450MW class unit in situ.

Steam-powered turbogenerators still play an important role in the worldwide electrical power generation. The survey by Xiao et al. [1] presents the main components of a fossil-fuel power plant and its importance and share in the world power generation industry. Although renewable energy each year takes a bigger and bigger share in the power generation market, the safety of the power grid

requires large units in the system to balance the renewable sources in case of lack of wind or sun. Another important use of steam plants is supplying steam required by numerous industrial processes (e.g. chemical or paper plants). In such cases, renewal energy fails to provide such a media as was presented in the report [2].

1.2. Fundamentals of rotating machinery measurements

Large turbomachinery like steam and gas turbogenerators should operate as long as possible without stoppage or interruptions of power generation. This approach forces a new monitoring standard for steam and gas turbogenerators. Monitoring and assessment of their technical condition are carried out using signals from vibration sensors. There are three basic types of sensors measuring turbine vibrations. These sensors are: accelerometers, velocity sensors, and eddy current sensors. Their designs and principles of operation vary from one type to another. In the next sections the main differences between these sensors will be presented.

1.2.1. Accelerometers

Acceleration sensors are usually small and light, as can be seen in Figure 5 top. They can measure wide range of frequencies. Depending on size and dimensions their frequency response usually starts from 3Hz and spans up to 120kHz . Figure 5, bottom, presents typical components of accelerometer. The main element is a piezoelectric element sandwiched between base of the sensor and the seismic mass. Stress applied to the piezoelectric element causes an electric charge to be generated. The sensor's electrical response is directly proportional to the vibration's acceleration to which it is subjected. The force acting on piezoelectric element is obtained by the mass and preload spring configuration in the sensor. Often, they have integrated preamplifiers to rectify and enhance the signal output. Also, they need an external power supply to work.

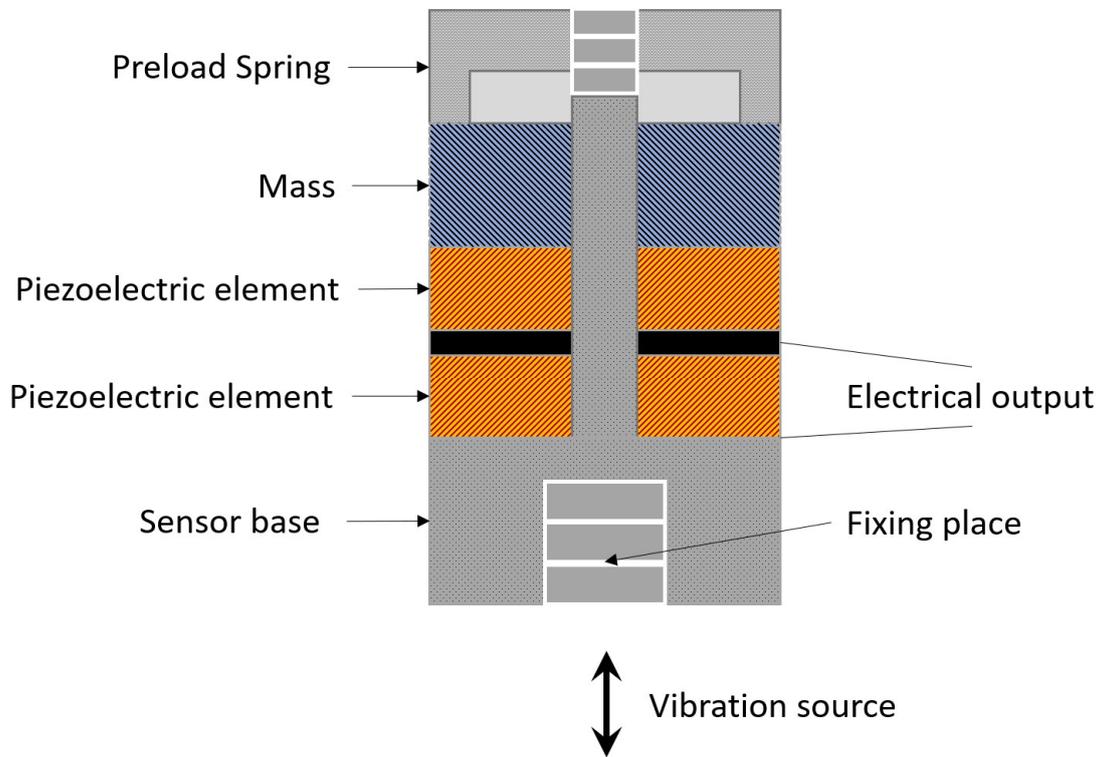
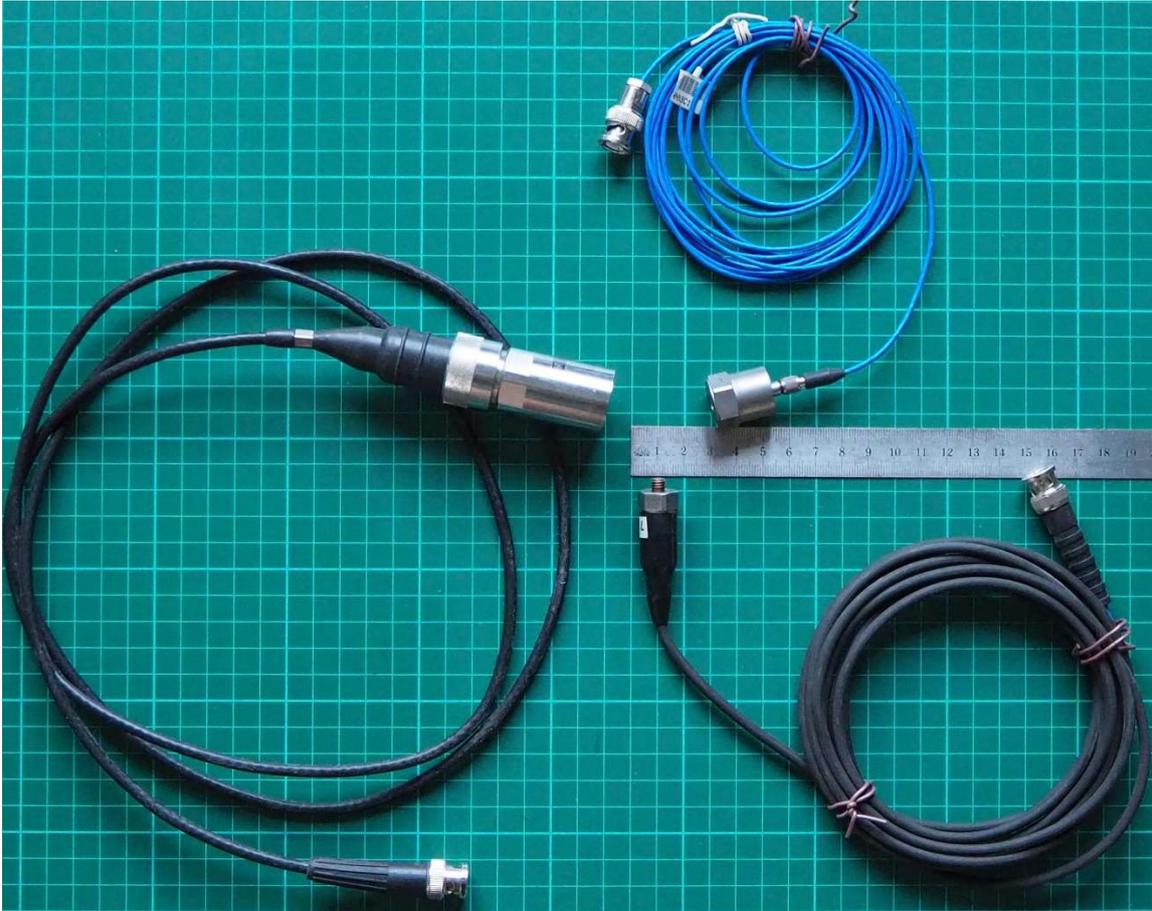


Figure 5. Typical accelerometer sensor: top – real sensors examples; bottom – sensor components schematic.

Due to their high-frequency bandwidth, these sensors are often used to monitor rolling bearings and planetary gears.

1.2.2. Velocity sensors

Velocity sensors are generally the biggest and the bulkiest of all types of vibration measurement sensors. They measure absolute velocity of the stationary structures. Due to the moving mass, i.e., permanent magnet, the usable frequency sensor output range spans from approximately 10 ÷ 15Hz up to 1500Hz, and they are sensitive to the mounting orientation. Figure 6, top left, presents examples of velocity sensors used to monitor turbine 's bearing pedestal casings, and top right – velocity sensor mounted on the bearing pedestal on-site. The components of the typical velocity sensor are presented in Figure 6. The signal is generated by the permanent magnet moving inside a coil. The signal output is proportional to a velocity of vibration to which the sensor is subjected. They generate signal output without external power supply.

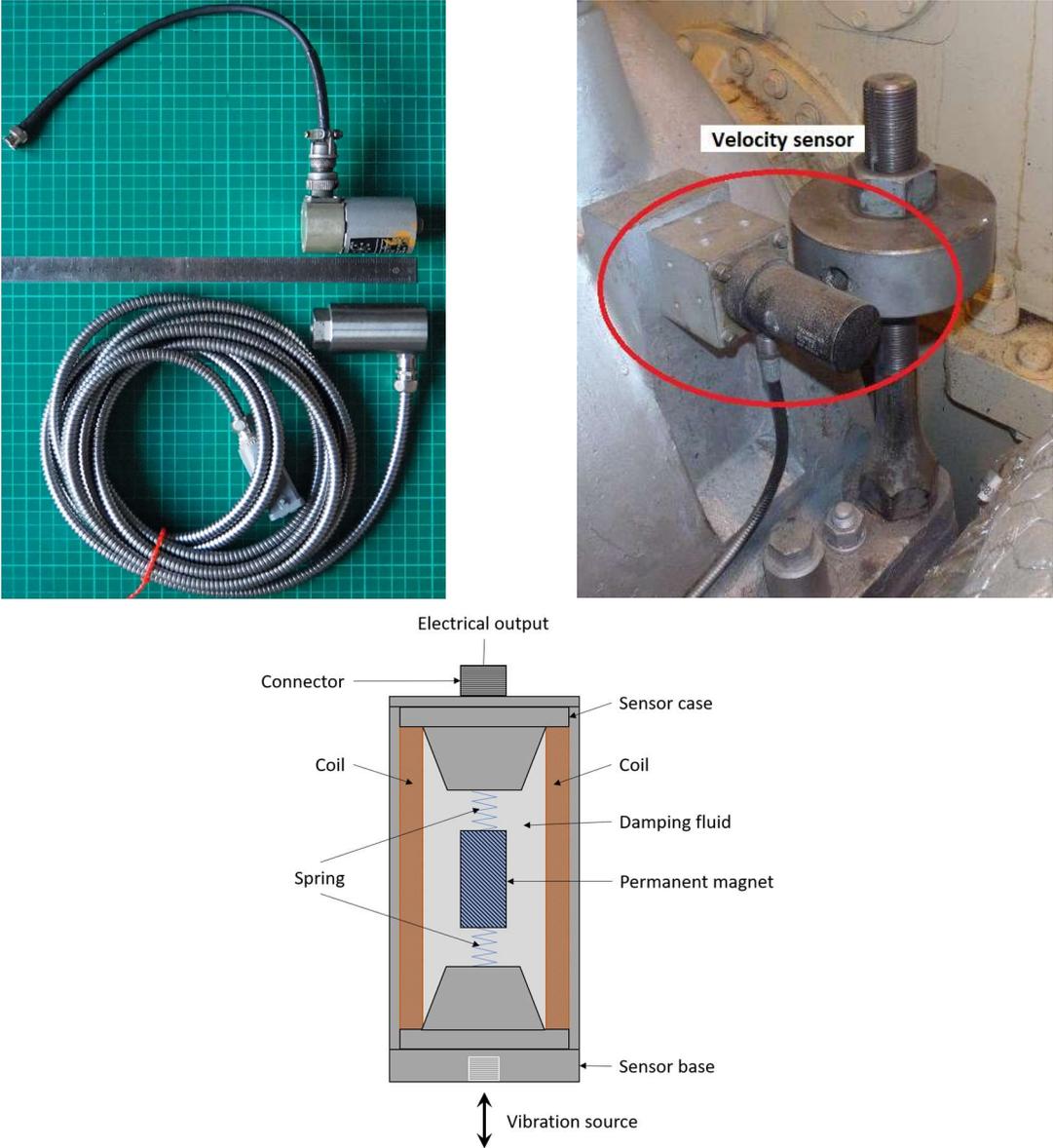


Figure 6. Typical velocity sensor: top left – real sensors examples, top right – on-site velocity sensor assembly; bottom – sensor components schematic.

Velocity sensors are often used in large turbomachinery monitoring and diagnostics as a supplementary measurements to eddy-current probes.

1.2.3. Displacement sensors

Shaft's motion in relation to the bearing casing in fluid-film bearings is the vital monitoring parameter. This motion is used in diagnostic purposes as the most important source of information. Eddy-current probes can measure both static and dynamic motion of the shaft. It measures the gap between rotor and sensor. Static part of the signal is a rotor position in the bearing and the dynamic part is the relative motion from actual static position. Critical machines are equipped with eddy current probes that measure the shaft's vibration inside a journal bearing in two perpendicular axes (relatively to one another). In the large turbomachinery equipped with fluid-film bearings, relative bearing-to-rotor displacement measurement provides essential information about the behavior of the rotor . Figure 7, top left, presents a schematic of their arrangement, and Figure 7 bottom depicts the field assembly.

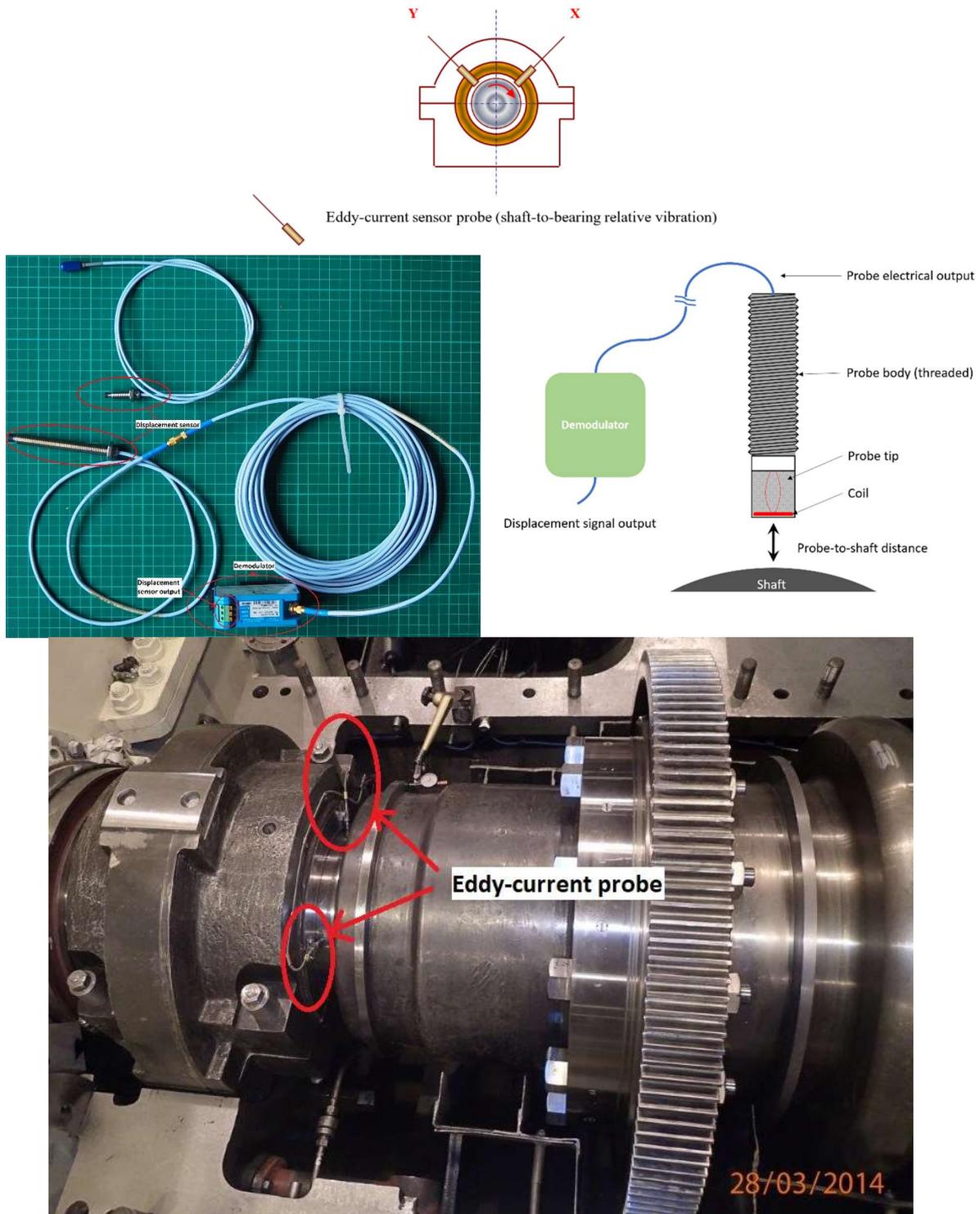


Figure 7. Arrangement of eddy-current probes in the bearing housing; top– schematic, middle left – physical sensors example, middle right – typical sensor components; bottom – field assembly on a journal bearing.

There are several different applications of eddy-current probes in the turboset measurements apart from vibration measurements. One of the most important is the measurement of the turbine's axial displacement. It measures the position of the turbine shaft to the reference point, often found inside the thrust bearing pedestal. This measurement is critical and is part of the machine safety system. Other measurements that use such sensors are the relative expansion of the rotor to the turbine

cylinder body, as depicted in Figure 8. Another measurement is the rotational speed, phase angle, and eccentricity (or mechanical runout of the shaft in front of the rotor in a radial direction) measurements used for diagnostic purposes.



Figure 8. Relative expansion sensor assembly.

In Chapter 6 of [3], Eisenmann describes various types of sensors, including eddy-current sensors, for machine health monitoring and protection systems. Figure 9 top shows the arrangement of shaft axial position, reference one-per-revolution point (called Keyphasor by Bently [4]), and relative vibrations sensors. The bottom part of Figure 9 depicts the real-life assembly and placement of the probes in the thrust bearing pedestal.

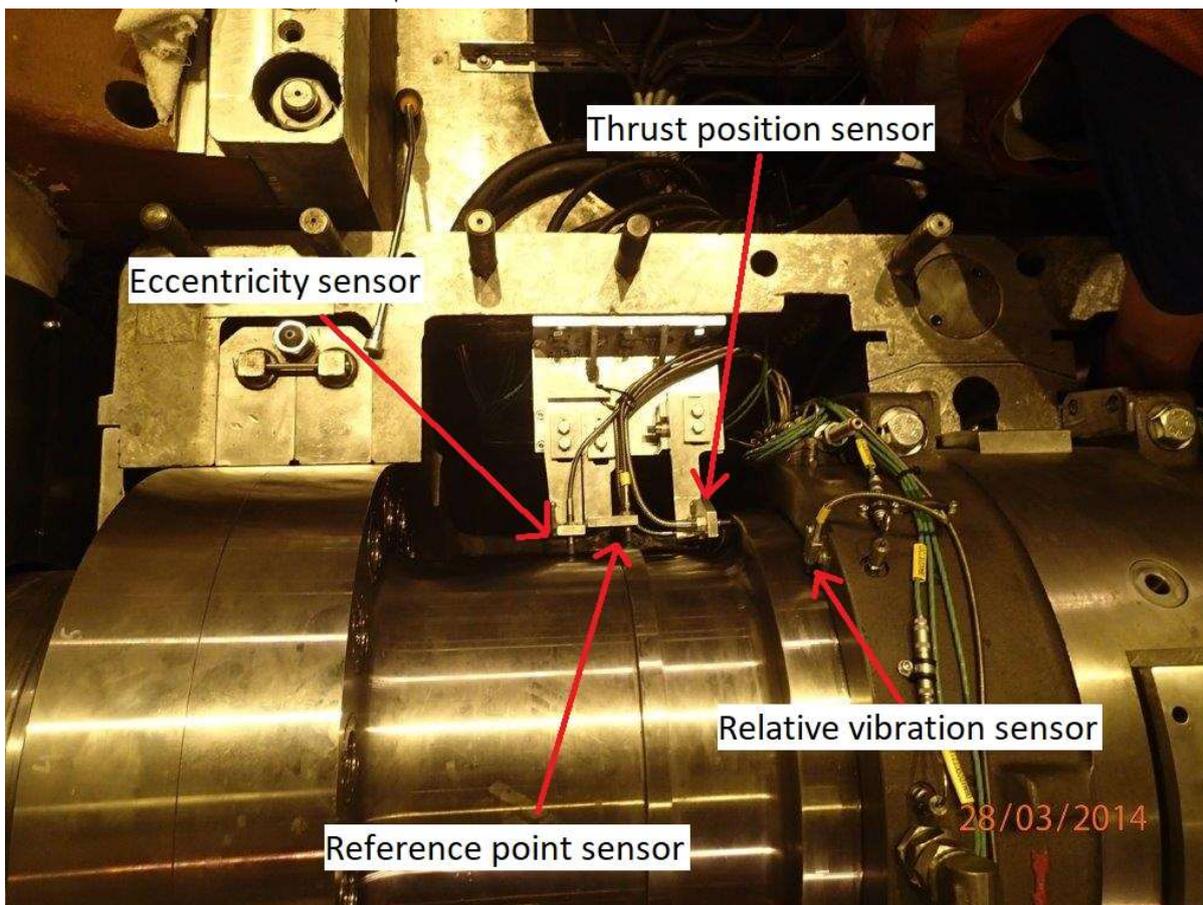
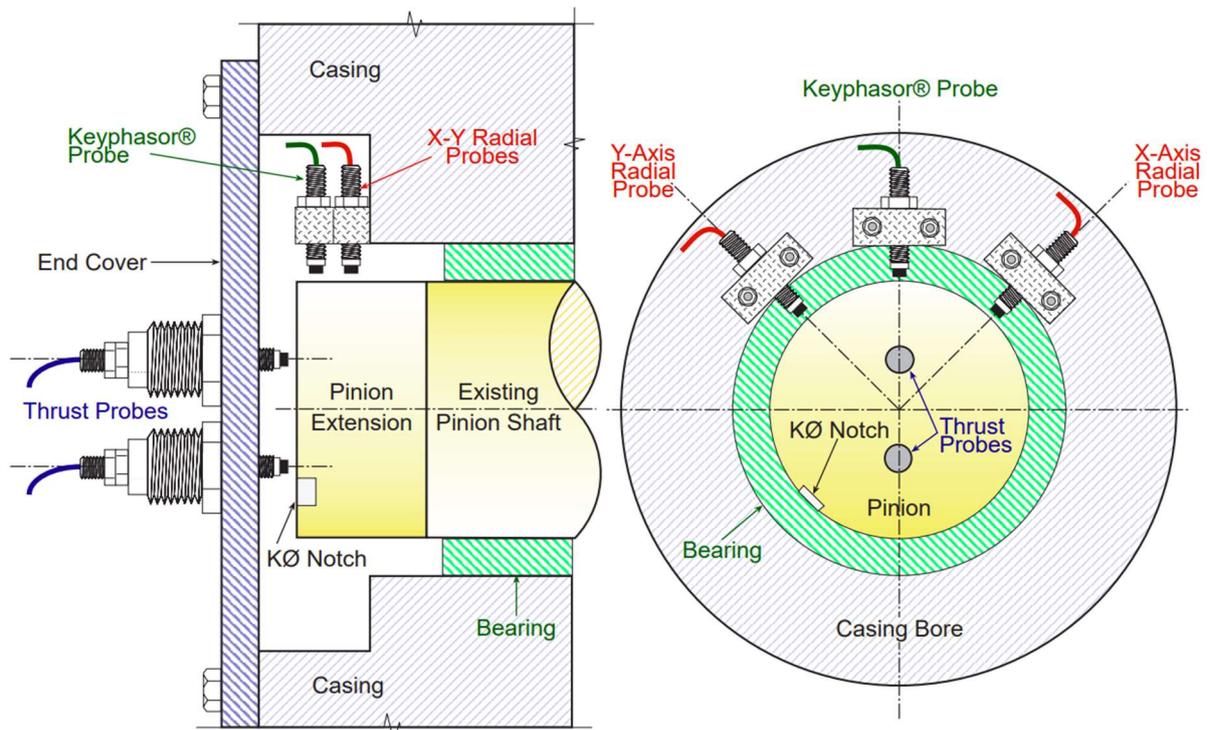


Figure 9. Eddy-current probe installation example; top – schematic arrangement by Eisenmann [3], bottom – physical assembly of the eddy-current sensors inside thrust bearing pedestal for different type of measurements.

Performing a vibration analysis involves considering several parameters, e.g. broadband features, frequency selective features, and harmonic vectors. The most critical parameters to be assessed are

the natural frequencies of a shaft line. Vo and Ton-That [5] present an extensive study in this respect. The monitoring and protection systems of most large rotating machinery equipped with sliding fluid-film bearings typically use two types of sensors, namely eddy current and velocity sensors.

As assessed following ISO 20816-1, the technical condition of rotating machinery should fall within the 10-100Hz frequency range. Due to the features of the sensors presented in Sections 12-15, eddy current sensors are the primary sensors for monitoring and protection systems of turbogenerators. Each bearing of the machine is equipped with this type of measurement. It is the main quality parameter determining the technical condition of the machine. In addition, the data from vibration velocity factors are often used as supplementary ones. Unfortunately, turbo-sets are often not equipped with velocity sensors in a repeatable and unambiguous manner due to economic reasons. This fact disqualifies measurements from these sensors for automatic evaluation and implementation on the entire machine park.

1.2.4. Amplitude and phase lag measurement

To extract diagnostic information from the vibration signal, a set of two parameters can provide the best results. These are the amplitude and frequency of a vibration component of a known frequency. For example, in rotating machinery with sliding bearings, the critical feature in signal extraction is the phase lag of signal components. This parameter determines the timing difference between pre-defined events. The example of the physical arrangement for this type of sensor is depicted in Figure 9. For example, the once-per-revolution mark determines the 360° of rotation, and the time between the consecutive marks determines the angular speed of the machine presented in Figure 10. The once-per-revolution mark is one of these events, and the second one is the closest distance between the sensor probe and rotor present in the figure. A diagnostic system uses an eddy-current sensor to produce this reference point.

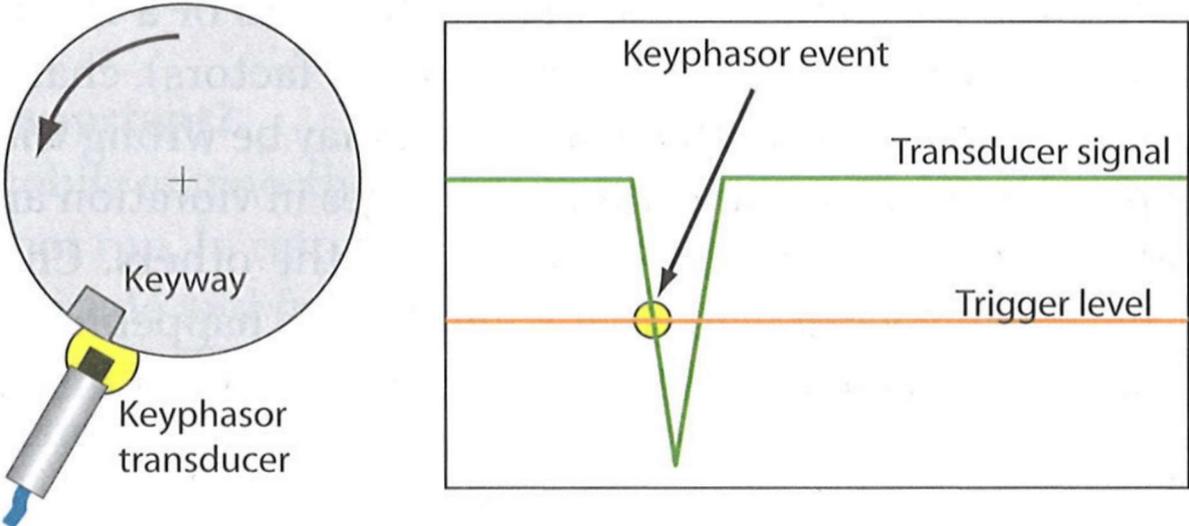


Figure 10. Once-per-revolution event. Timing difference and rotational speed measurement [4].

It becomes a reference point for the whole shaft-line vibration measurements. All features from the vibration probes across the machine train reference to this signal. This reference enables obtaining

specific characteristics of the dynamic response of the shaft-line. Bently and Hatch [4] presented these concepts in an easy and straightforward way. Figure 11 depicts extracting a phase lag and amplitude from a simple vibration signal. The top chart presents "the raw" vibration signal from the sensor. The middle part of the figure describes a one-per-revolution reference point from the lag angle sensor (called Keyphasor by Bently [4]). The bottom chart shows the combination of the upper two to produce the wanted vibration feature. Time T is the time between the reference mark events.

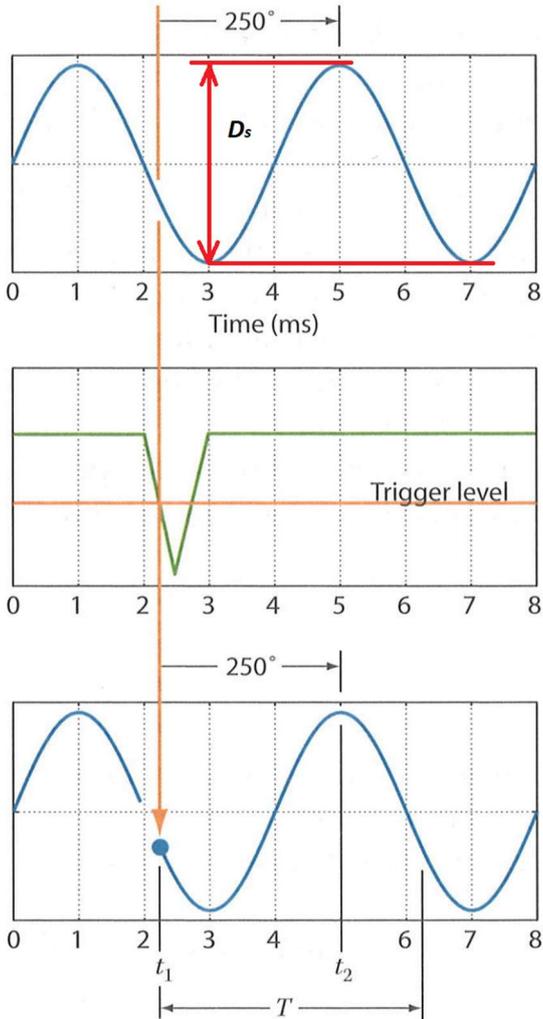


Figure 11. Phase lag angle, angular velocity, and amplitude extraction procedure [4].

The difference $t_1 - t_2$ is the time between the reference point and the highest point in the waveform since the reference point occurred. It is the amplitude D_s of the vibration vector represented by the Equation (7), and depicted in Figure 11. Equation (1) presents the phase angle β_s derivation.

$$\frac{t_1 - t_2}{T} = \beta_s \quad (1)$$

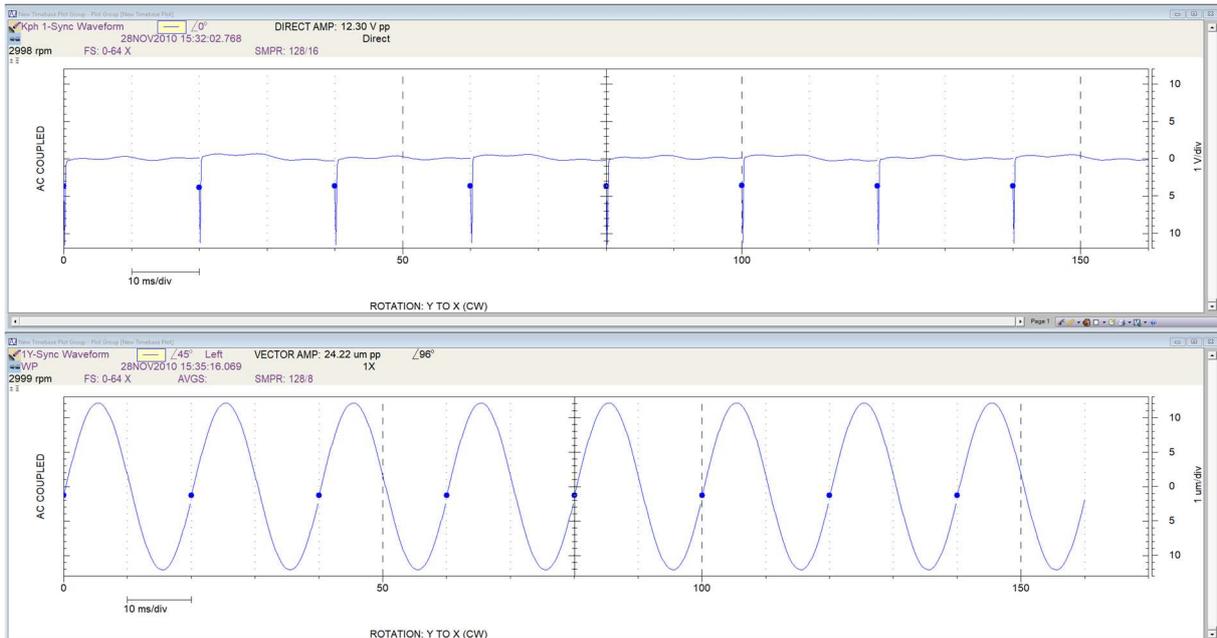


Figure 12. Reference mark and vibration signal during field measurement.

An example of extracting the characteristics of a vibration signal during field measurements is shown in the Figure 12.

1.3. Selected aspects of rotordynamics

In this thesis, the author focuses on determining the baseline behavior of the machine in a healthy state, detecting an anomaly when it occurs during the transient state, and identifying the malfunction – when detected. To introduce the reader into the subject, the key notions from the field of rotordynamics will be presented in the following section.

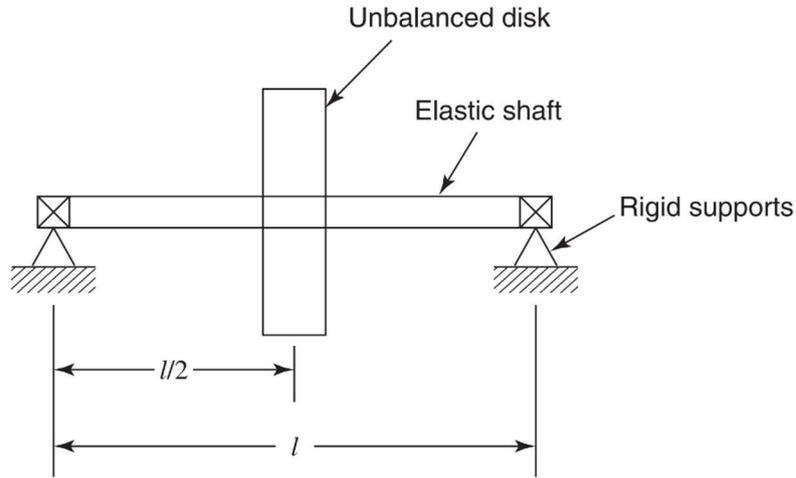
1.3.1. Transient rotor response

The most fundamental concept of estimation of rotating machinery is the model of the Jeffcott rotor described by Vance et al. [6] and presented in Figure 4. Starting from this model, Ehrich [7] presents the response of the rotor-bearing system to excitation by response for the synchronous excitation, i.e., imbalance during a transient state, in his in-depth study of rotordynamic topics.

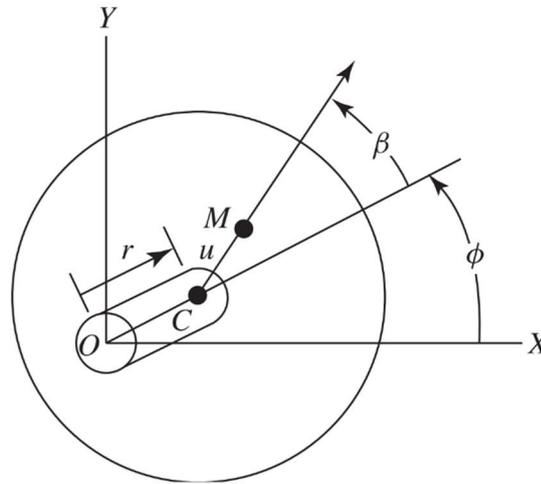
Thus the equations of motion for the Cartesian coordinate for the rotor model in Figure 4 can be listed as follows [6].

$$m\ddot{X} + c\dot{X} + kX = m\omega^2 \cos \omega t \quad (2)$$

$$m\ddot{Y} + c\dot{Y} + kY = m\omega^2 u \sin \omega t \quad (3)$$



(a)



(b)

Figure 13. The Jeffcott rotor model (a) [6]; End view of the Jeffcott rotor and its coordinates (b) [6].

With a solution of:

$$X = \frac{\omega^2 u}{\sqrt{\left(\frac{k}{m} - \omega^2\right)^2 + \left(\frac{c\omega}{m}\right)^2}} \cos(\omega t - \beta_s) \quad (4)$$

$$Y = \frac{\omega^2 u}{\sqrt{\left(\frac{k}{m} - \omega^2\right)^2 + \left(\frac{c\omega}{m}\right)^2}} \sin(\omega t - \beta_s) \quad (5)$$

$$\beta_s = \tan^{-1}\left(\frac{c\omega}{m\left(\frac{k}{m} - \omega^2\right)}\right) \quad (6)$$

$$D_s = \sqrt{X^2 + Y^2} \quad (7)$$

Equations (6) and (7) can be drawn for the angular speed of the rotor ω . Such graphs will provide helpful information on the behavior of the rotor during transient state operation. Figure 14 presents the solution to equations presents the solution to the equation (2) and (3) depicted visually throughout the transient state. Analyzing Figure 14, well below critical rotor speed ($\omega = \sqrt{k/m}$), phase angle (phase lag β_s) is in phase with the unbalance force. The shaft behaves like a rigid body within this zone, i.e., it does not deflect. As a rotor approaches its critical speed, whirling amplitude D_s approaches its maximum with a phase angle reaching 90° , and the rotor deflection reaches the maximum value. With increasing rotation speed, phase lag will increase from 90° up to 180° , which will cause whirling amplitude to reduce after its critical speed range.

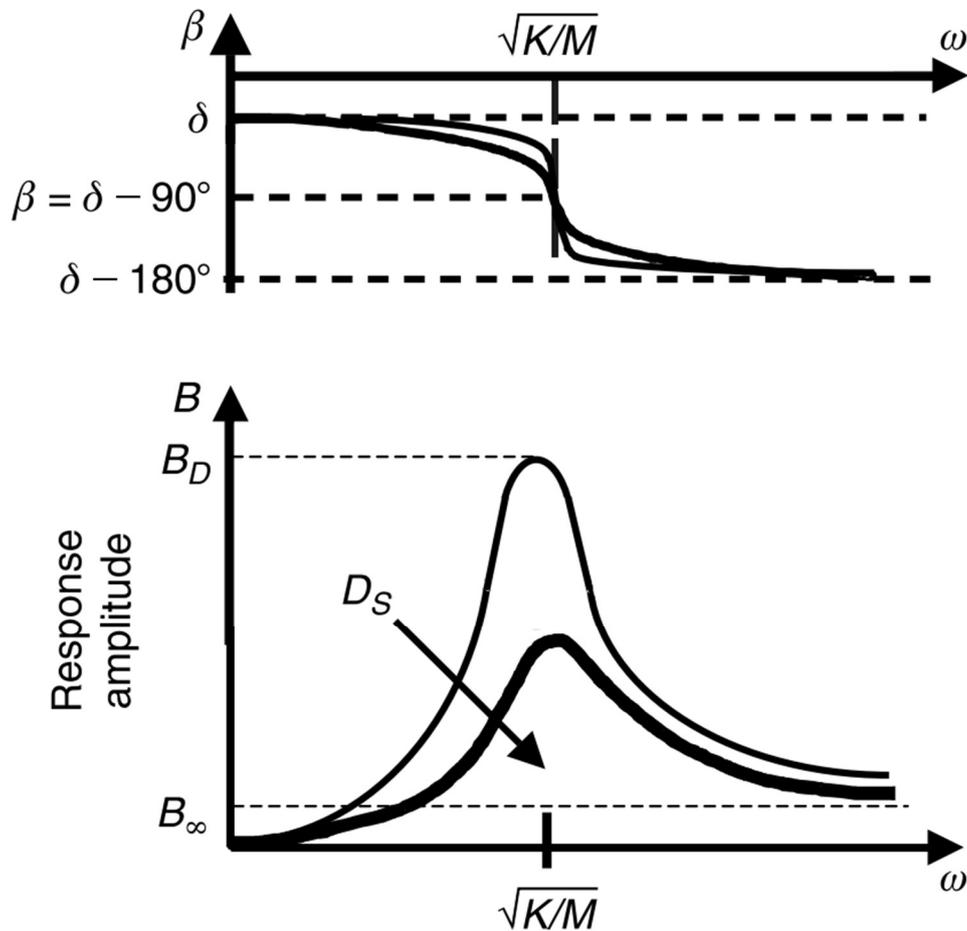


Figure 14. Bode plots of the rotor response phase (top described by equation (5)) and amplitude (described by equation (3), (4), and (6)) [8].

Similar behavior of the system is observed during measurements with portable instruments connected to eddy current sensors. Figure 15 presents an example of a Bode plot from a single eddy-current (i.e., relative vibration) sensor obtained during field measurements.

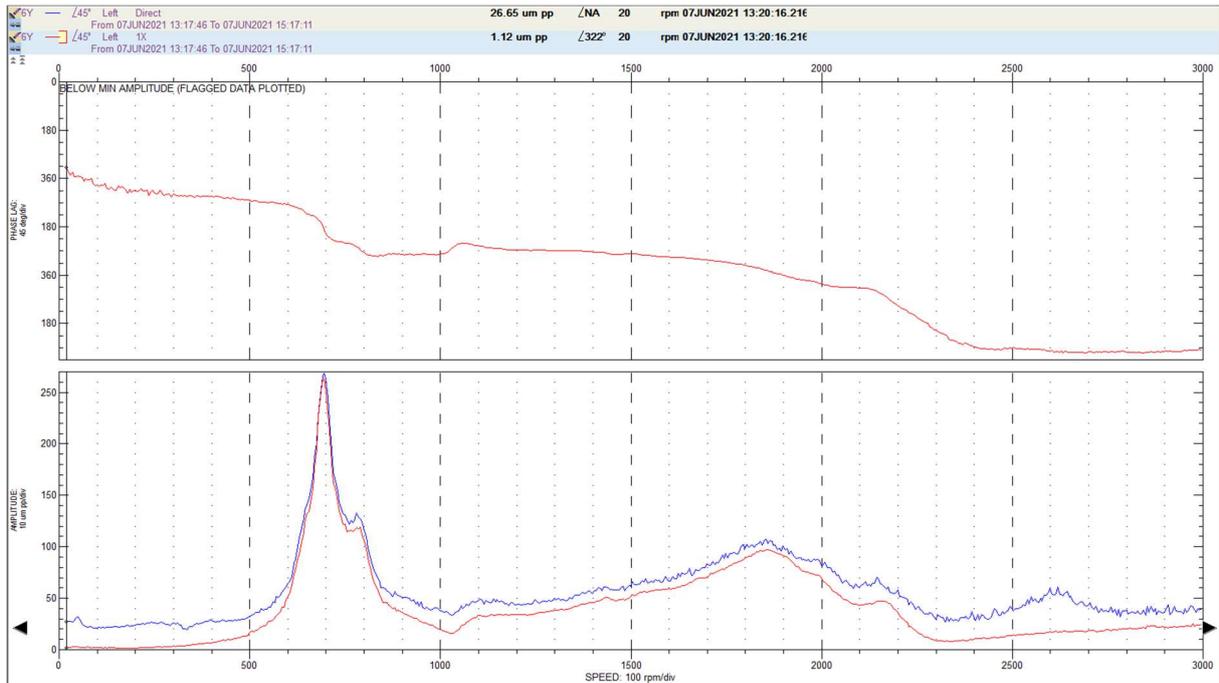


Figure 15. Bode plot of the real system.

The upper diagram shows the change in phase lag, and the lower chart shows the amplitude for individual revolutions. The graphs for Figure 15 show the amplitude-phase response of the system, at the measuring point, to the variable synchronous excitation, which is the centrifugal force.

1.3.2. Unbalance response of the system undergoing transient states

The most common malfunction in rotating machinery is the unbalance of one of its rotors. The force that generates the unbalance on the rotor is related to the centrifugal force shifting the center of mass of the rotating rotor away from the inertia axis of the rotor. For example, in Figure 13, marked as point *C*, the radius *r* on which the unbalance mass *u* acts with rotational speed ω creating a rotating centrifugal force. The force creates rotor lateral movement perpendicular to the rotor axis inside the bearing. It can be approximated by the simple equation, i.e., $F = mr\omega^2$. The force generated by the rotor imbalance is proportional to the square of the rotational speed at which the rotor spins.

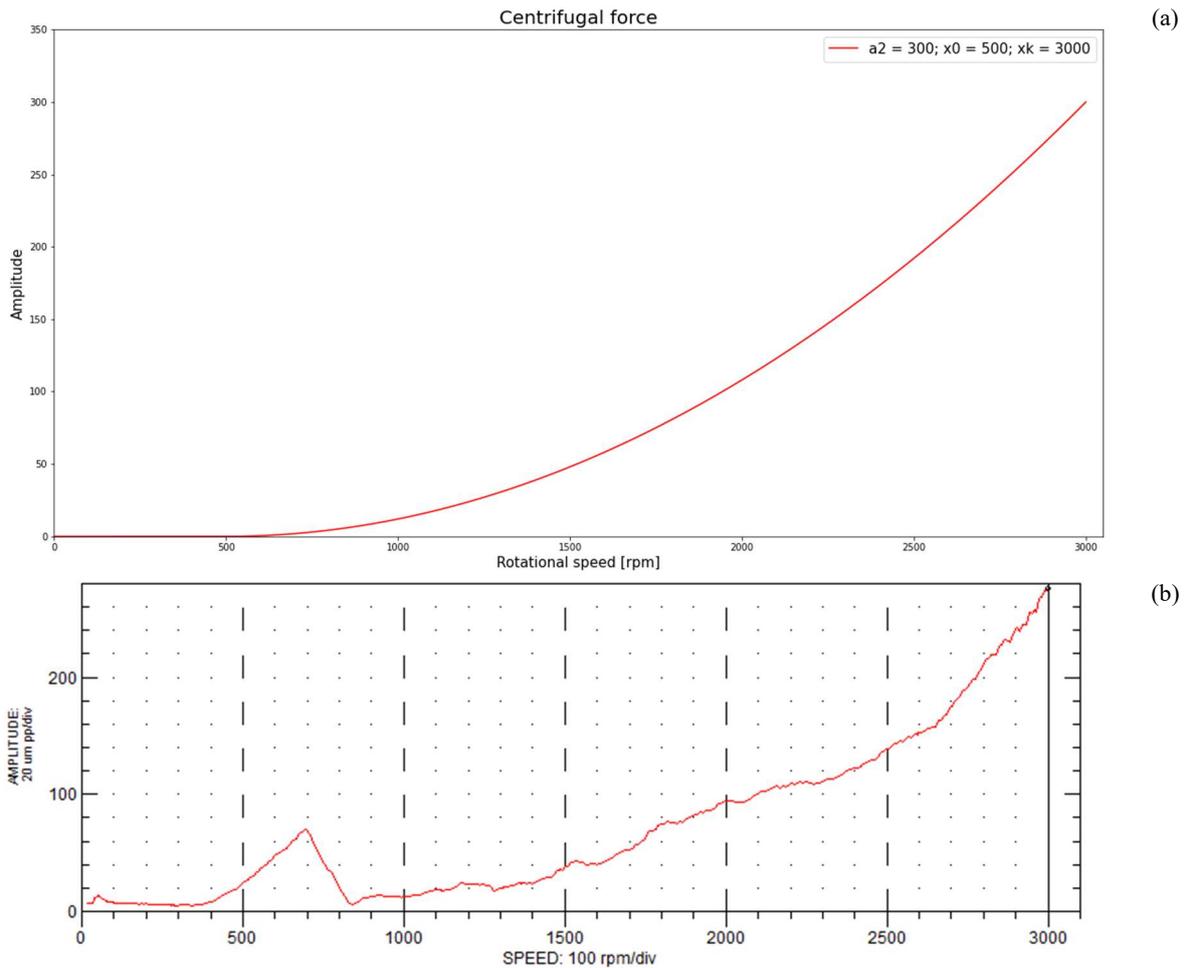


Figure 16. (a) – centrifugal force simulation; (b) – real data example of the unbalance rotor system response.

Figure 16 (a) shows an exemplary diagram of the centrifugal force acting on the rotor during a change in rotational speed, i.e., during a transient state and the Table 1 summarize the simulation basic values together with their description.

Table 1. Rotational speed simulation parameters.

Description of the unit used in the simulation	Unit	Value
The highest rotational speed value during simulation	x_k	3000rpm
Centrifugal force acting at the highest rotational speed value	a_p	300
Centrifugal force entry point (since when centrifugal force contributes to the system response)	x_0	500

The similar nature of the system's response to excessive imbalance force has been confirmed many times by the author's research on real objects. Figure 16 (b) shows an example of the system response in the form of a relative shaft vibration measured at one end of an unbalanced rotor.

The centrifugal force from imbalance only makes a significant contribution starting from 900RPM due to the rotor transient's nature, as shown in Figure 16 (b). From around 400RPM to around 900RPM, the rotor passes through its resonance speed range. During transient states, the rotors of steam and gas turbine sets go through at least one, and sometimes even two, regions of resonant speed intervals.

1.4. Problem formulation

After presentation of the basics of the large rotating machines and an outline of the nature of the response of their transient states, we can define the motivation for the actions taken in the thesis.

1.4.1. The motivation

Fault detection and identification is not an easy task for condition monitoring of the large rotating machinery. The reliable vibration measurements of high-power energy machines requires an advanced signal processing system, including spectral analysis and order analysis. Tracking filters are used during start-ups and coast downs, which set the synchronous component to the current rotational speed of the turbine set. Due to fluid-film bearings used as the supporting structures of the machine, the complexity of the rotor-to-stationary part relations is even more complex. One can find an excellent introduction to the subject in the widely recognized book by Bently and Hatch [10]. Interested readers can also refer to the work of Vance et al. [6], who presents the entire course of rotordynamic analysis. In-depth and very detailed rotor modeling examples, in turn, can be found in the books of Muszyńska [8], Kiciński [9], and Eisenmann [3].

Along with regular rotor models, they propose and explain several malfunctions, such as fluid induced instabilities presented in [8] and [10]. Complex and accurate models, mainly focused on 200MW class turbines, were proposed by Kiciński in [9] and [10]. These advanced models were based on the FEM (Finite Element Method) to analyze a rotor's behavior. On the other hand, a rotor rub fault requires a different approach to modeling, and the measurement of torsional vibration appears to be very helpful in this process. Interested readers can refer to the work of [11], where the author presents the study in a very detailed manner.

According to the works of Muszyńska [8], Bently and Hatch [4], Adams [11], and Eisenmann [3], the primary and most common malfunctions are:

- unbalance,
- misalignment,
- rotor rub,
- oil-related and steam-related instabilities in bearings and seals, respectively,
- bearing overload/ underload.

Even though these failures are the most frequent to be encountered, it is tough to distinguish them without extensive experience and expert knowledge. Some of them can exhibit similar features during steady-state operation. In such cases, transient signal analysis is the best way to analyze vibration data to produce a reliable outcome. The raw vibration signals are very complex, and the measurement database size is enormous.

A few types of signal features describe machine dynamic state, such as overall level of vibration, both relative (measured in peak to peak or 0 to peak, respectively) and absolute (measured in RMS – Root Mean Squared, or 0-pk – zero-peak). Other commonly used features are synchronous system response (1X), second super synchronous system response (2X), and sub-synchronous, as Bently and Hatch [4] and Eisenmann [3] described. It is important to track both amplitude and phase for the harmonic

features. Eddy-current sensors provide all the necessary information to obtain the aforementioned features.

Due to the rarity of transients in the machine's life cycle, the author had a limited amount of them collected. Although the author could label some transients from his measurements as defective, they did not account for even five percent of all measurements collected throughout the author's career. This made it impossible to use machine learning and artificial intelligence methods that need training and test data to train the models. It is especially important regarding the data from faulty states.

To overcome this problem, the collected data were used to create the foundations of a system based on data-driven methods that turned out to be effective and robust.

The system proposed by the author will operate in the background, using the data from the turbine set monitoring system. It is activated after every coast down or start-up. The system will automatically assess the data. It will allow further operation, as long as vibration response during transient will be qualified as correct. In the other case, it will perform the fault identification and inform about possible cause of a problem.

The methods proposed in the dissertation below are relatively simple. This is because it should be used in practice by non-experts. There is a theoretical disadvantage that system may not be able to detect the smallest changes and deterioration of dynamic state and will send alerts only when the condition of the machine deteriorates noticeably. The simplicity (and hence resistance to disturbances) and robustness are decisive advantages of the proposed approach. The system will be able to report the dynamic state change and to avoid false alarm indications. This will contribute to the increased trust to the indications of the proposed FDI (Fault Detection and Identification) system.

Let us first focus on the main component of the authors research – the transient state. During a change of the rotating speed, excitation force (centrifugal, synchronous force) will vary, changing the stiffness and damping relations of the system. These changes are an inherent part of the flexible rotor design. Machines in a correct state also experience the same mechanisms, although their vibration response is very different. If we assume a rotor-to-bearing system without malfunctions, the transient curve will have a set of parameters to reflect its behavior. This set will include the peak of the resonance response, width of the critical speed interval, vibration amplitude at Full Speed No Load (FSNL) state, etc.. That same system with developing malfunction will produce a different response, and the parameters reflecting its behavior will have different values.

The majority of methods available in the literature consider only steady-state operation of the machine. Brito et al. in [12] showed advantages in unsupervised learning and its incorporation in rotating machinery fault pattern detection and diagnosis. It constitutes a methodology to detect a fault mode and predict its trend. In fault diagnosis, they used the black-box model approach and shapely additive explanations method. He used unsupervised classification and root cause analysis to produce a diagnosis. They present several advancements in fault detection, diagnosis, and prognosis in rotating machinery. From this perspective, automatic detection of common malfunctions can become an interesting complement to standard monitoring equipment.

In large turbomachinery, this is not the case. Machine Learning approach is very hard to apply. The major problem is that fault mode data are often not available. Furthermore, there is no available training data set due to the low rate of transient states during the machine's lifetime (and even less with a fault). Additionally, a few state-of-the-art anomaly detection algorithms are examined. Thus, there is a shortage of techniques dealing with the transient states, especially for large turbogenerators, for which the transient data sets become very large both in terms of data points and in terms of time, depicted in Section 2.1 and shown in Figure 19 as coast down, and in Figure 20 as start-up.

The author in [13] and in [14] proposed foundations and the basic considerations of automated turbomachinery fault detection and identification, respectively. The concept was proposed for a single channel only with unprocessed data and without any severe malfunctions. For such a relatively simple case, one probe is sufficient. However, taking into account the whole turbogenerator shaftline relying on a single sensor often lacks essential information. Further-more, different malfunctions can exhibit itself in different parts of the machine during different circumstances. Therefore, for large turbogenerator sets, multichannel analysis is a necessity. Analysis should use different features from a single sensor (i.e. overall vibration amplitude, its first harmonic and phase, second harmonic, subharmonics, and others). Also, the investigation can incorporate different sensors from the same bearing (oriented orthogonally in the bearing plane). Finally, the research can use sensors at different axial locations along the shaftline. The above-described analysis challenges are why the original method, proposed in [13] and [14] for a single sensor, must be extended to a multidimensional case.

A maintenance strategy that enables detecting malfunctions at the early stages of their evolution should play a crucial role in facilities using these types of machinery. The best data source for assessing the technical condition is the transient data measured during start-ups and coast downs. Most of the automated methods proposed in the literature are applicable to small machines with a rolling element bearing, during a steady-state operation with a shaft considered a rigid body. Large power turbomachinery express a very different behavior. They operate above their first critical rotational speed interval, and thus their shafts are considered flexible. To make the case more complex, these turbines are equipped with hydrodynamic sliding bearings. Such an arrangement introduces significant complexity to the analysis of the machine behavior, and consequently, analyzing such data requires advanced rotordynamics knowledge and field experience. Typically, after each transient machine data should be investigated to check whether the dynamic state is satisfactory and the machine can be operated safely. Such a step requires advanced measurement equipment, which is not always accessible, and availability of a skilled expert, which is costly and must be scheduled according to the availability.

The goal of author's dissertation was the research of a method to diagnose large rotating equipment in an autonomous way, so that the load on experts can be reduced to really important cases. In the dissertation the author proposes the anomaly detection method which he named the Operating Envelope (abbreviated as OpEn) and the fault identification method, which he named the Multidimensional Data Drive Decomposition (abbreviated as MDDD or MD3). Combination of both methods extend the single sensor concepts proposed in [13] and [14] to a functional and autonomous multi-channel fault detection and identification system.

Some works heading in the same direction were published in recent years. For instance, Bielecki et al. in [15] proposed a simple yet effective method for unsupervised monitoring of rotating machinery for failure detection in the early stages. Lei et al. in [16] incorporated unsupervised feature learning on big data set to diagnose the motor and locomotive bearing faults patterns. Wang and Sun [17] used the combination of wavelet decomposition sparse filtering networks and a support vector machine to establish fault diagnosis in the motor bearings. Authors adopted the decomposition concept to their research. All these works consider smaller machines during their steady-state operation. There are lack of works that take into account transient states of a large turbomachinery.

The data-driven methods (OpEn and MD3) developed in this dissertation for the analysis and automatic diagnostics of failures are driven by the type and nature of data obtained during large turbomachinery measurements. Therefore, the methods proposed by the author in the doctoral dissertation are a compromise of the amount of available data and the accuracy/repeatability of the results. The dissertation is also a result of over 13 years of industrial practice combined with experience and expert knowledge in the field of signal processing, rotor dynamics and sliding bearings.

Data-driven methods have advantages and disadvantages. The advantage of the methods proposed in this work are their simplicity. The advantage of simplicity is clarity in interpreting the results and a straightforward implementation process. No expert knowledge is required to read the method's indications contrary to, for example, FEM methods, where expert knowledge is essential for diagnosis. The disadvantage of simplicity is the possibility of "insensitivity" and overlooking the nuances of the early changes in the behavior of the turbine set.

Such a limited set of data available for analysis makes it impossible to use artificial intelligence methods and algorithms and advanced novelty detection algorithms. Nevertheless, the methods proposed by the author turned out to be robust and were positively validated on both laboratory data and data from real objects.

1.4.2. Goal of the thesis

The main scientific goal of the dissertation is proving of the statement:

It is possible to detect and identify faults of the large turbomachinery by an automated algorithm using analysis of transient data.

This goal was achieved by dividing the whole work into several steps, which are listed below:

- Development and proposal of system architecture.
- Development of data preprocessing methods.
- Analysis of the correct dynamic state databases and selecting reference data for baseline evaluation.
- Invention, creation, and development of the fault detection (OpEn) method.
- Establishing the upper and lower values for the OpEn 2D case and the ellipsis axis values for the OpEn 3D case.
- Establishing the severity parameters for the OpEn 2D and 3D cases.
- Invention, creation, and development of the fault identification (MD3) method.
- Proposition the algebraic representations of the decomposed functions.

- Defining the fitness functions for the Differential Evolution algorithm.
- Estimation of parameters of decomposed functions (adopting Python's DE algorithm code to find the decomposed function parameters).
- Implementation of the complete system in Python.
- Planning, preparation and execution of experiment on a test rig.
- Validation of proposed methods on laboratory data.
- Validation of proposed methods on real object data.

As a result, the dissertation aims to create the foundation framework, methods, and procedures of automated vibration data assessment to enhance maintenance reliability. Automatic fault detection and identification (FDI) tool can help maintain the reliability and safety of equipment in industrial plants.

In addition, such a tool will facilitate work of operating personnel in turboset dynamic state assessment. It can also enable them to react faster to changes in the dynamic response to avoid critical failure and reduce the downtime to a minimum. Furthermore, information on the dynamic state of operation and its condition based on transient response can help management personnel plan essential repairs ahead. As a result, it can minimize the risk of long unplanned stand-still, overhaul, and repair. Such a situation can reduce the company's financial results and make the enterprise non profitable. Last but not most important, increasing the reliability resulting from information from the procedures and methods contained in this doctoral dissertation may increase the safety of devices and people operating these devices by minimizing catastrophic damage to machines.

1.4.3. Scope of the thesis

The doctoral dissertation is structured as follows. The first chapter introduces the subject of large utility rotating machines. It includes a guide to vibration measurements of large rotating machines. Then, the most important values for assessing the dynamic condition of turbine sets are presented and described. It also includes an introduction to selected aspects of rotor dynamics needed for the research. The chapter ends with the formulation of the problem and the thesis, and the aim of the doctoral dissertation.

The second chapter describes the work research object in greater detail. It shows the subjects on which the author performed the research. The specificity of the measurement of vibrations of large rotating machines and its influence on the selection of measuring equipment is discussed. Also presented are instruments used in measuring large turbine sets in the industry.

The third chapter describes the process of collecting and processing transient data points. First, the measuring equipment used during diagnostic tests is described. Then, the entire configuration of the measuring equipment is presented and discussed. Next, the transient data points sampling rates are described in detail. The chapter also covers the most critical data structures that are needed and used in automatic fault detection and identification systems. The last part of this chapter discusses the issue of interpolation and its use in the data preprocessing process.

In chapter four whole fault identification system is laid out. First, the author describes essential aspects of an automatic fault detection system. Then, the term Operating Envelope (OpEn) method is

introduced. This chapter describes the transient selection method for baseline measurements from which the acceptance region is calculated. Next, the term OpEn 2D is introduced, the OpEn method, which applies to only one vibration signal feature. Next, the severity parameters for the OpEn 2D method are given and entered. Then the author introduces the concept of OpEn 3D as an OpEn method for vibration signals consisting of two features. Finally, the method of obtaining baseline measurements for this case and severity parameters, which are used to assess the severity of the detected malfunction, are given.

In chapter five, the entire system for identifying malfunctions is presented. In the beginning, the author introduces the method of parameter identification. Then the methods of identifying inequalities are discussed. Next, the concept of the Multidimensional Data Driven Decomposition (MD3) method is described. Finally, the author presents the application of the Differential Evolution (DE) algorithm to identify the parameters of decomposed functions.

The sixth chapter describes the architecture of the automatic detection and identification system for large rotating machines proposed by the author. First, the author proposes a place for his system in the monitoring and diagnostics systems of the existing power plant systems. Then the pseudo-codes of the most important modules of the whole system are presented and explained. Then, the Python code for all the most essential functions and procedures is presented and described.

The seventh chapter describes the validation of the OpEn method for both the 2D and 3D cases. First, the baseline measurement is presented. The baseline measurements are a prerequisite for validating the results of both methods. There are tables with baseline values for the entire turbo set consisting of high-pressure (HP), intermediate-pressure (IP), and low-pressure (LP) cylinder and the generator rotors. Then the OpEn method validation process for the 2D case is presented. Finally, the OpEn 3D method is validated on another dataset used in the identification method's validation process to validate the entire diagnostic system.

Chapter eight describes the entire MD3 method validation process. First, the author used the method validation on test-rig model data. In the first part of the chapter, the author discusses the model data in detail, as well as the simulation results and method testing. Then he used the validation data of the OpEn 3D method from the previous chapter. Finally, the author gives a set of function parameters of decomposed functions. For this case, a scenario is indicated, i.e., a set of parameters of decomposed functions that best match the real transient.

In the ninth chapter, the author presents the conclusions of the following work. Detailed conclusions regarding the OpEn as well as MD3 methods are presented. Finally, the author outlines the directions for further research and potential improvement of the system.

2. Measurement data of large turbomachinery

The vibration data taken at large turbomachinery are very specific. The data acquisition equipment is highly specialized and is seldom used for any other machinery, due to high cost, set of functionality and required reliability. It also has several features, which requires specialized pre-processing before further data processing methods can be applied. These aspects will be presented in this chapter.

2.1. Design and operation of large steam turbines

The research subject is a steam turbine rigidly coupled to a generator of high output power. The research in this doctoral dissertation focuses mainly on machines of the +200MW class. To validate the author's proposed method +500MW class turboset research was also included.

The share of large turbomachines in professional power sector markets varies depending on a specific country and grid settings. As far as Poland is concerned, an essential part of the turbogenerators are units of +200MW type (the 13K215 with 13MPa live steam pressure, 535°C of live steam temperature, condensate operation unit, and 215MW of power output and its modifications up to the 13K242 – the same parameters, but almost a 15% increase in power output). This unit type is the most common large turbo-set found in Poland's power generation industry. More than 50 units of this type operate until now, and they contribute to almost half of the national power production capabilities. The 18K360 are the second most crucial type of turbo-sets. There are 16 such units in Poland. Additionally, a few units were built in the last several years with a very high-power output ranging from 800MW to +1,000MW. Due to the Polish power generation structure, my objective is to focus on +200MW units at first, and afterwards extend to units with a similar transient behavior.

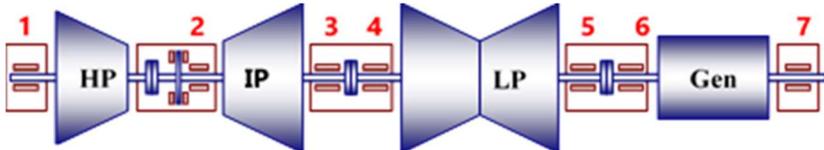


Figure 17. The layout of a 200MW type turbo-set. The bearings are numbered from the HP side (the figure prepared by the authors).

A typical 200MW class turbogenerator consists of three cylinders (HP for high pressure, IP for intermediate pressure, and LP for low pressure) and one generator Figure 17 presents the entire setup of this type of machine. The author based the validation of the method on a turbines with a power output of +500MW. They have a similar kinematic scheme and the dynamic behavior. Figure 18 presents the layout of its arrangement.

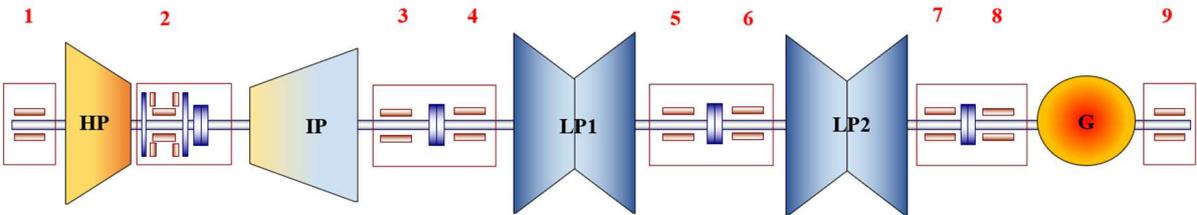


Figure 18. The layout of a +500MW type turboset.

Turbines of this type have one HP cylinder, one IP cylinder, and two LP cylinders. They are also rigidly coupled with a generator. Despite the extension of the shaft line with an additional LP cylinder, the nature of the generator rotor response is very similar. The author confirmed the similarity of their transient responses throughout over 13 years of research on these two types of devices. The validation of the MD3 method results was performed on the responses of the unbalanced rotor of the + 500MW turbine generator set.

In the power generation critical machinery such as turbogenerators are assumed to run smoothly for the whole lifetime period – often more than 30 years. Especially in large units, which are equipped in oil lubricated hydrodynamic bearings, vibration severity criterion is one of the most essential monitoring parameter as for the machine's mechanical condition. Mechanical vibration is the source which contains most information about the health of the component.

Unambiguous data are needed for comparison, to determine the baseline measurements, perform the anomaly detection task, and identify the parameters of the decomposed functions. The best data that defines the correct condition of the turbogenerator are vibrations of the shaft journal in the bearing bushing. The data for these machines' research is characterized by substantial changes during the operation of the device. Therefore, they can take different values depending on the machine's operation point. For example, during an idle run, i.e., Full Speed No Load (FSNL), the turbo-set may have a different vibration response than in the full-load operation of the generator. This is because the turbine's rotational speed is at its nominal value on the idle run. Still, the generator does not produce electricity (the generator is not synchronized with the power grid yet). Therefore, the turbine can move relatively freely because of moderate steam-related forces only. Finally, when the turbogenerator works at full load, the shaftline is subjected to maximum forces (the generator rotor load, steam-related thrust force, temperature vector, etc.). To the contrary, during the coast down the turboset is unaffected by any forces other than the inertia. Figure 19 shows the coast down curve of a 200 MW class turbogenerator. As shown in the picture, the duration of the coast down may exceed 120 minutes (time from 19:28 to 21:52).

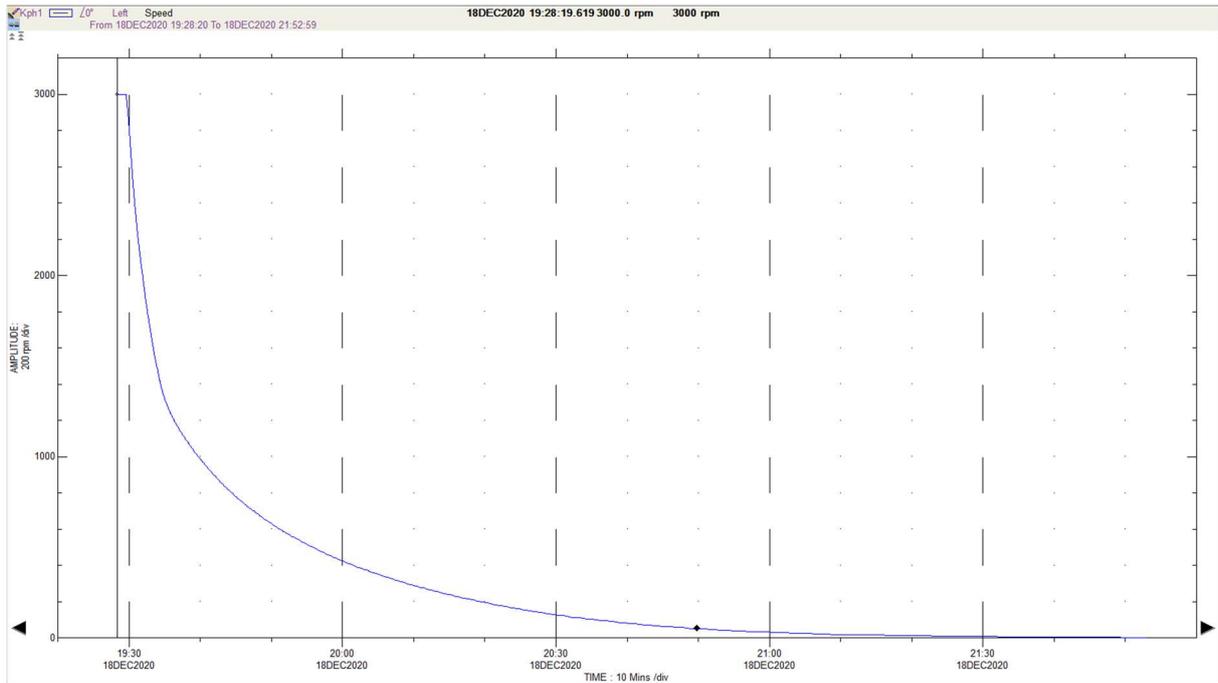


Figure 19. Shaftline inertia curve (coast down example) of the 200MW class unit.

Similarly, when starting the machine (i.e., during run-up), the forces from the steam are also not significant. Therefore, in the author's opinion, based on the experience gathered during many years of research on energy machines, transients' states (i.e., both run-ups and coast downs) are the best ways to assess the correctness of the technical condition of high-power energy devices.

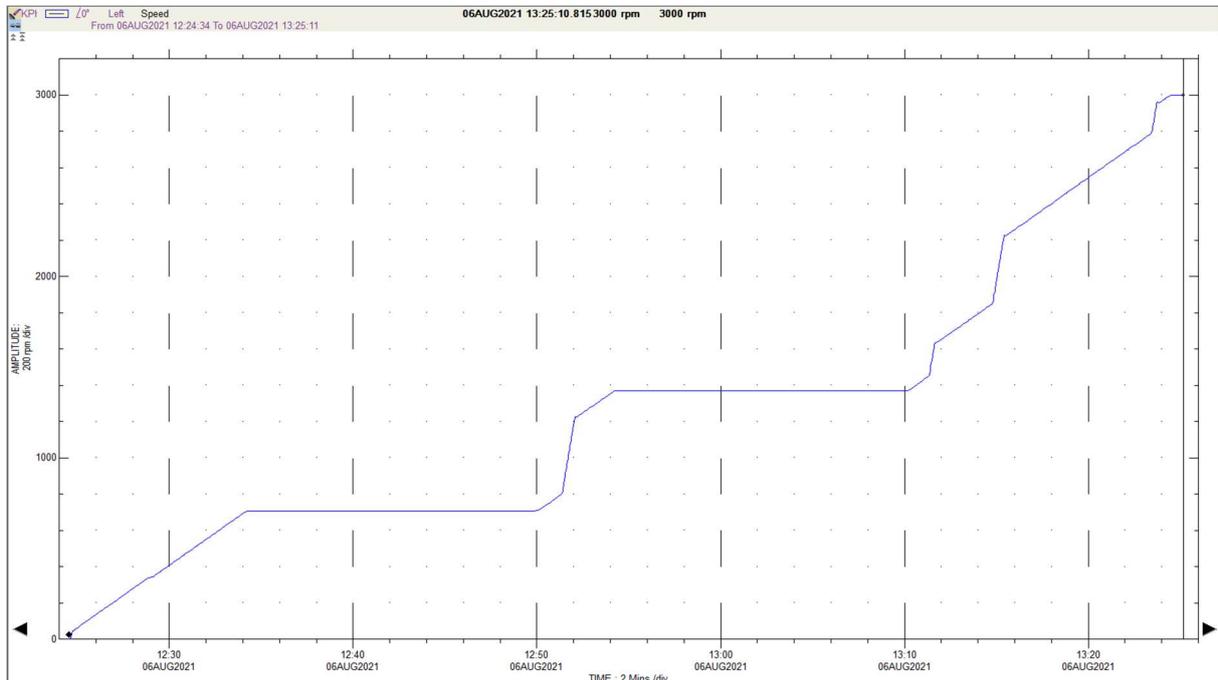


Figure 20. Start-up curve of the 200MW class unit

There are several methods to measure vibration during the transient state. Each turbo-set is equipped in an online monitoring system. Such systems continuously measure vibration and calculate simplest features (peak amplitude, rms, gap, etc.). When vibration level exceeds configured level of alarm, the monitoring system will send the trip signal to the turbine controller, which will cut off the steam and

cause the turbine to trip. Although they may bring some help to analyze the machine condition, measurements from the monitoring systems do not provide much diagnostic information. For example, they often do not calculate vibration features, despite having such capabilities. Neither, they do not store the history of operation, nor raw vibration signals for a more thorough analysis. Therefore, there is another method used to collect the data for transient analysis, namely advanced portable data collection systems. Measurements with portable diagnostic equipment are widely used for in-depth analysis and diagnostics of the technical condition.

The experience in configuring the equipment of monitoring systems in power plants in Poland and abroad allows the author to assume that with a small amount of work on the part of the devices' users, it will become the basis for the implementation of the proposed autonomous FDI system. The system proposed in this dissertation can work as extension of the standard monitoring system. Furthermore, the FDI system will become an invaluable help during unexpected shutdowns during which no measurements by portable devices are gathered. Finally, such an arrangement will help assess the machine's condition on an ongoing basis and indicate places of the evolution of potential malfunctions.

The prerequisites to the work in research of autonomous algorithms are access to real life data and extensive experience in the field of transient data analysis. During the author's 13 years of experience in research on large rotating machines, he developed plans and schedules for the commissioning of power equipment and diagnostic tests in terms of measurement, vibration analysis, and evaluation of the dynamic state of turbine sets. He carried out diagnostic tests by analyzing data from portable equipment. The data was collected after a previously prepared and agreed plan for conducting measurements and plant operation. This is worth to emphasize, as such plans may affect a country power generation system and must be agreed upon with the plant and the grid management authorities. Thus, the author has a unique position and ability to undertake the dissertation task.

In the second scenario, the automated FDI methods can be used on the data collected by portable equipment. Then, with implementation of automated analysis of transient data, collected by portable diagnostic equipment, manual analysis by a human expert can be limited to only vital examinations. This also helps to reduce the time and costs associated with planning and diagnostic measurements, reducing the maintenance costs.

2.2. Portable measurement systems

Since the dissertation is based on data collected by portable systems dedicated for large turbomachinery, it is important to present the specifics of these data acquisition and analysis systems. There are several companies worldwide who manufacture advanced portable data collection systems, suited for large turbomachinery. Figure 21 shows examples of data acquisition units offered by several market-leading companies.

The acquisition unit should be configured according to the ISO standards (ISO 20816-1). This standard defines only the fundamental frequency range for assessing the technical condition of a machine based on its vibrations of shafts and bearing caps. To perform full diagnostics of shaft lines and bearing supports, more advanced and detailed settings of the measuring equipment are often needed. In this

Section, the author gives the complete Digital-to-Analog Interface Unit (DAIU) configuration for the full diagnostics of large rotating machinery.

The measuring instruments, used in the diagnostics of large rotating machines, are industrial computers with specialized signal conditioning and high-end data acquisition. These are expensive devices that can cost over 100kUSD. These DAIUs are equipped with fully configurable analog-to-digital converters. The inputs of these devices can measure various signals:

- Static signals - values from temperature, pressure, and valve position sensors.
- Dynamic signals - coming from all types of sensors (described in Section 1.2). In addition, some DAIU's input cards have built-in accelerometer sensor power supplies.
- They can collect more than 20 dynamic signals simultaneously.
- They are equipped with tachometer sensor inputs. They can determine the rotational speed of the machine. It is used to synchronize all the sensors features to the first harmonic of rotational speed and its factors. Additionally, these inputs can be used to set the tracking filter, calculating the harmonic components. The tachometer inputs are fully configurable, and they can handle outputs from different sensors like eddy-current probes, laser and optical sensors.

Along with the device, an advanced software for configuration and analysis of rotor dynamics is delivered, dedicated to a given unit, as depicted in Figure 21, top left and right.



Figure 21. Examples of DAIUs from leading manufactures. The top: left – Siemens [18], right – Meggitt [19]; in the middle: left - Bently Nevada [20], right – OROS [21]; in the bottom: left – Emerson [22], right - Bruel and Kael [23].

Depending on the equipment's configuration and needs, measurement data can be saved concerning time or the change of rotational speed increment or two of these parameters simultaneously and independently. These computers can write substantial amounts of data to their internal disks. They can also convert the necessary diagnostic features on-the-fly and present them. In addition, depending on the software, the tachometer input and analog-to-digital converter inputs can be used for modal analysis.

The author works for a company equipped with measuring equipment from Bently Nevada. Hence, the author has the most experience with such equipment. However, each of the other companies listed in Figure 21 offers comparable products.

3. Transient data preparation and preprocessing

This chapter describes the various steps of data preparation that will be used by the automatic fault detection and identification system. First, it contains the structure of diagnostic data used in the system. Then, it presents the individual stages and results of data processing.

3.1. Portable equipment used for data collection

After each diagnostic measurement, the data are saved in the acquisition unit memory as a database. Then the database has to be downloaded from Digital-to-Analog Interface Unit (DAIU). DAIU is an industrial computer capable of acquiring, processing, and presenting collected dynamic signals, as presented in Section 2.2. In addition, it can export processed data as text files. This operation enables further analysis of vibration features.

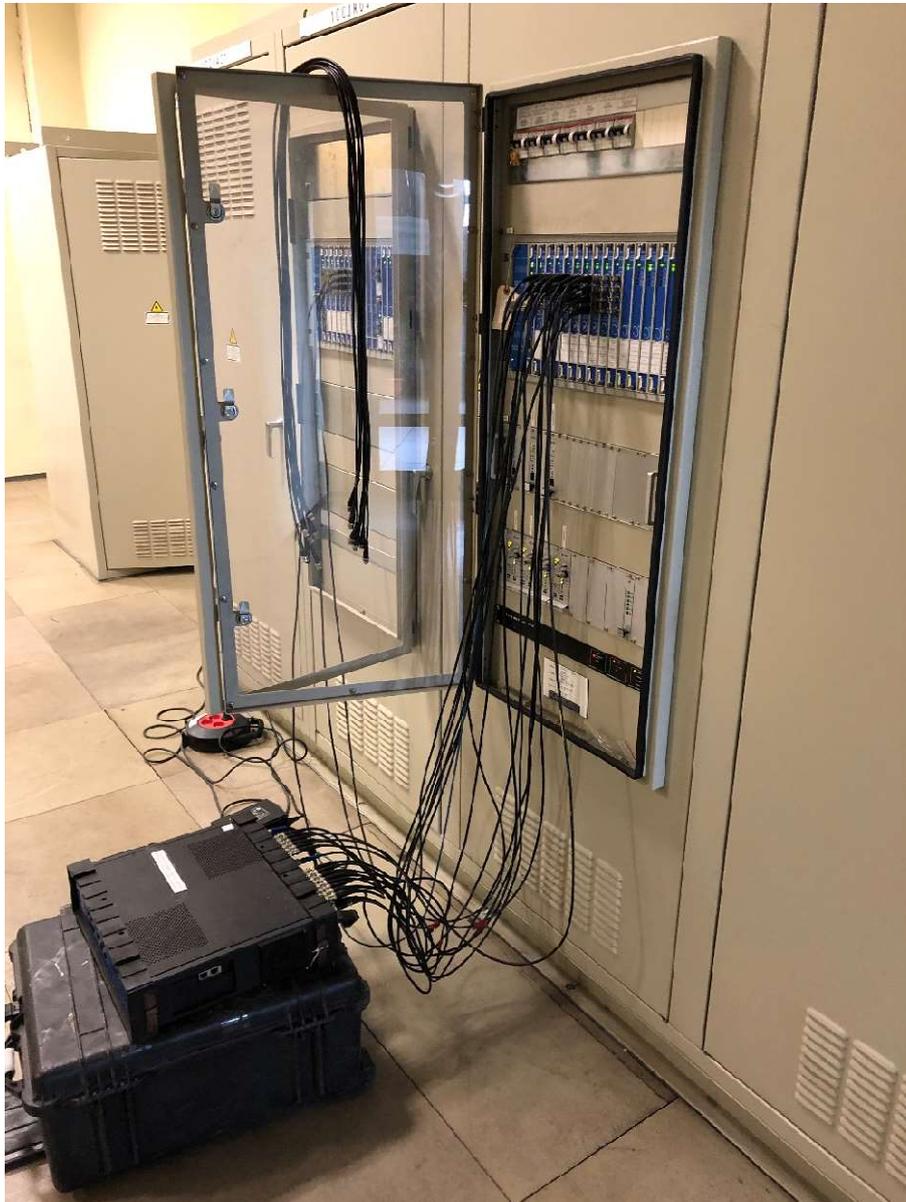


Figure 22. An example of the vibration measurements setup – acquisition units (two DAIUs) connected to the stationary monitoring system in the control room.

The DAIU unit is connected to a stationary monitoring system to collect data from eddy current sensors. Figure 22 presents an exemplary setup of DAIU during the measurement course. DAIU is connected to buffered outputs of a stationary monitoring system.

3.2. Measurement system configuration

In this section the configuration and preparation of measurements with portable instruments is presented. The software for configuration, acquisition, processing, saving, and presentation of data used by the author is a part of the measuring equipment supplied. This program is the ADRE Sxp. However, any other apparatus, described in Section 2.2, with the same settings can reproduce the author's research repeatedly. This section is quite technical but is necessary to fully understand the structure of the data used for further analyses..

Figure 23 and Figure 24 presents the dynamic channel and one-per-revolution reference mark setup details used during each measurement course, respectively. Table 2 and Table 3 summarize the most important configuration steps and options set up by the author during his research.

First, in the General tab window, one can set the general properties for all the sensors participating in the measurement. Next, the "Location" column describes the placement of each sensor on the rear panel to be easily identified. "B" letter stands for the Box in this column, i.e., an acquisition unit. Letter "S" - means slot, i.e., this identifies the dynamic sampling card in the Box. The best-equipped boxes can hold three "slots" (i.e., three dynamic sampling cards). Each card can contain eight channels. Finally, "C" is the channel number on the "slot". Finally, the "Channel Name" column specifies a unique sensor name during the measurement course used in the analysis process. The "Machine Name" column is not obligatory but describes the placement of the sensor. In the example presented in figure 2, the 1X sensor measures HP vibration. The values in the "Keyphasor" column are the most important ones to set up in this tab. It assigns the reference one-per-revolution mark sensor described in Section 3 to a dynamic sampling channel that measures vibrations. Only the reference mark sensor enables vibrations feature extraction that the author uses in this dissertation. Finally, the "Transducer Orientation" column sets up the angular orientation of the sensors to an arbitrary reference direction - in this case, "Up". For example, suppose we set the value in the "Direction" column to "45-Right-Radial". In that case, the sensor location is 45 degrees from the vertical axis in the right direction, perpendicular to the shaft axis. The "Transducer" tab sets all the sensor properties up, like Transducer type, type of the measurement, the minimal and maximal value measured by the sensor, its sensitivity, coupling, the sampling mode, and the bandwidth filter properties. Next, the "Variables" tab configures the low and high frequencies for the bandpass filters and the number of its poles on the input. The NX-1, NX-2, and NX-3 columns set up the sub-synchronous feature of the system response to 0.5 times the whirling frequency, super synchronous features of the three and four times, respectively. In the final step, one can configure different types of waveforms.

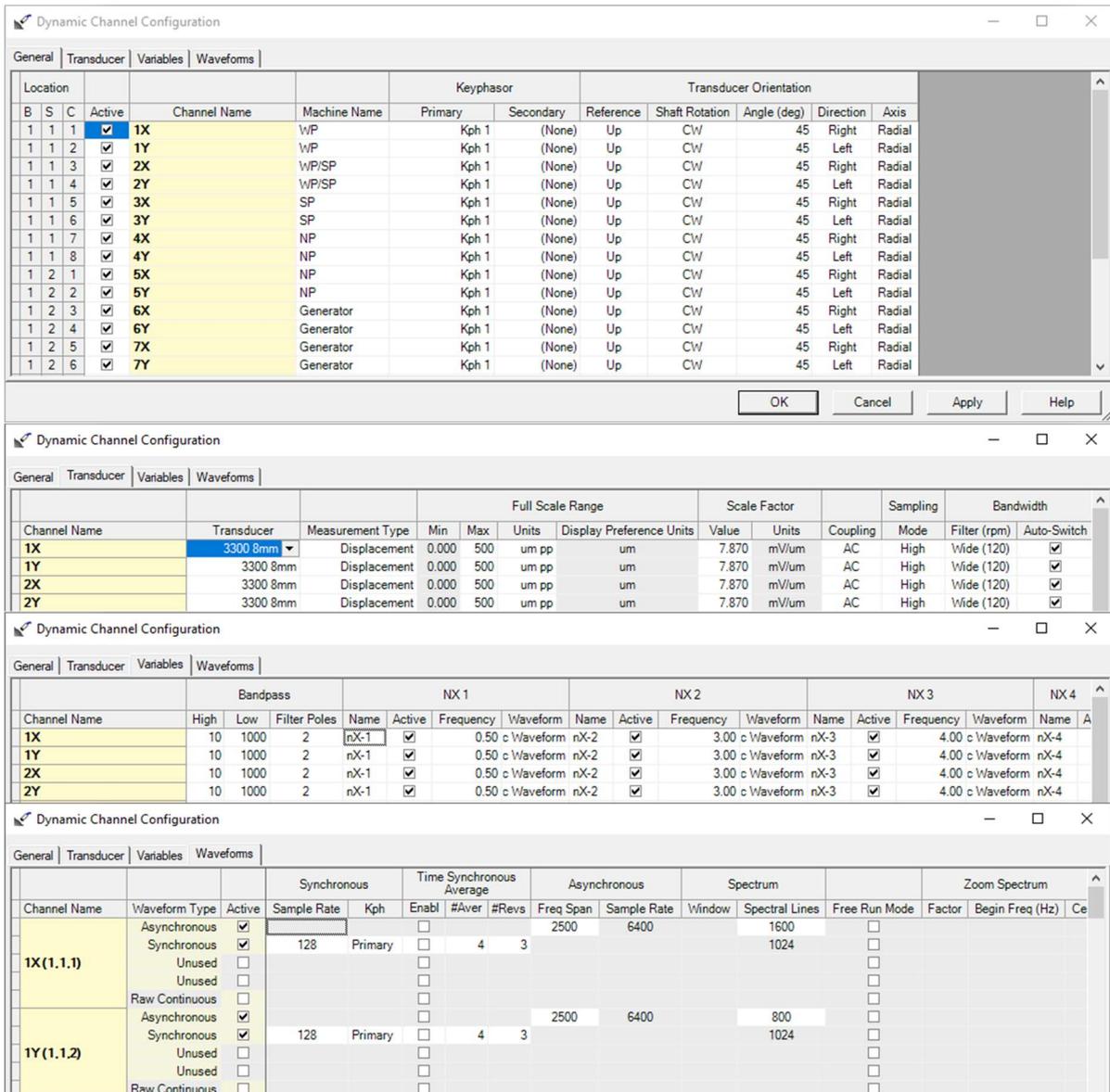


Figure 23. ADRE Sxp dynamic channels setup.

After years of experience in research and analyzing the machines in question, the author decided to introduce asynchronous and synchronous waveform settings to the measurement procedure. It allows additional accuracy in the spectrum and orbits analysis and becomes the foundation to formulate the baseline values as the turbine reference behavior.

Setting up a one-pre-revolution reference mark configuration is similar to that for dynamic channels. The eddy current sensors described in Section 4 are used for the reference mark measurement. General tab configuration extends only by the maximum value of the measured rotational speed. In the "Transducer" tab, the sensor type, measurement unit of the sensor, its sensitivity, and Coupling and Sampling are set. Using DC coupling for the constant synchronizing measurement for the eddy current sensor produces finer results.

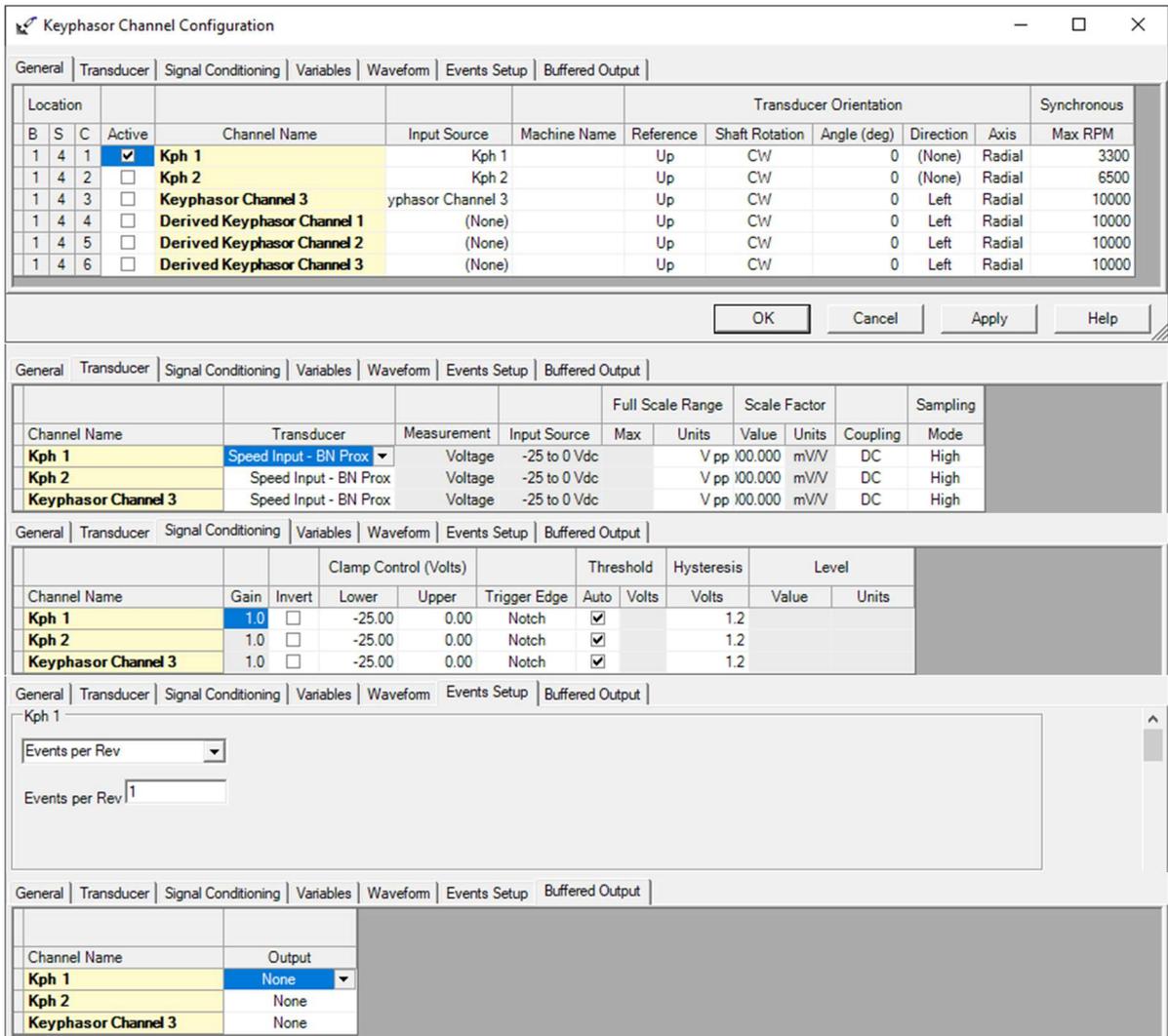


Figure 24. One-per-revolution reference mark configuration tabs.

In the "Signal Conditioning" tab, the minimum and maximum voltage configuration, the type of edge that will trigger the measurement, and the threshold hysteresis can be set.

Usually, large turbogenerators have only one notch on their circumference. That is why "Event per Revolution" is set up to one.

Table 2. Dynamic card configuration table.

Dynamic Channel Configuration			Notes	
General		Relative sensor name:	1-7 X or Y	depending on the sensor direction
		Machine Name:	HP, IP, LP or Generator	depending on the relative sensor placement
		Keyphasor:	Name fo theone-per-revolution mark	Mandatory for feature extraction Assigning the relative sensor to one-per-revolution-mark sensor (confirured separately)
	Transducer Orientation	Reference:	Up	
		Shaft Rotation	CW (Clockwize)	
		Angle (deg)	45	
		Direction	Right or Left	depending on the relative sensor placement
Transducer		Axis	Radial	Perpendicular to the shaftline axis
		Transducer	3300 8mm	Type of the relative sensor
	Full Scale Range	Measurement type	Displacement	Type of transducer measurement
		Max	500	Maximum expected sensor's value
	Scale Factor	Units	umpp	Unit of measurement (micrometers peak-to-peak)
		Value	7.87	Sensor sensivity
	Sampling	Coupling	AC	Sensor coupling
		Mode	High	128 or 256 samples per revolution depending on waveform configuration
	Bandwidth	Filter (rpm)	Wide (120)	Wide Bandwidth filter
		Filter poles	2	Numner of Poles for the bandpass filter
Variables	Bandpass	High	10	High bandpass filter frequency
		Low	1000	Low bandpass filter frequency
		Filter poles	2	Numner of Poles for the bandpass filter
	NX 1	Name	nX-1	
		Active	yes	Activates this signal feature
		Frequency	0.5	Feature frequency Coefficient times whirling seed
		Waveform	1# Sync Waveform	Waveform to obtain the signal feature from
	NX 2	Name	nX-2	
		Active	yes	Activates this signal feature
		Frequency	3	Feature frequency Coefficient times whirling seed
	NX 3	Waveform	1# Sync Waveform	Waveform to obtain the signal feature from
		Name	nX-3	
		Active	yes	Activates this signal feature
Frequency		4	Feature frequency Coefficient times whirling seed	
Waveforms	Waveform type	Synchronous		
		Active	Yes	
	Synchronous	Sample Rate	128	
		Keyphasor	One-per-revolution mark sensor synchronizing the Waveform	
	Time Synchronois Average	Enable	Yes	
		#Aver	4	Averaging 4 consecutive Waveforms
		#Revs	3	Number of revolutions to be averaged
	Spectrum	Spectral Lines	192	Number of spectral lines
	Waveform type	Asynchronous		
		Active	Yes	
	Asynchronous	Frequency Span	1000Hz	According to ISO standards
		Sample Rate	2560	Sampling rate for the asynchronous Waveform
		Spectrum	Spectral Lines	1600

The last tab is used when using other portable equipment. It is not used during this research. Other tabs are not significant for the type of research the dissertation considers. Therefore, the "Variables" and "Waveform" tabs are left in default modes.

Table 3. Configuration of one-per-revolution reference mark.

		One-per-revolution Marker Channel Configuration		Notes
General		Relative sensor name:	1-7 X or Y	depending on the sensor direction
		Machine Name:	HP, IP, LP or Generator	depending on the relative sensor placement
	Transducer Orientation	Keyphasor:	Name fo theone-per-revolution mark	Mandatory for feature extraction
		Reference:	Up	Assigning the relative sensor to one-per-revolution-mark sensor (confirured separately)
		Shaft Rotation	CW (Clockwise)	
		Angle (deg)	0	
	Synchronous	Direction	None	Default settings
		Axis	Radial	Perpendicular to the shaftline axis
Max RPM		3300	Maximum rotational speed expected during research	
Transducer		Transducer	Displacement sensor	Type of the relative sensor
		Measurement	Voltage	Type of measurement received from sensor
	Full Scale Range	Input Source	-25Vdc to 0Vdc	Negative polarity sensor
		Max	500	Maximum expected sensor's value
	Scale Factor	Units	Vpp	Unit of measurement Volts peak-to-peak)
		Value	1000 mV/V	Sensor sensitivity
	Sampling	Coupling	DC	Sensor coupling
		Mode	High	128 or 256 samples per revolution depending on waveform configuration
Signal Conditioning		Gain	1.0	Default settings
		Invert	No	Trigger acts at the falling edge of the notch
	Clamp Control (Volts)	Lower	-25 VDC	minimum value of sensor voltage - sensor passes on the notch
		Upper	0 VDC	maximum voltage of sensor voltage - sensor in contact with the metal
		Trigger Edge	Notch	
	Treshold	Auto	Yes	Automatic selection of thershold selection on the notch slope
		Volts		Default settings
	Hysteresis	Volts	1.2	Hysteresis for the theshold level

Thanks to the measurement equipment's configuration and setting summarized in Table 2 and Table 3, the author obtained the vibration features used in the following dissertation. Data collected with the portable equipment are a much richer source of information than only a single parameter used by monitoring systems in the control room. The power plant's monitoring and protection systems use an overall vibration amplitude. As a result, they miss most of the critical signal components used for diagnostic purposes. On contrary, portable systems can calculate much broader set of features. Typical list of these features is given below:

- Rotational speed of the shaft;
- Probe-to-shaft average distance in DC voltage (called Gap)
- Overall vibration amplitude (called Direct);
- The amplitude of the first harmonic of the signal (called 1X_Amplitude);
- The phase angle of the first harmonic (called 1X_Phase);
- The amplitude of the second harmonic of the signal (called 2X_Amplitude);
- The phase angle of the second harmonic (called 2X_Phase);
- The amplitude of the sub&super-harmonic of the signal (called nX_Amplitude);
- The phase angle of the sub&super-harmonic (called nX_Phase).

Figure 25 presents vibration features during measurements course in tabular view window. There is a brief summary of measurement configuration and status in the upper part of the tabular view. It describes channels, their name, location, the sensor's status, location and angular orientation, the timeframe of the measurement, the unit of speed, and units in which the amplitude and phase lag are shown in the table window below.

Ch#	Channel Name	Machine Name	Status	Angle	Direction	Run Type	Date	Speed Units (P)	Speed Units (S)	Amp Unit	Phase Unit
1	1V		OK	45°	Left		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
2	1X		OK	45°	Right		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
3	2V		OK	45°	Left		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
4	2X		OK	45°	Right		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
5	3V		OK	45°	Left		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
6	3X		OK	45°	Right		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
7	4V		OK	45°	Left		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
8	4X		OK	45°	Right		09mar2021 21:19:32.174 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
9	5V		OK	45°	Left		09mar2021 21:19:32.173 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
10	5X		OK	45°	Right		09mar2021 21:19:32.173 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
11	6V		OK	45°	Right		09mar2021 21:19:32.173 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
12	6V		OK	45°	Left		09mar2021 21:19:32.173 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
13	7V		OK	45°	Left		09mar2021 21:19:32.173 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg
14	7X		OK	45°	Right		09mar2021 21:19:32.173 To 03mar2022 19:06:50.873	rpm	rpm	um pp	deg

Ch#	C... Sample#	Date	Spe... Direct	Avg Gap	1XAmpli... 1X ...	2XAmpli... 2X ...	nX-1Amplitude	nX-1 Phase	nX-2Amplitude	nX-2 Phase	nX-3Amplitude	nX-3 Phase	Bandpass				
1	1V	1	09mar2021 21:19:32.174	3001	40.03	-7.191	29.10	358	6.560	92	1.686	266FNX	1.249	153BMA	1.438	255BMA	38.51
2	1X	1	09mar2021 21:19:32.174	3001	24.67	-10.088	17.17	159	3.936	266	0.774	74BMA	0.421	303BMA	0.712	39BMA	23.74
3	2V	1	09mar2021 21:19:32.174	3001	58.45	-8.624	25.62	192	10.28	124	3.176	310FNX	1.787	94FNX	1.582	333FNX	36.30
4	2X	1	09mar2021 21:19:32.174	3001	40.82	-10.173	27.43	295	13.12	328	1.508	101FNX	0.937	181BMA	1.473	135BMA	39.06
5	3V	1	09mar2021 21:19:32.174	3001	40.72	-7.165	34.21	248	7.365	78	0.018	270BMA	1.994	84FNX	0.437	139BMA	39.07
6	3X	1	09mar2021 21:19:32.174	3001	36.64	-10.312	29.11	341	6.256	190	0.253	134BMA	1.244	199BMA	0.446	148BMA	34.18
7	4V	1	09mar2021 21:19:32.174	3001	11.8	-6.000	97.0	302	23.21	353	2.514	163FNX	3.916	125FNX	3.377	203FNX	11.3
8	4X	1	09mar2021 21:19:32.174	3001	33.80	-9.953	24.00	22	9.971	119	0.526	185BMA	0.675	209BMA	0.018	270BMA	31.85
9	5V	1	09mar2021 21:19:32.173	3001	108	-5.929	95.3	90	22.45	351	1.654	340FNX	2.380	290FNX	1.229	317BMA	102
10	5X	1	09mar2021 21:19:32.173	3001	61.4	-10.075	58.5	132	15.23	101	0.963	308BMA	0.979	328BMA	1.197	308BMA	61.9
11	6V	1	09mar2021 21:19:32.173	3001	39.82	-9.783	29.64	75	4.104	345	0.679	220BMA	0.820	266BMA	0.786	224BMA	37.26
12	6V	1	09mar2021 21:19:32.173	3001	60.0	-5.629	3.997	0	48.11	9	2.106	50FNX	2.594	227FNX	1.942	77FNX	59.6
13	7V	1	09mar2021 21:19:32.173	3001	63.8	-6.378	55.6	244	15.73	137	0.877	215BMA	2.880	87FNX	1.407	235BMA	61.3
14	7X	1	09mar2021 21:19:32.173	3001	53.7	-9.931	36.19	347	20.63	309	0.735	162BMA	1.034	174BMA	0.640	159BMA	46.19

Figure 25. Portable equipment diagnostic features capabilities.

Finally, the lower part of the tubular window presents the essential features collected by the system and listed above. There is one more column called Bandpass. It presents the overall vibration amplitude within the bandpass filter range. This filter is set during configuration procedure depicted in Figure 23 in the third step (the Variable tab) and summarized in Table 2 in the Variables and Bandpass rows.

Figure 26 provides an overview of all the most important parameters needed to assess the technical condition of a machine during a transient state. Evaluation of such a complex object is not a trivial task. The analysis results are often displayed in the form of trends or graphics. They illustrate the behavior of the total vibrations, or they can be depicted as a graph of vibration vector composed of individual vibration features.

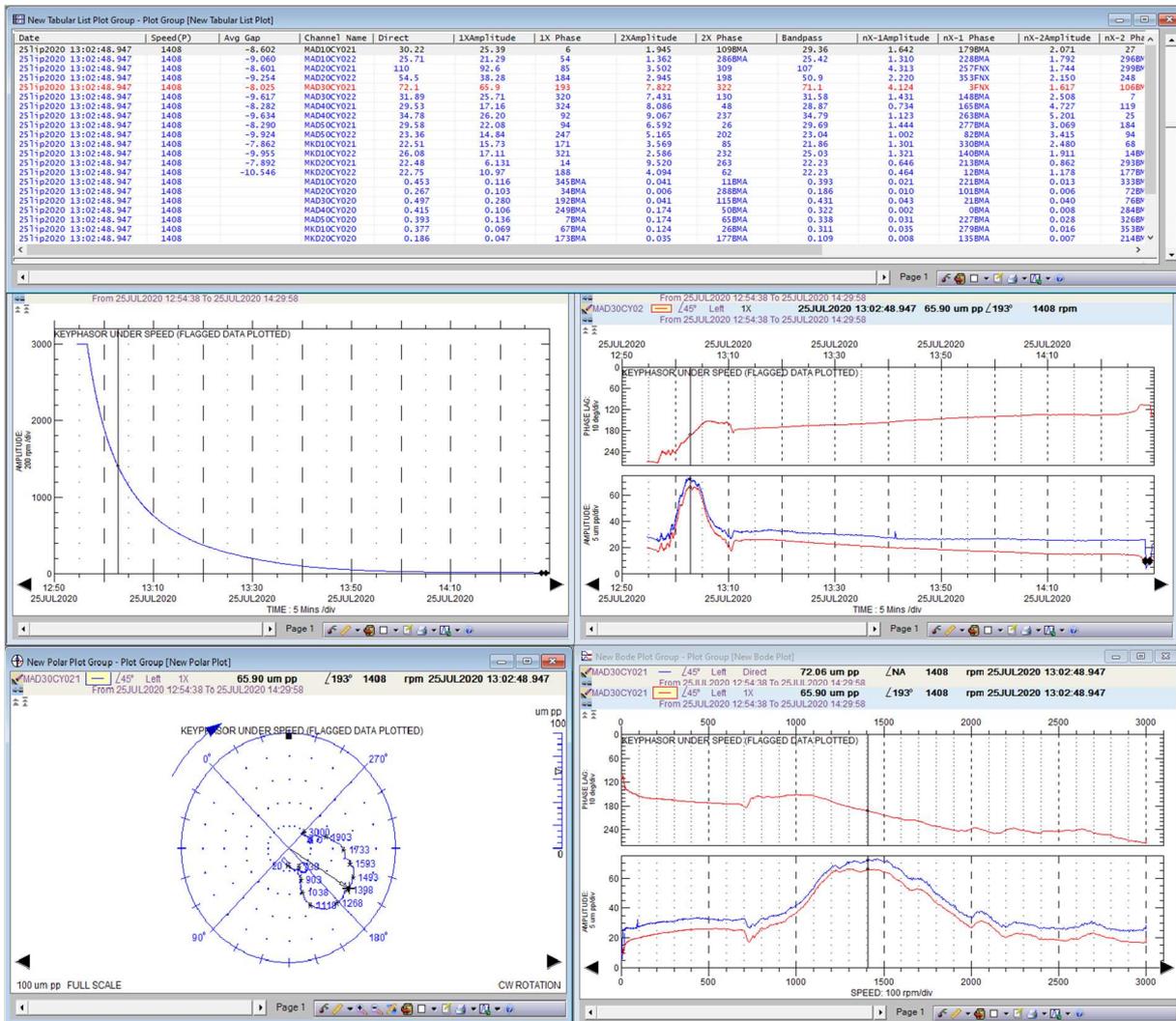


Figure 26. Coast down chart examples.

The top figure shows the tabular view with all available vibration features. The selected data point taken at speed of 1408RPM is synchronized across all the charts below. Red values in the tabular window describe the analyzed sensor. The coast down curve with the currently analyzed rotational speed instance is depicted in the middle-left figure. The right-hand side graph in the center shows the phase lag trend (upper, red chart) and both the overall and synchronous vibration trend (blue and red, respectively). The bottom charts are the Nyquist and Bode charts on the left and right. The Bode plot represents the amplitude and phase response of a system vibration signal separately during a transient state. The response is strictly related to rotational speed. The Nyquist plot shows the system response during a transient as a trajectory of the vibration vector in polar coordinates. The vector consists of the first harmonic amplitude and the corresponding phase lag. The set of these values for individual rotational speed increments is the trajectory of the vibration vector during coast down.

3.3. Transient data point sampling

Each monitoring and acquisition system records data with different resolutions in terms of time and rotational speed intervals. It is a result of the design of monitoring systems used in the field, and

numerous datasets have such a feature. This data were already taken in the past and it is not possible to repeat these measurements. Since the data is essential for the algorithm preparation, the data sampling must be tackled in the first place.

Portable measurement system configuration has two different triggering options, according to the change of speed and according to the elapsed time. Typical values are 20–60 s for time intervals and 5–50 RPM for the rotational speed change. Depending on the trip time instance, the measurement systems record the transient process at different points in time and speed. Therefore, direct comparison of the transient vibration parameters in an automated way is not possible and the data must be preprocessed to allow automated processing.

The following assumptions must be introduced in terms of sampling and triggering data to collect transient data properly:

- sample after each $\Delta_{rpm} = \pm 10rpm$ (since the trigger is activated by each change in the rotational speed of ± 10 revolutions per minute, the data will be stored in the acquisition unit. It does not matter when the trigger was activated along the rotational speed span regardless of the time elapsed between the samples),
- sample after each $\Delta_{time} = 20sec$ (after every 20 seconds, samples are to be stored regardless of the rotational speed change).

Due to the type of the tested object, which is a turbine set, the transition time is very long. The entire shaft line of the largest vessels can weigh over 500 tons, and their run-on time, as shown in Section 1.1, may take more than 120 minutes. Due to the shaft line inertia during a transient state (coast-downs especially) depicted in Figure 19, the data from 3,000 RPM to approximately 1,300 RPM will often be collected by Δ_{RPM} trigger, and from 1,300 RPM to the turning gear (which is activated at $\omega_{rpm} < 20$ RPM) the main trigger will be Δ_{time} . As a result, the data is not evenly distributed in terms of both time and RPM value. Several issues must be overcome to use the actual data from the measurement:

- The “raw” data points are unevenly spread (the reason is the configuration of the sampling which is explained in the previous paragraph),
- Due to the operational reasons, the start-point and end-point of the measurement are not always the same,
- Depending on the type of a transient (start-up or coast-down), the data is not shown in order (the rotational speed vector may start from the lowest rotational speed or the highest one).

Even if the nature of the transient is the same, the starting point of triggering each transient is not always the same, particularly for the rotational speed. The most important aspect is whether the starting point varies between particular measurements. Such complicated trigger procedure generates a different set of samples every time a transient is recorded. Data points are placed close to each other RPM-wise (comparing transient-to-transient), but not identically with respect to rotational speed mark.

Due to the fact that acquired data are field measurements and also that we measured many machines (of the same type, but still different units), the noise reduction was also an important matter. The data noise reduction considerations are presented in Section 3.5.1.

3.4. Structure of measurement data

The data collected by DAIU, which are needed for the analysis has a complex structure. It is similar to a three-dimensional matrix. Figure 27 presents the structure of the transient data matrix.

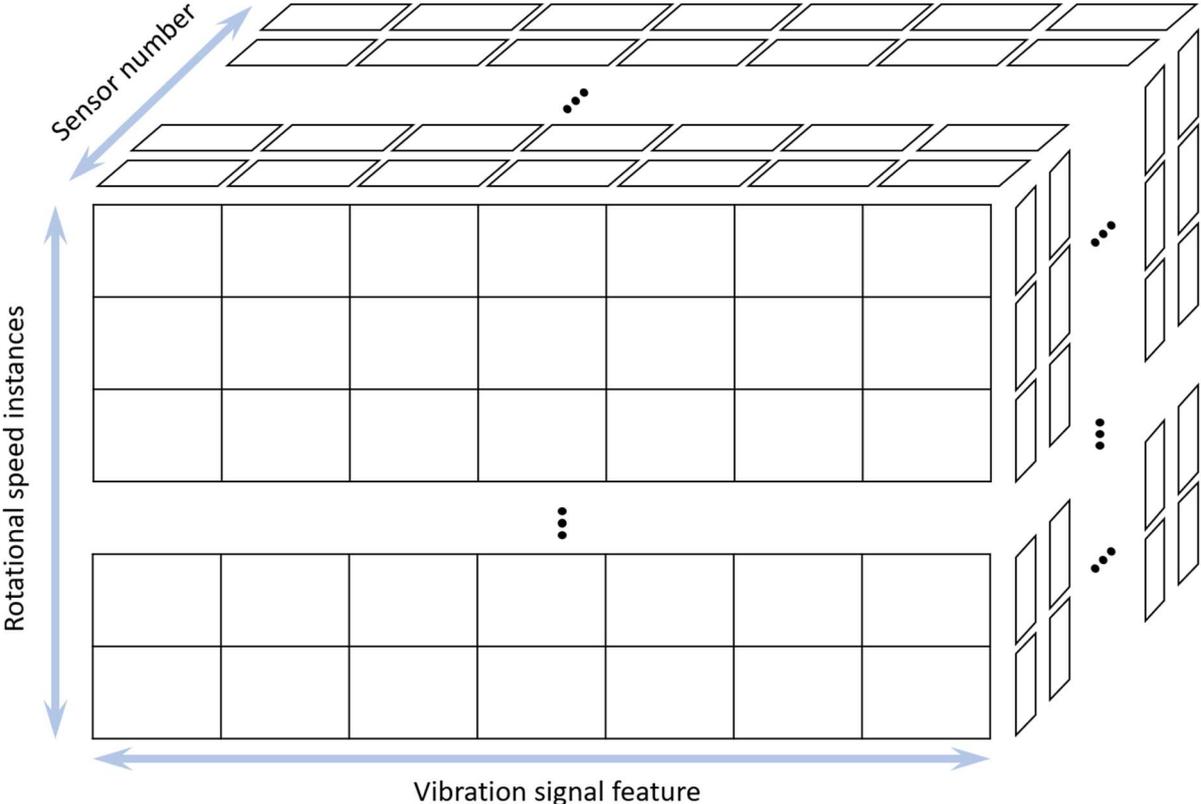


Figure 27. Transient data structure.

On the first axis (vertical), there are i -rows of subsequent rotational speeds for which data has been collected. There are j -columns with vibration signal features on the second axis (horizontal), and the third k axis (depth) contains the set of all sensors.

Each transient has 500 to 750 data points and up to 15 features per a single data point. Figure 28 shows a fragment of the data points matrix (rotational speed vs. signal features) from a single sensor during a transient state. This is the i and j axis of the transient data matrix. The third, k axis of the matrix includes all vibration sensors. In this example dataset there are 14 sensors.

Channel...	Date	Speed(P)	Direct	Avg Gap	1XAmplitude	1X Phase	2XAmplitude	2X Phase	nX-Amplitude	nX-1 Phase	nX-2Amplitude	nX-2 Phase	nX-3Amplitude	nX-3 Phase	nX-4Amplitude	nX-4 Phase
1Y	251p2020 16:53:37.423	3000	66.0	-8.611	63.2	40	1.088	166BMA	1.906	216BMA	2.095	14	2.063	313	2.035	204
1Y	251p2020 16:53:37.422	3000	67.1	-8.611	63.5	39	1.014	152BMA	1.816	203BMA	2.030	15	2.107	313	1.984	206BMA
1Y	251p2020 16:54:06.392	2997	66.9	-8.610	62.4	38	2.182	145	1.608	198BMA	2.074	14	2.031	312	2.011	206
1Y	251p2020 16:54:06.132	2992	66.6	-8.603	63.0	38	2.125	145	1.492	212BMA	2.103	13	2.062	314	2.046	207
1Y	251p2020 16:54:06.916	2987	67.3	-8.598	63.3	38	2.511	149	1.992	210BMA	2.120	12	2.080	314	2.089	205
1Y	251p2020 16:54:07.020	2987	67.5	-8.594	64.0	38	2.461	149	1.986	208BMA	2.105	12	2.072	313	2.058	205
1Y	251p2020 16:54:08.465	2977	67.6	-8.591	64.2	38	2.650	156	1.739	209BMA	2.105	12	2.054	314	2.095	204
1Y	251p2020 16:54:09.251	2972	69.1	-8.589	64.4	38	2.757	154	1.772	198BMA	2.076	12	2.045	314	2.043	203
1Y	251p2020 16:54:10.040	2967	68.0	-8.589	64.0	38	2.981	154	1.981	198BMA	2.093	12	2.036	315	1.993	204BMA
1Y	251p2020 16:54:10.828	2962	68.1	-8.589	65.1	37	2.502	151	1.772	209BMA	2.060	13	2.088	315	1.941	205BMA
1Y	251p2020 16:54:11.609	2957	69.5	-8.590	65.1	37	2.165	149	1.847	205BMA	2.063	14	2.015	315	2.021	205
1Y	251p2020 16:54:12.473	2952	69.3	-8.592	65.3	37	2.112	150	1.798	205BMA	2.080	14	2.131	315	2.048	203
1Y	251p2020 16:54:13.306	2947	69.5	-8.594	65.5	36	1.929	149BMA	1.529	218BMA	2.095	14	2.166	314	1.964	204BMA
1Y	251p2020 16:54:14.141	2942	69.4	-8.595	65.5	36	1.732	152BMA	1.929	218BMA	2.119	14	2.187	313	2.036	207
1Y	251p2020 16:54:14.999	2937	69.1	-8.597	65.7	36	1.574	158BMA	1.988	214BMA	2.122	14	2.140	314	2.084	204
1Y	251p2020 16:54:15.857	2932	69.6	-8.599	65.5	36	1.625	174BMA	1.879	213BMA	2.157	14	2.175	315	1.979	205BMA
1Y	251p2020 16:54:16.738	2927	68.8	-8.601	65.5	36	1.908	174BMA	1.765	208BMA	2.182	12	2.140	314	2.099	205
1Y	251p2020 16:54:17.421	2923	69.2	-8.602	65.6	36	1.888	182BMA	1.703	214BMA	2.216	12	2.175	314	1.995	206BMA
1Y	251p2020 16:54:18.300	2918	69.6	-8.603	65.8	36	1.989	184BMA	1.661	205BMA	2.189	12	2.166	315	1.954	207BMA
1Y	251p2020 16:54:19.182	2913	69.4	-8.604	65.7	35	2.128	180	2.006	207BMA	2.159	9	2.140	315	2.111	204
1Y	251p2020 16:54:20.064	2908	69.5	-8.606	65.4	35	2.475	180	1.671	208BMA	2.163	10	2.133	313	2.028	206
1Y	251p2020 16:54:20.937	2907	70.8	-8.608	66.2	35	2.580	187	2.003	210BMA	2.170	9	2.187	312	2.132	206
1Y	251p2020 16:54:21.827	2902	69.2	-8.610	65.9	35	2.687	183	1.633	208BMA	2.166	8	2.159	312	2.044	207
1Y	251p2020 16:54:22.719	2897	70.1	-8.612	66.2	35	2.625	188	1.900	212BMA	2.182	9	2.188	311	1.987	208BMA
1Y	251p2020 16:54:23.612	2892	70.1	-8.613	66.0	35	2.905	186	1.655	199BMA	2.187	7	2.198	311	2.104	206
1Y	251p2020 16:54:24.506	2887	69.9	-8.614	66.3	35	3.010	184	1.979	208BMA	2.150	8	2.186	312	2.110	207
1Y	251p2020 16:54:25.405	2882	70.7	-8.616	66.5	35	2.948	191	1.912	201BMA	2.154	8	2.152	312	2.051	204
1Y	251p2020 16:54:26.305	2877	69.2	-8.618	66.3	35	3.263	188	1.950	201BMA	2.147	7	2.126	312	2.037	203
1Y	251p2020 16:54:27.206	2872	70.4	-8.619	66.8	35	3.186	189	1.877	205BMA	2.150	7	2.150	313	2.110	207
1Y	251p2020 16:54:28.109	2867	70.1	-8.620	66.8	35	3.597	186	1.728	209BMA	2.131	6	2.150	313	2.084	204
1Y	251p2020 16:54:29.033	2862	70.0	-8.621	66.8	35	3.617	182	1.679	215BMA	2.108	6	2.185	313	2.071	206
1Y	251p2020 16:54:30.040	2857	71.5	-8.623	67.2	35	3.579	185	2.221	204FNX	2.102	8	2.229	312	2.038	206
1Y	251p2020 16:54:31.049	2852	70.4	-8.624	67.0	35	3.644	183	2.015	198FNX	2.147	7	2.211	313	1.995	206BMA

•••

1Y	251p2020 17:39:48.746	142	31.30	-8.838	24.23	325	4.534	74	9.110	118FNX	3.105	29	6.377	226	5.540	92
1Y	251p2020 17:39:57.269	141	31.48	-8.828	24.75	325	4.554	79	9.528	115FNX	3.079	30	6.253	229	5.320	72
1Y	251p2020 17:40:17.768	137	30.98	-8.819	24.61	324	4.951	80	9.230	114FNX	3.013	30	5.910	226	6.550	49
1Y	251p2020 17:40:19.306	137	31.49	-8.819	25.05	325	4.925	79	9.536	114FNX	2.964	27	6.117	225	6.676	53
1Y	251p2020 17:40:37.267	134	30.97	-8.820	24.56	324	4.849	79	9.332	114FNX	3.049	30	5.460	223	6.276	47
1Y	251p2020 17:40:54.002	132	30.93	-8.822	24.01	324	4.969	80	9.825	113FNX	2.833	28	5.988	224	6.483	32
1Y	251p2020 17:40:57.266	131	30.01	-8.823	24.16	324	5.088	80	9.203	113FNX	2.828	28	5.210	227	6.110	27
1Y	251p2020 17:41:15.265	128	30.57	-8.825	23.77	323	4.990	77	9.117	112BMA	2.807	26	5.847	226	6.027	38
1Y	251p2020 17:41:23.453	127	30.64	-8.826	23.64	323	4.835	77	9.305	111FNX	2.834	25	4.726	223	6.385	19
1Y	251p2020 17:41:37.263	125	30.96	-8.827	23.30	322	4.888	76	9.198	110FNX	2.959	26	4.541	225	6.149	11
1Y	251p2020 17:41:57.262	122	31.13	-8.828	22.84	322	4.899	77	8.972	109FNX	2.927	28	4.108	223	5.958	6
1Y	251p2020 17:41:57.691	122	31.13	-8.828	23.05	322	4.892	77	9.947	109FNX	2.712	26	4.324	226	5.403	7
1Y	251p2020 17:42:17.681	119	31.74	-8.828	22.75	320	4.949	77	8.933	108FNX	2.729	27	4.027	226	5.368	5
1Y	251p2020 17:42:33.866	117	31.21	-8.829	22.68	321	4.868	76	8.805	107FNX	2.664	22	4.952	226	4.952	352
1Y	251p2020 17:42:37.260	116	31.44	-8.828	22.58	321	4.906	76	8.900	105FNX	2.685	27	3.840	227	5.205	352
1Y	251p2020 17:42:55.262	112	31.45	-8.831	22.54	320	4.758	74	8.772	104FNX	2.704	28	3.549	228	5.058	348
1Y	251p2020 17:43:09.572	112	31.42	-8.832	21.95	320	4.817	76	8.668	103FNX	2.501	22	3.804	226	4.793	341
1Y	251p2020 17:43:17.258	111	31.41	-8.833	21.93	319	4.857	73	8.601	103FNX	2.597	26	3.427	226	4.102	336
1Y	251p2020 17:43:37.257	108	29.69	-8.835	21.47	318	4.425	74	8.682	100FNX	2.623	30	3.431	225	4.466	336
1Y	251p2020 17:43:47.947	107	29.46	-8.837	21.40	318	4.614	75	8.464	99FNX	2.587	25	3.550	232	4.056	330
1Y	251p2020 17:43:57.256	106	29.28	-8.837	21.84	318	4.639	75	8.382	100FNX	2.527	28	3.150	232	3.908	332
1Y	251p2020 17:44:17.255	103	29.00	-8.838	20.95	317	4.835	74	8.320	97FNX	2.499	26	3.270	233	3.644	328
1Y	251p2020 17:44:27.589	102	28.15	-8.840	20.73	318	4.868	77	8.046	96FNX	2.464	26	3.426	233	3.941	325
1Y	251p2020 17:44:37.253	101	28.03	-8.841	20.60	317	4.638	72	8.243	97FNX	2.374	25	3.203	233	3.596	319
1Y	251p2020 17:44:57.252	98	27.95	-8.843	20.08	317	4.740	73	8.023	98FNX	2.389	26	3.159	233	3.588	322
1Y	251p2020 17:45:09.239	97	27.44	-8.846	19.77	316	4.925	72	7.515	95FNX	2.455	26	3.100	233	3.514	320
1Y	251p2020 17:45:17.251	96	27.18	-8.847	19.49	316	4.750	74	7.542	94FNX	2.284	25	2.972	231	3.466	316
1Y	251p2020 17:45:37.250	94	26.79	-8.847	19.23	315	4.963	70	7.623	93FNX	2.412	29	2.923	233	3.162	316
1Y	251p2020 17:45:52.445	92	25.94	-8.848	19.16	315	4.661	72	7.683	91FNX	2.236	27	2.931	233	3.203	308

Figure 28. Transient data points for one sensor.

Figure 29 shows the j and k axes of the transient matrix at one point of the rotational speed instance. The rotational speed column called “Speed(P)”, depicted on the figure shows fixed value but sensor column called “Channel” presents all sensors during the measurement course. On average, one transient matrix

mark as the reference signal enables us to compare data repetitively and reliably. Data prior, during, and past critical speed can be unequivocally identified and compared. Resonance peaks and shaftline unbalance response depend heavily on rotational speed and should be appropriately compared independently on a trigger setting.

Gathering a sufficient amount of data during transient states requires using two kinds of triggers: one is rotational speed-dependent, and the other one is time-dependent. Both triggers are independent from each other. Figure 19 presents the inertia of the system during a typical coast down. It depicts revolutions per minute versus time elapsed. Typical coast down of the large turbomachinery can last more than an hour. Rotational speed starting from the trip point (at the full rotational speed of the machine) down to approx. 1/6 of nominal speed (500 revolutions per minute) changes quickly compared to the time elapsed. The RPM-dependent trigger is needed during this first stage of a coast down. During the second stage, rotational speed changes are much slower. At that stage, the time-dependent trigger will provide more samples (i.e., information). Such a trigger procedure (widely used in engineering practice) generates a different set of the database every time a transient occurs. Data points are placed close to each other (comparing transient-to-transient) but not identically concerning the rotational speed mark.

Unequally spaced data points along the rotational speed axis can introduce difficulties in implementing processing algorithms. Figure 30 presents an example of data shown for one sensor with ten transient states. Each transient has a different scatter color. The data from the transients have different rotational speed values, making the comparison of signal amplitude values inaccurate and cumbersome, and direct automatic evaluation of such transients is impossible. To tackle this issue, a preprocessing method is required. Such a method is part of this dissertation.

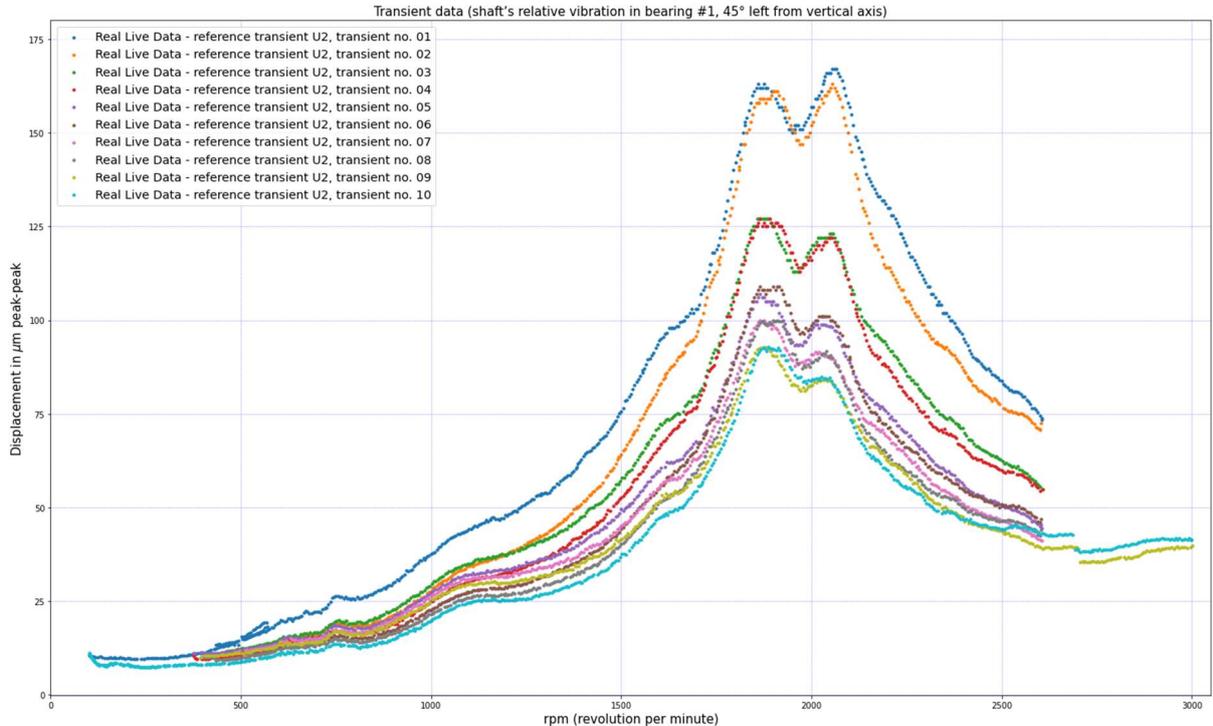


Figure 30. Non-equidistant spaced transient data points.

Measurements of transients in large machines are not frequent events. Transient states cannot be repeated at will, due to cost-related, time, and stress for the machine. Therefore, interpolation is needed to find function values at places not specified in the transient data point matrix.

To start with, the algorithm which converts the measured data to the RPM-equidistant data is necessary. At first, a function creates an equally spaced vector of rotational speeds and each particular transient. In other words, the first step is resampling the speed vector to generate a data vector having the same RPM values for all the transients.

Figure 31 presents set of “raw transients” data points of one signal feature with the data obtained throughout one of the field measurement courses performed by the author. Upper plot shows the whole four transients recorded during one of the measurement course. The bottom figure presents the zoom for an example part of the rotational speed interval of [1820, 1900] RPM. The zoom reveals that not all the data points are collected for the same RPM values. It creates a significant problem in analyzing and comparing transients data with respect to rotational speed increments. In order to correctly compare values between transients, the values for individual transients must be interpolated.

There are various interpolation methods. The trivial, linear interpolation introduces significant errors. Trigonometric interpolation is used to approximate periodic functions. However, the transient curve is not a periodic function, so the interpolation task cannot use this method. Polynomial interpolation such as first and higher-order is used to find function values at points beyond the points collected during measurements. However, the polynomial of interpolated a function has a degree one lower than the number of points in the data set. For a single signal feature during a transient, the collection of points averages 500-700 data points. The extremely high order of the polynomial makes this approach impractical for this application. An interesting solution is spline interpolation, especially its variation - a third-degree polynomial called cubic spline (CS). This interpolation fits the third degree (cubic) polynomials between each of the two consecutive data points in data set. This process produces third degree polynomials of one less than the data points in the data set. Each of the spline function from this set has a first and the second derivative specified fit the piecewise function without breakage of its continuity. The third degree polynomial assures that the line connecting all the data points will be continuous and will have smooth non-erratic shape.

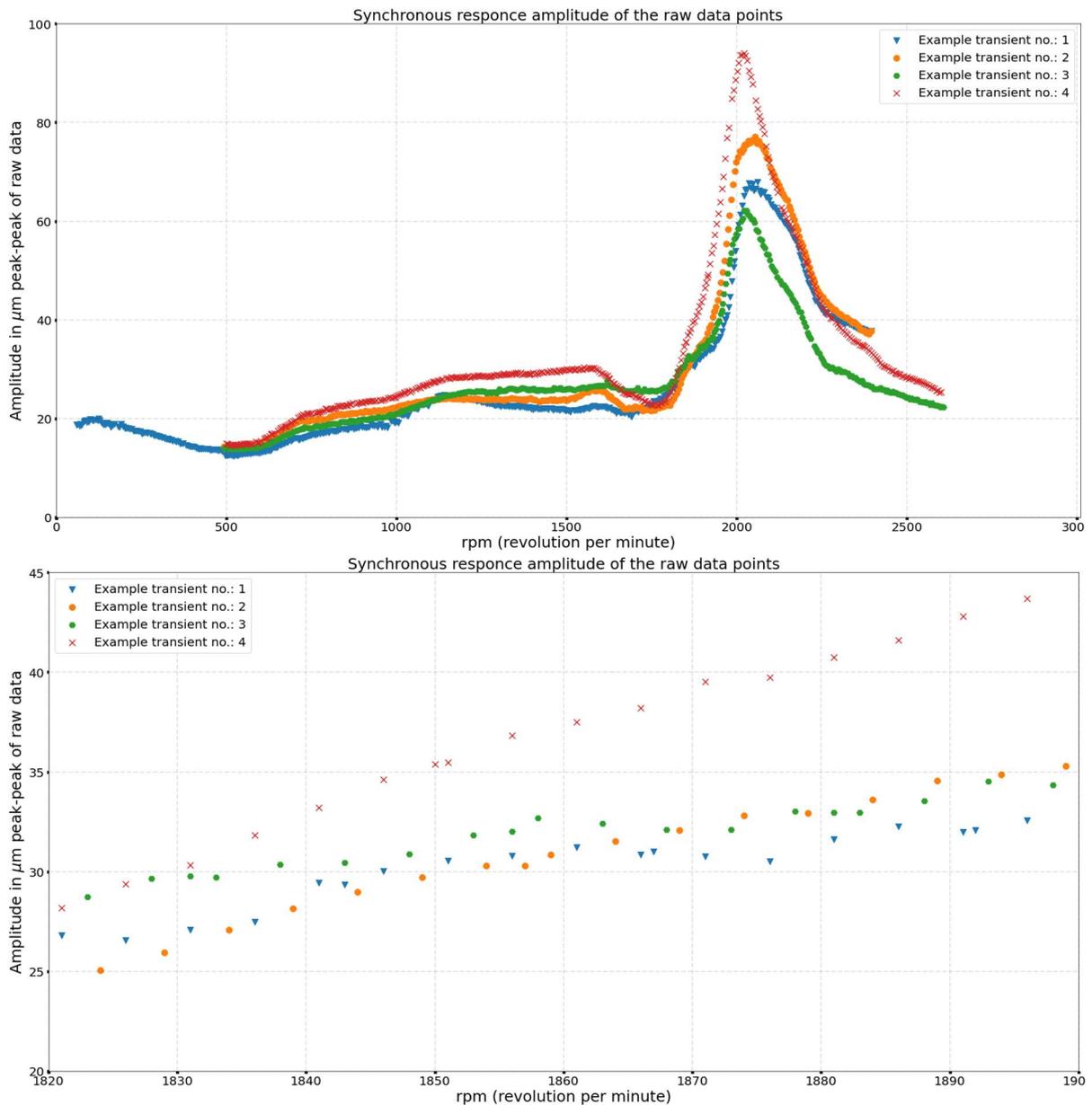


Figure 31. Raw transient data example: top – whole transient data point set; bottom – zoom of the transient representative interval.

Thus, the author decided that the most suitable method for this application will be cubic spline interpolation. CS interpolate the transient data points and helps to generate the data points for the same rotational speed points for each analyzed transient. Additionally, it also handles “cropped” transients, i.e., transients which do not start at “0” RPM and finish at the FSNL (Full Speed No Load) point, i.e. at 3000 RPM. Finally, the CS interpolation applied to any new analyzed transient ensures that all the data points are located in the same places on the rotational axis as from the baseline transient. This operation enables data to be compared reliably. Schumaker [24] and Dyer [25] presents the advantages of equally spaced points/knots in polynomial spline functions.

3.5.1. Cubic spline

The Cubic Spline interpolation is a type of interpolation that handles the problem of oscillating edges of intervals with equally spaced interpolation points when using higher-order polynomial interpolation. Gerald and Wheatley described theoretical considerations and applications in [26].

Schumaker in [24] formulated a set of four general properties for a centerline of the cubic spline function s in the Cartesian plane for a set of points $(x_i, y_i), i = 1, 2, \dots, k$:

- 1 s is a piecewise cubic polynomial with knots at x_1, \dots, x_k ;
- 2 s is a linear polynomial for $x \leq x_1$ and $x \geq x_k$;
- 3 s has two continuous derivatives everywhere;
- 4 $s(x_i) = y_i, i = 1, 2, \dots, k$

Such a function produces more minor errors and improves accuracy. Schumaker in [24] and Gerald and Wheatley in [26] describes the theory, together with a process of creating and using spline interpolation. Finally, [25] shows a few examples of spline interpolation as a curve fitting method. The main idea of cubic spline is presented by Schumaker in [24]. The goal is to produce a set of the third-degree polynomial functions $s_i(x)$ that satisfy:

$$S(x) = \begin{cases} s_1(x), & \text{if } x_1 \leq x < x_2 \\ s_2(x), & \text{if } x_2 \leq x < x_3 \\ \vdots & \\ \vdots & \\ s_{n-1}(x), & \text{if } x_{n-1} \leq x < x_n \end{cases} \quad (8)$$

Where polynomial to be fitted across each interval $x_i \leq x < x_{i+1}$, is given by equation:

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (9)$$

where $i = 1, 2, \dots, n - 1$, and respectively, the first and the second derivative is given by:

$$s'_i(x) = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i \quad (10)$$

$$s''_i(x) = 6a_i(x - x_i) + 2b_i \quad (11)$$

for the same $i = 1, 2, \dots, n - 1$.

The matrix equation for the cubic spline interpolation is given by:

$$\begin{bmatrix} 1 & 4 & 1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 & & 0 & 0 & 0 & 0 \\ & & \vdots & & \ddots & & \vdots & & \\ 0 & 0 & 0 & 0 & & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \\ \vdots \\ M_{n-3} \\ M_{n-2} \\ M_{n-1} \\ M_n \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-4} - 2y_{n-3} + y_{n-2} \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix} \quad (12)$$

where: $M_i = s''(x_i)$, and $h = x_i - x_{i-1}$.

This is an under-determined system ($n - 2$ rows by n columns). To find unique solutions for the matrix equation (5) the following assumptions has to be made:

$$M_1 = 2M_2 - M_3 \quad (13)$$

$$M_N = 2M_{N-1} - M_{N-2} \quad (14)$$

This boundary conditions let us reduce the system matrix to a $n - 2$ by $n - 2$ dimensions:

$$\begin{bmatrix} 6 & 0 & 0 & & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \\ 0 & 0 & 0 & & 4 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ M_4 \\ \vdots \\ M_{n-3} \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-4} - 2y_{n-3} + y_{n-2} \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix} \quad (15)$$

Solving (6) yields the sought equally distanced interpolated data points.

3.5.2. Cubic Spline spacing

When applying the interpolation of the data, it is important to properly select the data spacing parameter. During research work for the dissertation, the author tested different settings of the spacing parameter, i.e., spacing between consecutive rotational speed values, for the CS interpolation. He performed these tests with 25, 25, 100, and 150 RPM between data points and applied it to the OpEn procedure. Figure 32 shows the OpEn method for each value of this parameter.

The spacing parameter value of 25, Figure 32 (a), gives outstanding results in replicating the transient function shape. Unfortunately, it consumes a lot of time and computing power. As shown in Figure 1a, for a rotational speed of about 2600RPM OpEn center line and acceptance region recreate the disturbance in the form of a sudden decrease in the amplitude value. This drop indicates that using this parameter value will be potentially sensitive to data noise.

Figure 32 (c) and Figure 32 (d), respectively shows the OpEn method output for spacing parameters 100 and 150. It takes much less time to complete the entire procedure than for 25 RPM parameter. It is less sensitive to data interference, which is shown in Figure 3. Unfortunately, the graphs also show that the method for these parameters is not able to correctly replicate the shape of the transient between 1750RPM and 2250RPM. This range is the most important part of the transient, which disqualifies the 100 and 150 RPM parameters for the OpEn method .

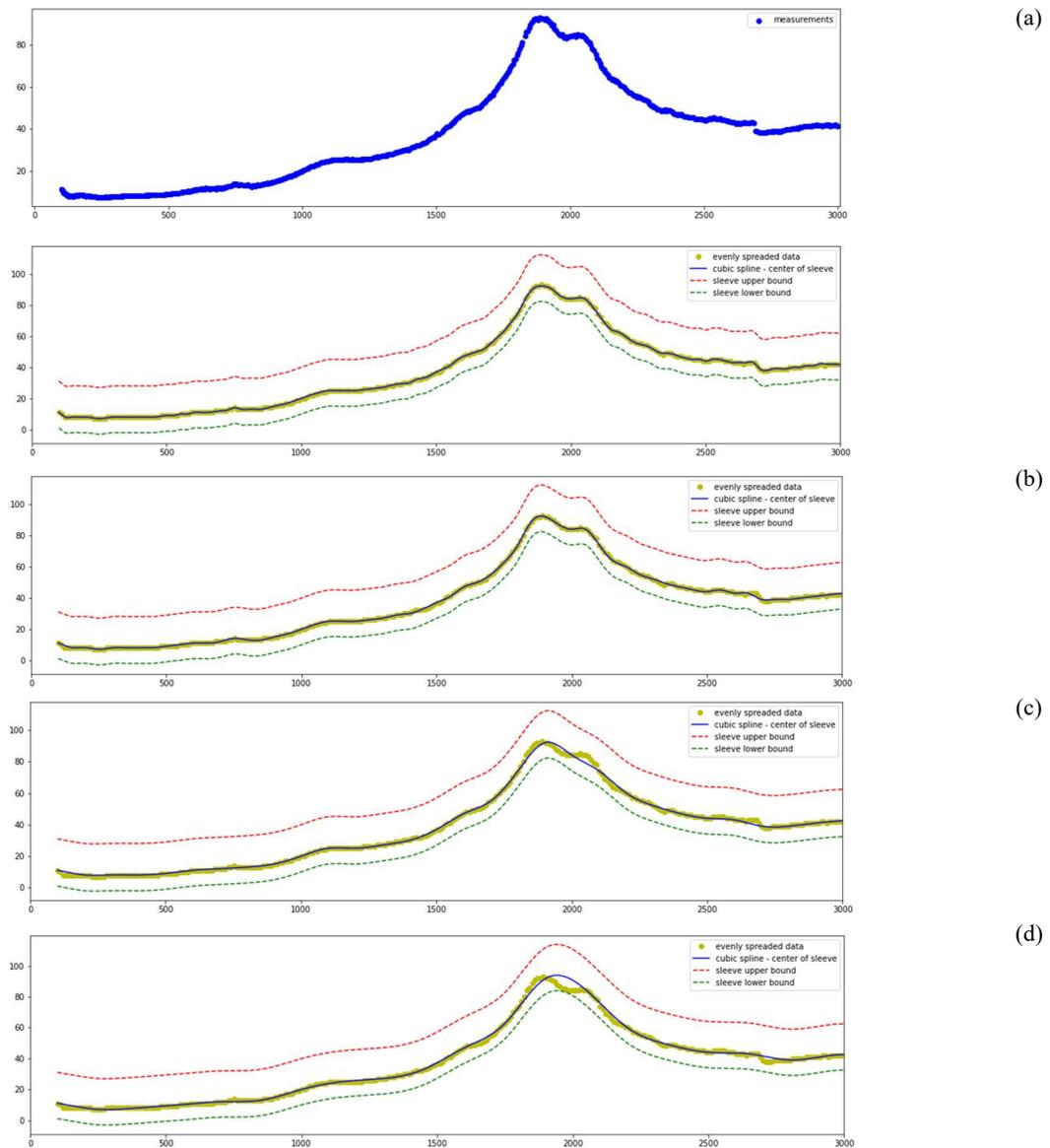


Figure 32. Examples of different settings of the spacing parameter interval between the consecutive rotational speed instances: (a) spacing: 25RPM; (b) spacing: 50RPM; (c) spacing: 100RPM; (d) spacing: 150RPM.

The best ratio of the data quality, i.e., reconstruction of the shape of the transient function, to the time and computing power has a value of 50RPM between samples. Setting the spacing parameter to 50 RPM allowed to reduce the susceptibility to the data noise with the correct reproduction of the shape of the transient curve in the entire rotational speed range.

4. Fault detection method

After the data are acquired and preprocessed, as proposed in the previous section, the algorithms for fault detection and identification can be executed. Fault detection is the first step of the proposed set of algorithms. Fault detection is checked as the first one and should check whether there is any change in the data. The ultimate application of the method is to check every transient state that the machine experienced. After implementation in an online system, the system would run in the background by default in parallel with the monitoring system. After each transient is stored in memory, the fault detection algorithm would be executed and screen the data whether a human expert should investigate the data.

4.1. Anomaly detection method

The automated assessment of complex technical systems was the subject of numerous research. Demetgul et al. in [27], highlight the fact that most industrial systems are non-linear and require appropriate analysis methods. Each such an attempt must include feature extractor and classifier. The authors have analyzed multiple generic methods for the diagnostic of the pneumatic systems of the material handling systems, starting from dimension reduction to clustering for classification.

In recent years, novelty detection algorithms for small machines equipped with rolling element bearing have gained more and more popularity. Dworakowski et al. in [28] test novelty detection algorithms. The principles of their operation are based on Artificial Neural Networks (ANN), feature space distance, and probability distribution. Their study included a healthy state of operation, different damage scenarios, and various feature vectors for elliptical gearboxes' autonomous and multidimensional monitoring. Wang et al. [29] sophisticated fault diagnostic scheme for planetary gearboxes. They use three vibration signal domains, i.e., frequency, time, and time-frequency, to extract fault features. A new method, incorporating a multi-objective evolutionary and decomposition algorithm, is proposed to enhance fault feature subsets. In addition, Dezert-Smarandache rules are applied to improve robustness and fault classification rate. Lis et al. [30] proposed an interesting approach to the novelty detection method. Their method used data over-hung centrifugal pump vibration data in the time domain. The data for the analysis is based on accelerometers. They introduced pump diagnostics based on the nearest-neighbor method reinforced by the new data preprocessing method. The proposed method has been validated on on-site pumps used in the industry. They argue that data-driven methods can be used in predictive maintenance strategies. The above methods and algorithms are very effective for small rotating machines equipped with rolling bearings. Unfortunately, large turbine sets have slide bearings, in which the wedge film suppresses a significant part of the vibration signal. As a result, it creates a barrier to high-frequency vibrations propagation, used for the analyses presented above. This makes the methods mentioned above ineffective.

Zhang et al. in [31], [32] presented an interesting approach to "the next level" of data-driven machinery diagnostics. He proposed a method that joins the domain gap across varying operating conditions. Although, his work implies to be effective in applications for the rolling element bearing in the train

industry, it can produce proper cross-domain fault diagnosis only with a balanced amount of different fault modes data available.

The author also analyzed the literature on rotating machines equipped with slide bearings. New condition monitoring strategies for these machines were well described by Jabłoński and Barszcz in [33] and Capelli et al. [34]. This new approach requires new ways of monitoring the turbines. Banaszekiewicz in [35] showed a concept lifetime assessment system for steam turbines that considers a wide range of operating condition changes in the scope of creep-fatigue damage. Zagorowska et al. [36] presented an interesting concept of exponential trend approximation with shape adaptation to monitor performance degradation during operation. Hanachi et al. [37] presented an interesting approach to improving prognostic accuracy in the compressor section of gas turbines by taking into account the effects of humidity, and Zohair et al. [38] proposed a modified Weibull distribution as a reliability estimator for gas compression turbines to reduce the failure risk. These works are of great value and present an improved way of monitoring the rotating machines; however, they are missing much information on machine dynamic conditions. This information comes from a transient state of the turbogenerator. New diagnostic technologies are being developed, e.g. state-of-the-art thermal and flow diagnostic of steam turbines (described by Głuch in chapter 3 of [39]) and introduced to power plants, still vibration response of the unit remains the fundamental method to assess the technical state of turbo-sets. In [40] authors perform analysis of a complex case of Gas Turbine vibrations. They confirm the fact that tedious analysis work and availability of experts is required for proper detection and identification of a large turbomachinery fault. These works are of great value and present an improved way of monitoring the rotating machines; however, they are missing a significant portion of the information on machine dynamic conditions. This information comes from a transient state of the turbogenerator.

The authors in [41] showed that such a method will be beneficial and can be beneficial in two ways: as a “health monitoring” parameter for the maintenance personnel and the planner and management personnel – to properly plan and execute machine inspections and overhauls. Bornassi et al. [42] highlight the importance of analysis of transients states in the case of large turbomachinery blades. The authors of the paper presented a combination of the 1DOF model with real blade vibration measurement data to identify the vibration parameters of blades during transient state.

The review of the state-of-the-art showed that there is a lack of the method to help the maintenance personnel quickly assess the state of machinery during turbine coast-downs and start-ups, ideally in an automated way. Therefore, creating a method to define a healthy pattern, i.e., reference or baseline) is of great value. Having such a pattern, together with some acceptance boundaries, can compare each transient to whether it represents a healthy condition. Due to the lack of skilled personnel, it should be reiterated that the method should be automated. The author proposed such a method and coined the name of Operational Envelope (OpEn). This idea is based on an envelope wrapped around a particular signal feature for the 2D case or features for the 3D case during the transient state of the machine.

The author proposed a method to detect an anomaly in transient behavior called Operational Envelope (OpEn). The method, depicted in Figure 33, consists of the following steps. First, the transient state

data are collected from the turbogenerator, i.e., real-object. Then the baseline transient is determined. Next, the CS interpolation is applied to these reference transients. Afterward, the baseline selection is established. The baseline determination procedure is described in Section 4.4 and 4.7. Next, the CS interpolation distributes the data points equally across the rotational speed range. This operation establishes the center points of the OpEn. In the following stage, the upper and lower values for OpEn are defined.

The proposed name of the "operational envelope", comes from the meaning of the actions involved. However, this term should not be mistakenly mixed with the concept of the "signal envelope," and its spectrum called the "Envelope Spectrum."

Table 4. Differences between OpEn and standard Spectrum Envelope.

	Operational Envelope	Envelope Spectrum
Function domain	RPM/CPM (revolutions/cycles per minute)	Hz
Rotational speed	Varying across large span	Constant
Number of amplitudes	1 st harmonic across whole RPM range (system's response to the centrifugal force)	N spectral lines (each refers to different frequency/amplitude) It contains sub-harmonics, harmonic and multiple of harmonic, and all between (depending on spectral resolution)
Tacho-sensor	Essential	Unnecessary
Attitude/lag angle	Used	n/a
Envelope's setpoints	Center of envelope ± arbitrary value(s)	n/a

These are two different methods, and there are several significant differences between the spectrum envelope and the OpEn proposed by the author. Table 4 summarizes the main differences between these two concepts.

As one can see, both methods are very different. OpEn uses a speed sensor to track changing RPM during transients. On the other hand, Envelope Spectrum assumes that the machine's revolutions are stable, i.e., the method is used for evaluating the machine's condition in the steady-state operation.

4.2. OpEn method

The first step in an automated anomaly detection method during transients of large turbomachinery is to collect data to create a baseline transient. For this purpose, the author analyzed over 150 databases with records recorded during 13 years of professional work. Next, from among evaluated transients, the author nominated approximately 25 to define the baseline transient. After that, the transient data must be CS interpolated to obtain strictly defined points in the field of rotational speed. Without interpolating the functions at predetermined points, it would not be possible to reliably compare transients due to the hurdles described in Section 3.5.

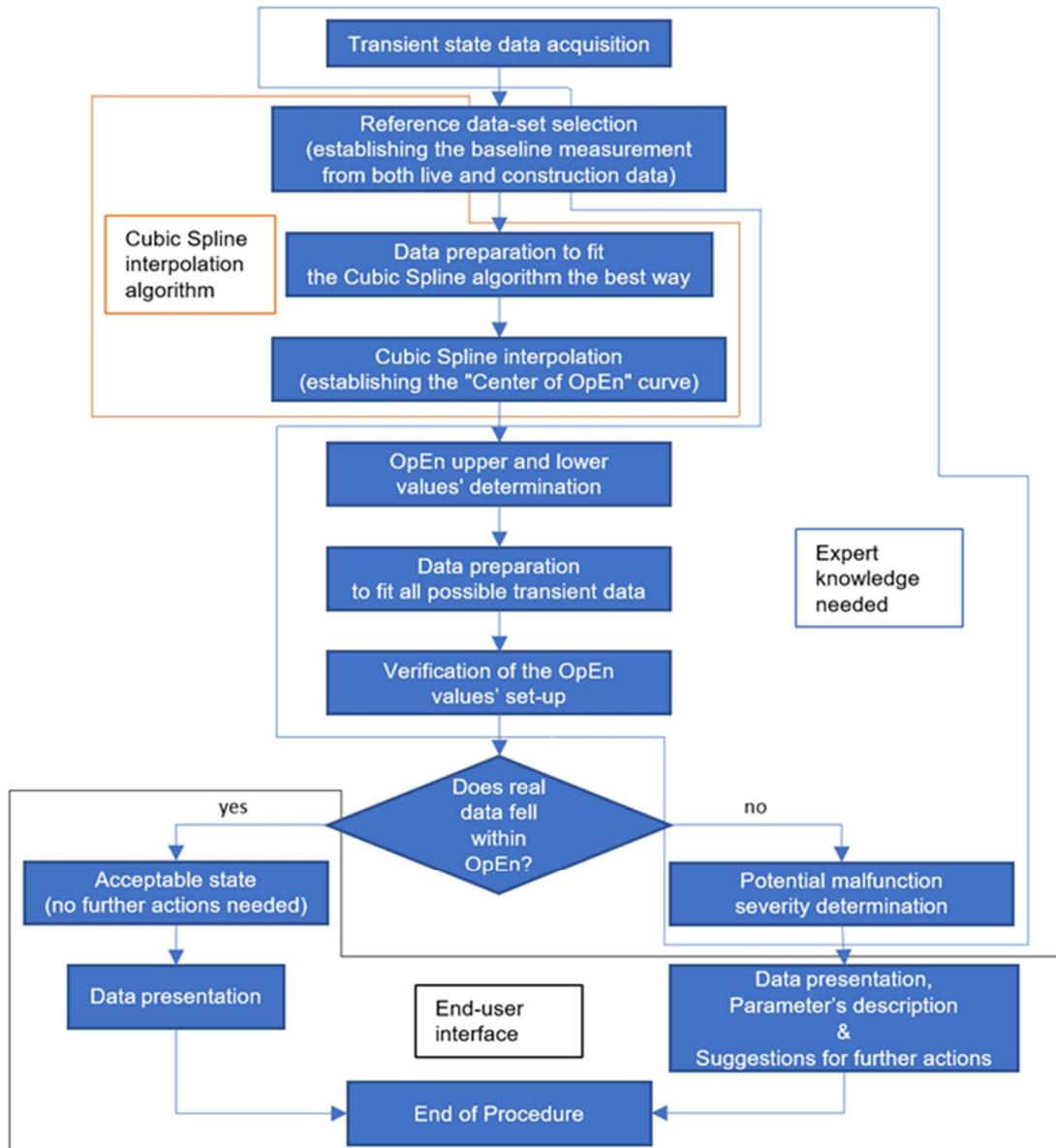


Figure 33. Automatic anomaly detection flowchart method.

This Section describes the Operational Envelope (OpEn) method in detail. The content of this Section was published by the author in paper [13]. The following description is only valid for one sensor. For the entire turbine set to be included in anomaly detection, each analyzed sensor must undergo this procedure. For example, this procedure will be performed six times for a turbine set equipped with three bearings with two eddy-current sensors for each bearing. On the other hand, for a nine-bearing turbine set with two sensors per bearing, 18 iterations of this procedure are required.

The method consists of several steps. The author collected real data from the transient state of the turbogenerator set. Then he used the acquired data and data from the turbine's design, i.e., engineering department where he works, to determine the baseline transient which suits the design best. Afterwards, the baseline transient is subjected to the CS interpolation to have equally distributed data points across the rotational speed range. Next, the center of OpEn and its upper and lower values are established.

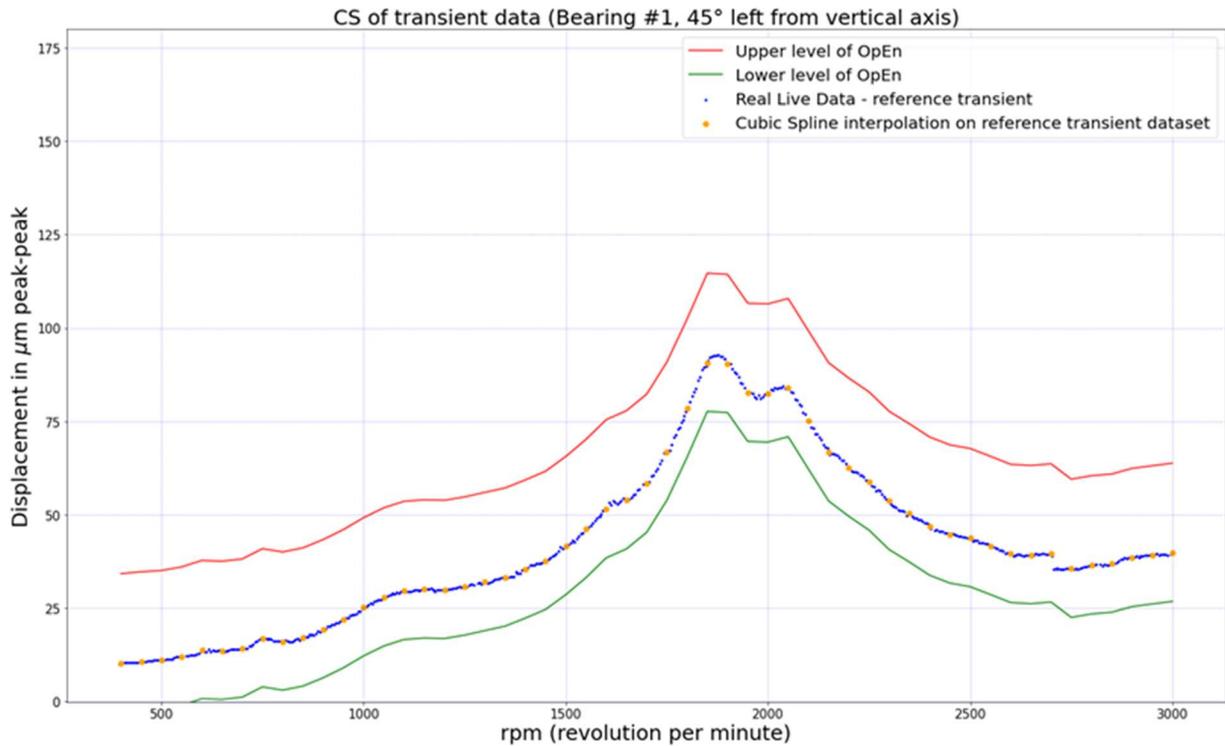


Figure 34. OpEn and its upper and lower values.

The upper and lower values of OpEn for individual rotational speed values constitute the acceptance region for that specific rotational speed. This is represented by red and green line values in the points of CS (orange scatter points) on Figure 34. A set of values for every rotational speed in which the turboset transient states are defined creates a 2D or 3D OpEn for a particular sensor. Now, for every rotational speed instance defined by CS interpolation, a new data can be quickly verified if it is inside the OpEr region. No further improvement actions are required if the OpEn contains all the data from a new transient state; however, if the data or even a few points from the transient lay outside the OpEn, further actions should be suggested to assess the severity of the malfunction.

After that, the amplitudes of reference transients are summed up for the consecutive rotational speed increment values (from RPM = 100 to RPM = 3000 every 50RPM). This operation gives 60 places where the algebraic mean of all amplitudes at each point is taken. Figure 35 shows an example of the reference data (for clarity, only five transients were used). Points for baseline transient are summed at specific places on the x-axis.

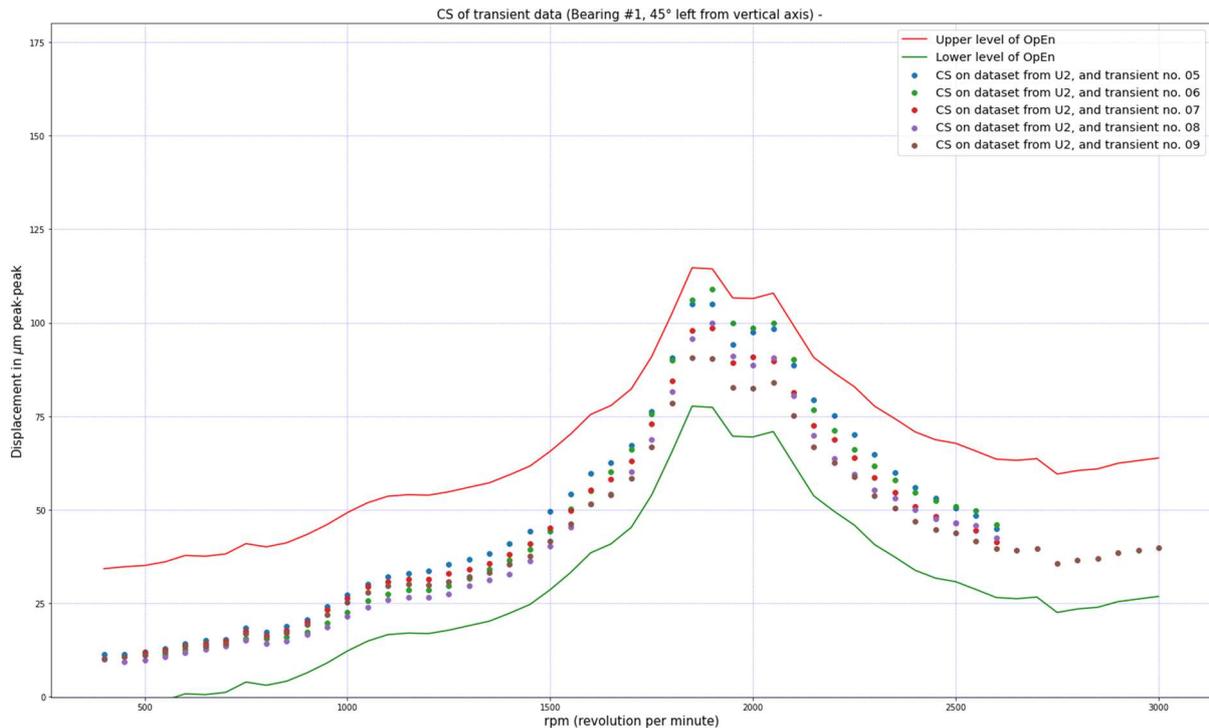


Figure 35. Example of reference transient data set with OpEn values given.

The next level of development of the OpEn method is to include additional vibration signal features. For example, applying the OpEn acceptance values concept to a vibration vector results in the OpEn acceptance region as an ellipse. Next, the ellipsis acceptance region is broadcasted on a predefined baseline transient vector for every rotational speed increment. In this way, the OpEn 3D defines the next level of acceptance region for complex vibration vectors.

4.3. Baseline measurements – selection and its consequences

In order to correctly detect an anomaly during transient, a reference transient is essential. However, selecting a single transient as the reference curve is not a good idea during the transient of large rotating machinery equipped with plain bearings. The turboset transient response under different operating conditions will have different values. An essential aspect is the oil temperature at the entrance to the bearings. In the author's experience, the transient of the same unit in winter and summer may differ significantly.

With this in mind, the author analyzed and made selections of the most representative transients from the set of all databases he had. The author analyzed over 250 databases of transient measurements. From among them, he selected and ranked transients according to the criterion of his methodology. He classified data included in the analysis into the following categories:

- dynamic condition: correct - helpful in the research,
- dynamic condition: acceptable but useless to define the baseline,
- data from runs with excessive misalignment of the HP-IP part,
- data with excessive unbalance, data with bearing oil instability 3,
- unusable data - from data points not covered by the test, with wrong sensor orientations.

He analyzed the correct dynamic state databases and selected reference data for baseline evaluation from them. The set of reference transient data set included approximately 25 transients containing all turbo-set's displacement vibration sensors, i.e., 14 sensors per transient, and all of the vibration signal features.

The author checked several averaging methods such as the median filter, standard deviation method, and the arithmetic mean. Due to the small amount of transient data points and a good results' correctness to the implementation ratio, he used the arithmetic mean.

4.4. 2D baseline

The data used to create the baseline is described in Section 4.3. A single feature of the vibration signal is used to create the baseline measurements. Figure 1 shows the flowchart of the method for calculating the baseline value for one sensor.

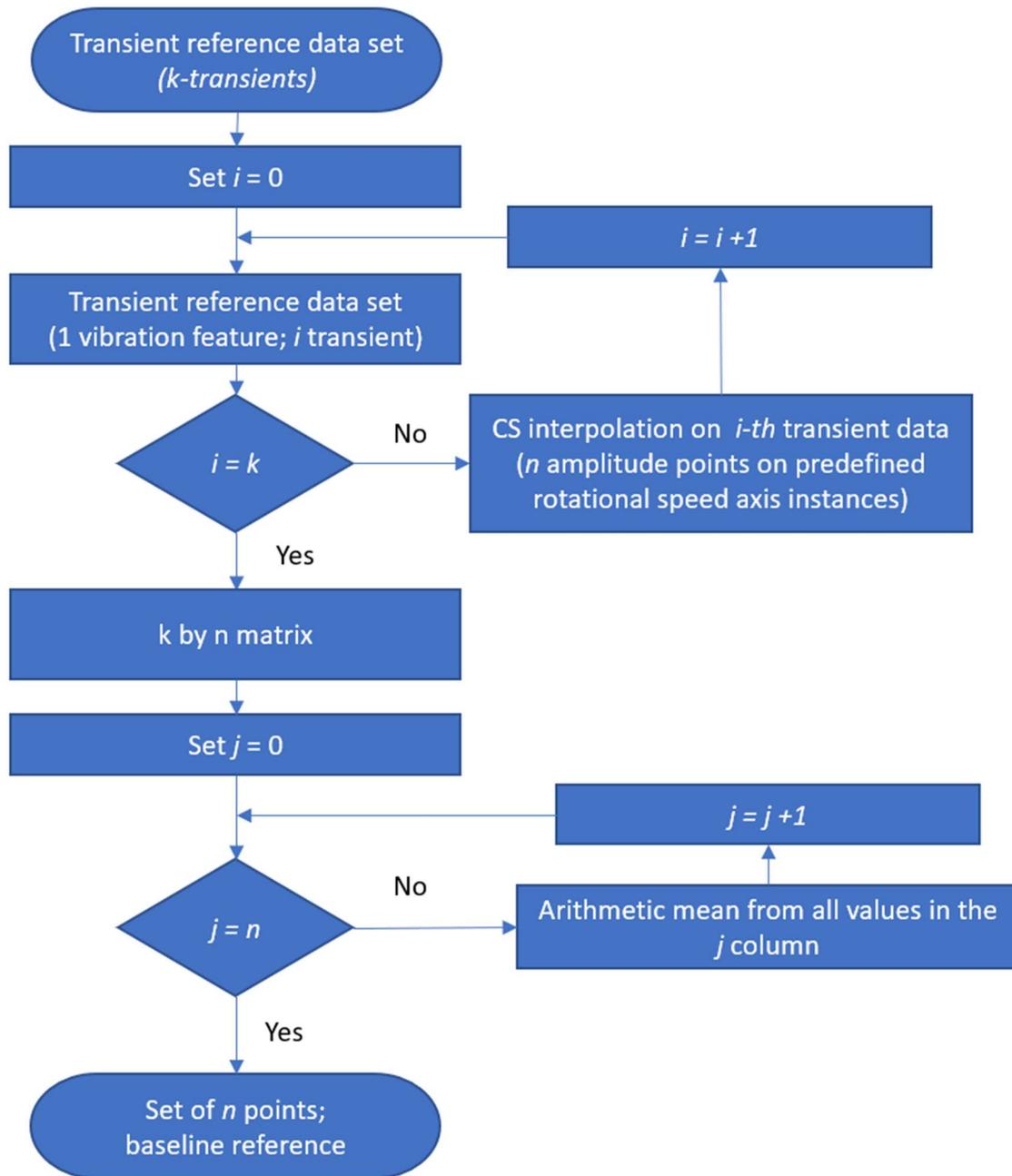


Figure 36. OpEn 2D baseline measurements flowchart.

First, the reference signal set is interpolated CS to determine the values of the individual reference transients in a predefined set of rotational speeds. The first part of this procedure creates a matrix of the k -rows of transients and the n -columns of rotational speed instances. Next, the arithmetic mean is performed column-wise, i.e., for each rotational speed instance. Finally, all values of the means create the baseline vector.

The procedure described above and presented in Figure 36 should be repeated as many times as sensors are included in the analysis to determine the baseline values for the entire turbine set. The set of all baseline values creates a matrix with dimensions l by k , where l is the number of sensors included in the analysis and k is the rotational speed increment. The result of the complete procedure is the

OpEn 2D Baseline Upper and Lower Limit matrix (OpEn BULL). It contains all acceptable values for individual sensors in all rotational speed instances.

4.5. Upper and lower OpEn values

Setting up an upper and lower bound for the OpEn is not a trivial task. The bounds mean the actual Operational Envelope above and below the centerline calculated as described in the previous section. We expect transients measured on healthy machines to stay within the area between lower and upper bounds.

Vance et. al. throughout their book [6] studied how different setups of the bearing applied to the same machine can produce dramatically different results. Eisenmann in [3] well described and explained how damping and stiffness affect response of the system during transient states. Thus, one needs to be aware of large effects caused by small changes.

The upper value should be set up higher, because of the non-linear nature of damping in bearing-rotor system. Non-linear nature of the rotor-to-bearing response is explained in by the numerous authors like Muszyńska [8], Eisenmann [3], Kiciński [9], Adams [11]. For instance, having properly aligned and balanced rotors on the same machine, different state of initial conditions (such as rotor and/or steam temperature, time of stand-still, etc.) can cause higher amplitudes, especially when whirling speed approaches to the resonant speed. Similarly, differences in inlet oil temperatures can produce differences in resonant peak amplitudes, and this is directly related to the oil damping properties.

The lower values are also important to analyze. The behavior of both static and dynamic response of rotor system changes together with crack propagation. Bachschmid described these phenomena in detail in [43]. Setting up lower value of OpEn can be a great help with shaft crack detection. As was presented in the report by Allianz[41] and Bently and Hatch [4] during evolution of a crack in the shaft its stiffness deteriorates. Such a phenomenon causes resonance frequency move to the direction of lower frequencies.

Based on the authors' experience, reinforced with suggestions from General Electric's engineering department following values were set up:

- The OpEn 2D case: The upper value is $24\mu\text{m}_{pp}$, and the lower value is $13\mu\text{m}_{pp}$.
- The OpEn 3D case: The amplitude value is $30\mu\text{m}_{pp}$, and the phase value is 20° .

4.6. OpEn 2D severity parameter definition

The core of the method is the detection of anomalies during transient states. The anomaly detection method can create two vectors having the same rotational speed values and different amplitudes in values. These vectors define the upper and lower bounds of the OpEn. In addition, they define the acceptance threshold of a new transient. Several norms are incorporated to automatically measure a distance between vector data points outside the OpEn for a new transient vector. The threshold will classify new vectors as they are measured. Only then can the method be proposed to machinery operators, and they will be able to use it without specialist knowledge and experience.

There is no single "silver bullet" method in order to appropriately assess the "good" or "bad" transient state. The author proposes a few metrics and compares their performance. The following metrics has been proposed for the OpEn 2D:

- *RMSE* – Root Mean Square Error from the whole transient;
- *KURT* – Kurtosis from the whole transient;
- *MAX_Oo_OpEn* – maximum distance above the OpEn upper value;
- *MIN_Oo_OpEn* – maximum distance below the OpEn lower value.

RMSE is a root mean square, as defined by Leon-Garcia in chapter 4 in [44], between the cubic spline interpolation of the reference transient (the OpEn centerline) and cubic spline interpolation of the actual data measured by a portable data acquisition system in the field on the same rotational velocity points. Figure 37 visualizes this norm. Equation defines *RMSE* norm.

$$RMSE_{OpEn} = \sqrt{\left[\sum_{i=l}^u \frac{(y_{ref_t_i} - y_{live_t_i})^2}{T} \right]} \quad (16)$$

Where:

- *RMSE_{OpEn}* – root mean square error of given transient
- *y_{ref_t}* – “healthy” value (reference transient data – center of OpEn)
- *y_{live_t}* – observed value (newly acquired, real transient data)
- *[l, u]* –rotational speed interval, common for *y_{ref_t}* and *y_{live_t}*
- *T* – number of common data points (samples at the same rotational speed points)

In the example on fig. 5 above *RMSE_{OpEn}* would be:

$$RMSE_{OpEn} = \sqrt{\left[\sum_{i=1800}^{2100} \frac{(y_{ref_t_i} - y_{live_t_i})^2}{7} \right]} \quad (17)$$

RMSE describes how far, on average, the newly acquired transient is from the OpEn, where only the centerline is considered. Thus, it measures the general average distance between these vectors.

KURT parameter is defined as the fourth standardized moment:

$$KURT_{OpEn}[X] = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2} \quad (18)$$

Where:

- *X* – is a vector of real data;
- μ – is the mean of *X*;
- σ – is the standard deviation of *X*.

The *KURT* parameter represents a distance between the two vectors with a higher weight of peaks, which should be detected automatically. For example, if a transient differs by a high value at only a

few data points, it cannot bring sufficient weight to the RMSE factor, but the KURT parameter detects it.

Equation (19) describes MAX_{Oo_OpEn} (abbreviation from Maximum Out of OpEn). It measures the highest distance above the OpEn upper value. This parameter measures and compares vector values at rotational speed like the previous ones.

$$MAX_{Oo_OpEn} = \max (\{abs (y_{OpEn_{uv_i}} - y_{live_{t_i}}) \mid i \in [l, u]\}) \quad (19)$$

- Where:
- $y_{OpEn_{uv_i}}$ – OpEn upper bound
- $y_{live_{t_i}}$ – observed value (real transient data acquired during transient)
- $i \in [l, u]$ – common rotational speed interval

In order to illustrate the above severity parameters, sample data from the real object transient will be presented below. The data comes from a +200MW class turbine set measured during commissioning after the outage. As shown in Figure 37 (top), the maximum value for this transient is 162 μ mpp, and the upper value of the OpEn in this rotational speed instance is given as 107.9 μ mpp. So, it equals 51.1 μ mpp. MAX_{Oo_OpEn} stays at zero as long as no point from the observed vector protrudes above the upper bound of the OpEn. Thus, it is a quick detection tool. It reacts to any violation of the upper bound.

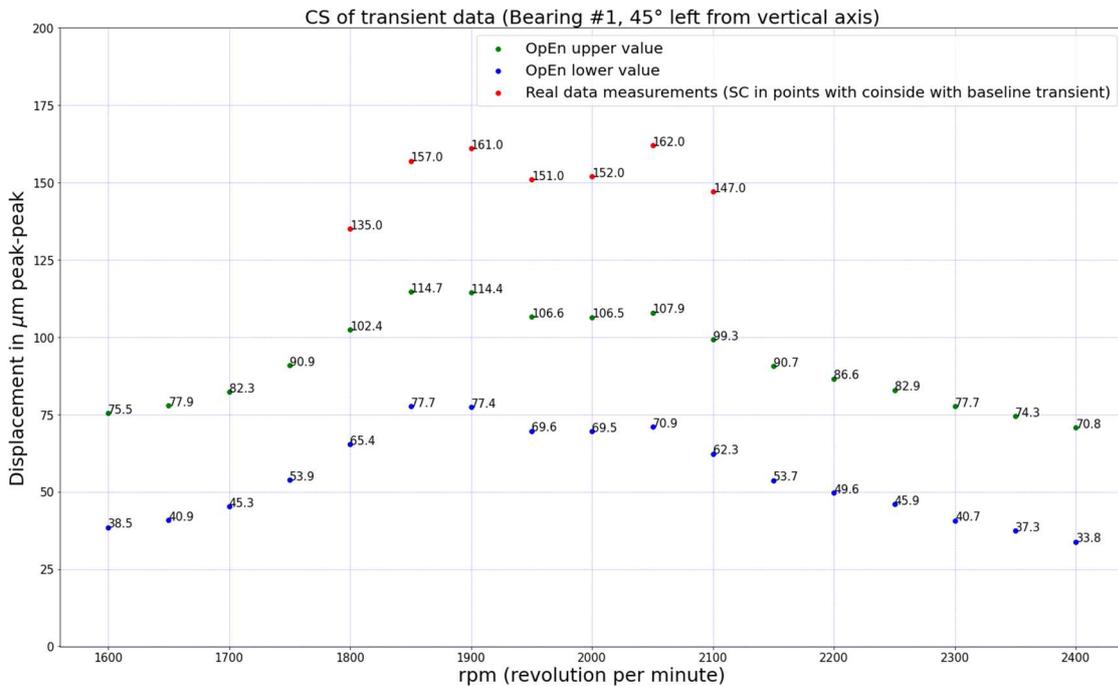
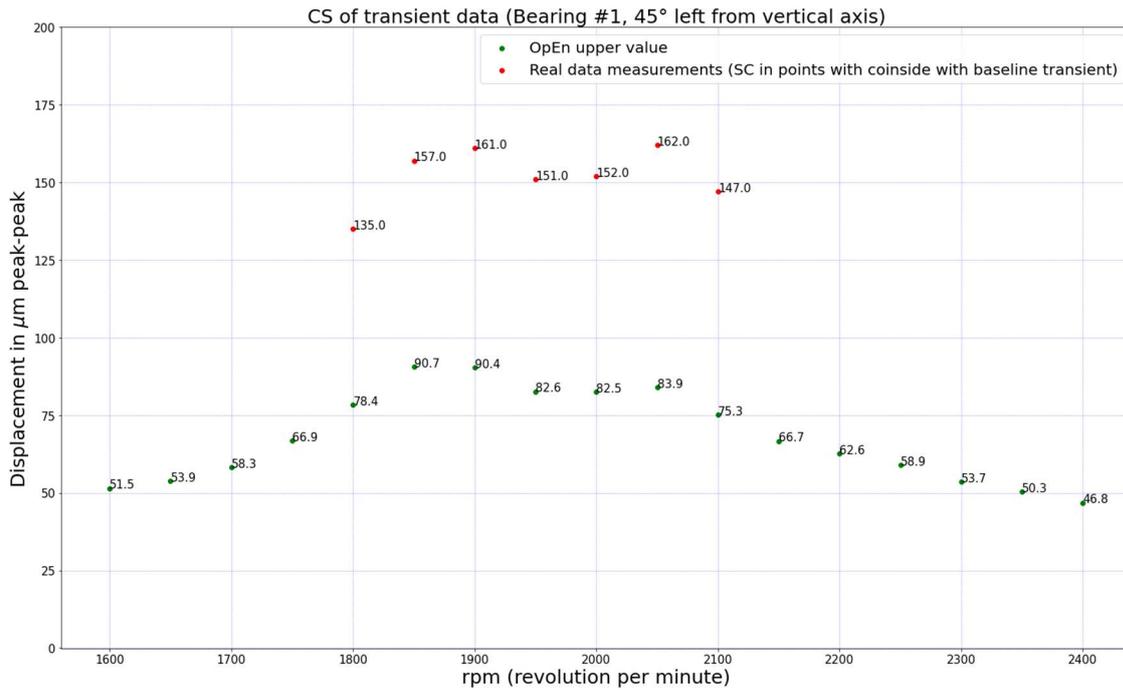


Figure 37. Severity parameter visualization versus real data values: top – RMSE norm; bottom – “Min Out of OpEn” and “Max Out of OpEn” values.

On the other hand, equation (20) describes the MIN_{Oo_OpEn} parameter (abbreviation from Minimum Out of OpEn). It is symmetrical to the previous measure but has a different value. The parameter measures and compares data at a common rotational speed interval.

$$MIN_{Oo_OpEn} = \max \left\{ \left| abs(y_{OpEn_lv_i} - y_{live_{t_i}}) \right| \mid i \in [l, u] \right\} \quad (20)$$

- Where:
- $y_{OpEn_lv_i}$ – OpEn lower value;

- y_{live,t_i} – observed value (real transient data acquired during transient);
- $i \in [l, u]$ – mutual speed rotation interval.

It is a measure of the highest distance below the OpEn lower value. The parameter measures and compares data at a common rotational speed interval. Figure 37 (bottom) presents the upper and lower values of the OpEn.

Due to the nature and the vibration response during transient states well described by Bachschmid et al. [43] and by Hajnayeb et al. [45] MAX_Oo_OpEn and MIN_Oo_OpEn parameter can help to detect a change in the bearing damping parameters, and developing of rotor cracks.

Together with the OpEn algorithm, the author proposes a set of parameters which can be used in order to automatically diagnose the transient. These parameters can be used in a conjunction with each other and other process data for better and more in-depth diagnostic purposes.

4.7. 3D baseline

Baseline measurements for the 3D configuration are different from those for 2D. Two vibration signal features must be combined into a vibration vector to determine the baseline value. The values of the vectors are the amplitude and phase of the individual harmonics described in section 1.2.4, describing vibrations in the complex plane. An example of such a vibration vector is the system's response to excitation with a synchronous force. The first coordinate for this vector is the signal amplitude, and the second is the phase lag value. A set of vectors for all predefined rotational speed instances creates a transient 3D vector response. The structure of the vector is similar to the n by two matrices. Each row in this matrix corresponds to a predefined rotational speed increment. Columns are amplitude and phase features of vibration signal, respectively. Figure 38 presents the flowchart of the method for calculating the baseline vector for the one sensor.

First, the two sets of transients composed of appropriately selected signal features are picked. Next, the CS interpolated is applied to each signal feature transient to obtain equidistant data points on the rotational speed axis. These operations create two matrices of k by n dimensions. One matrix describes points related to the amplitude and the second one - phase lag. Both matrices described above form the OpEn 3D Baseline Ellipsis Acceptance Region matrix (OpEn BEAR). These matrices are used to describe the acceptance region for the OpEn 3D method.

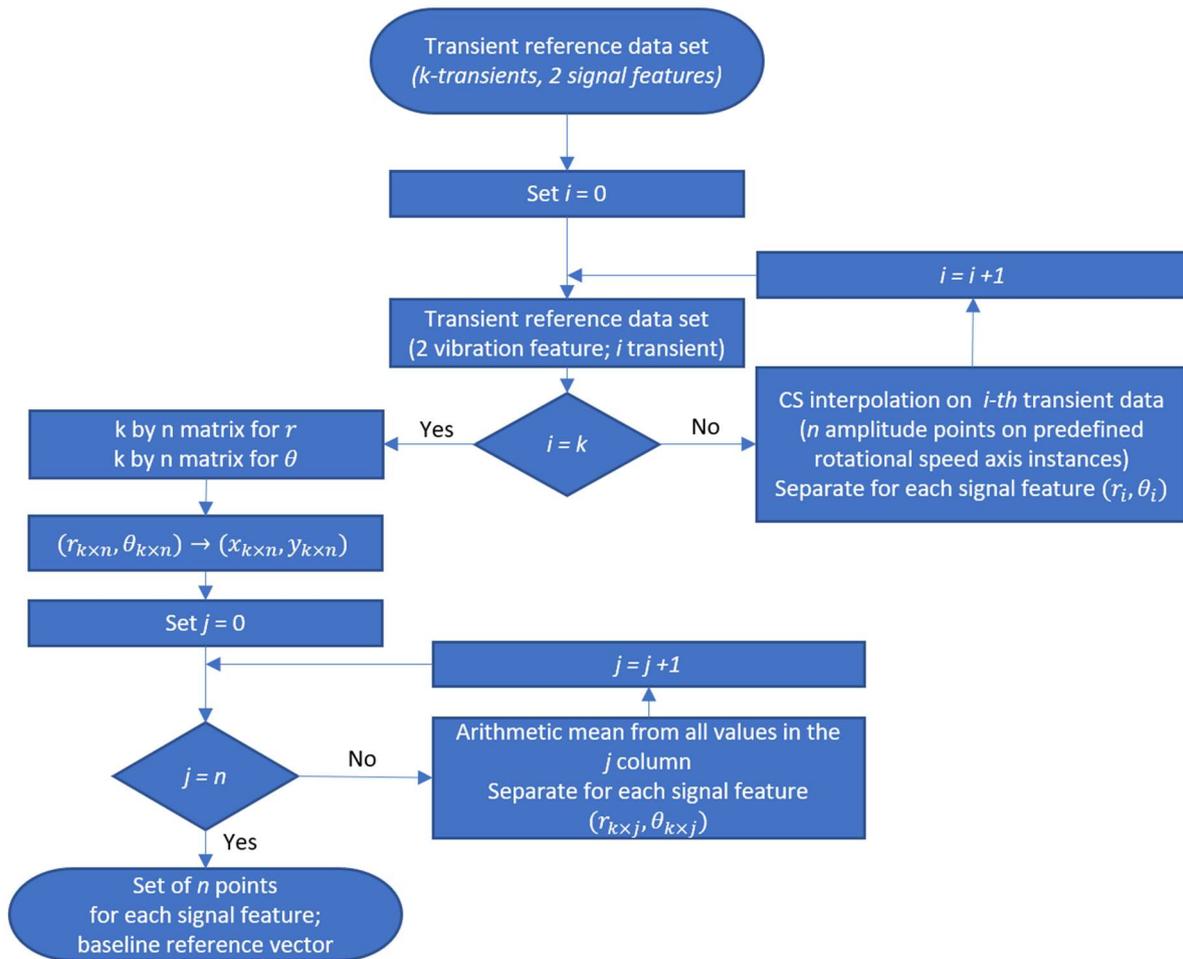


Figure 38. OpEn 3D baseline measurements flowchart.

Then, the amplitude and phase matrices of the vibration signal are converted to Cartesian coordinates. Each of the Cartesian coordinate matrix has the same dimensions as the r and θ matrices. Matrix rows correspond to individual transients and columns - rotational speed increments obtained after CS interpolation. Each of the matrices obtained is subjected to the arithmetic mean operation performed on all rows in a given column. The operation is performed for all columns of the matrix, that is, for all rotational speeds. The results of the above operations are two sets of Cartesian coordinates of the baseline vector. Each pair of the obtained coordinate sets is closely related to the corresponding rotational speed value. Associating Cartesian coordinate sets with the rotational speed domain creates a 3D vector. Figure 39 depicts two examples of baseline vectors obtained after procedure described above.

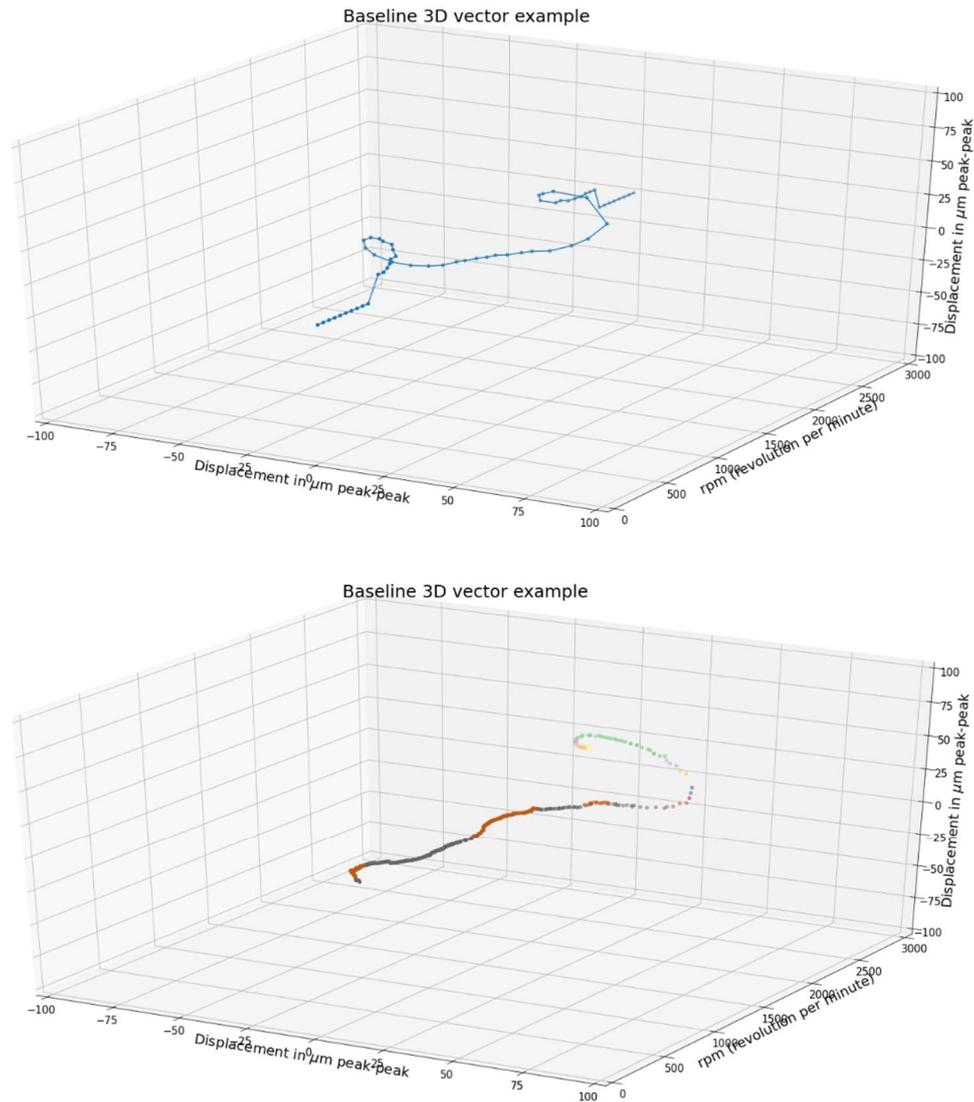


Figure 39. Baseline vector examples.

The point on the XY plane corresponds to the amplitude and phase of the vibration signal. The transient rotations are arranged on the Z-axis. This system creates a baseline vector trajectory for a particular sensor.

4.8. OpEn 3D severity parameter definition

One-feature only analysis may not be sufficient to diagnose potential malfunctions correctly. For this purpose, the author extends feature analysis on the vibration vector. This vector contains information about the amplitude and phase of the synchronous component.

When considering the overall vibration signal only, much information about the object is lost. For example, one can only tell if the vibrations are high or low. By extending the analysis of other features, the assessment of the machine state can be significantly improved. For example, information about the synchronous component and its phase lag is a significant dynamic parameter of the turbine set. Their change from one coast-down to another coast-down may indicate that evolution of malfunction

is taking place. Changing of the synchronous vector may indicate the appearance of additional force from unbalance and the need to take corrective actions. In turn, the change in the value of the second synchronous vector may be the symptom of the shaft crack propagation described in detail in the book [43] and research by Hajnayeb et al. in [45].

All rotational speed instances use an ellipse as an acceptance region for a given transient. The ellipse center is the baseline value described in Section 4.7. An exemplary set of ellipses constituting the acceptance region throughout the whole transient is shown in Figure 40. The top of the figure depicts the whole ellipsis set with its middle points as the baseline. The bottom chart presents these ellipses related to their rotational speed instance. This exemplary set refers only to one eddy-current sensor. Finally, to represent the baseline behavior of the shaftline fully, 13 more sets are defined.

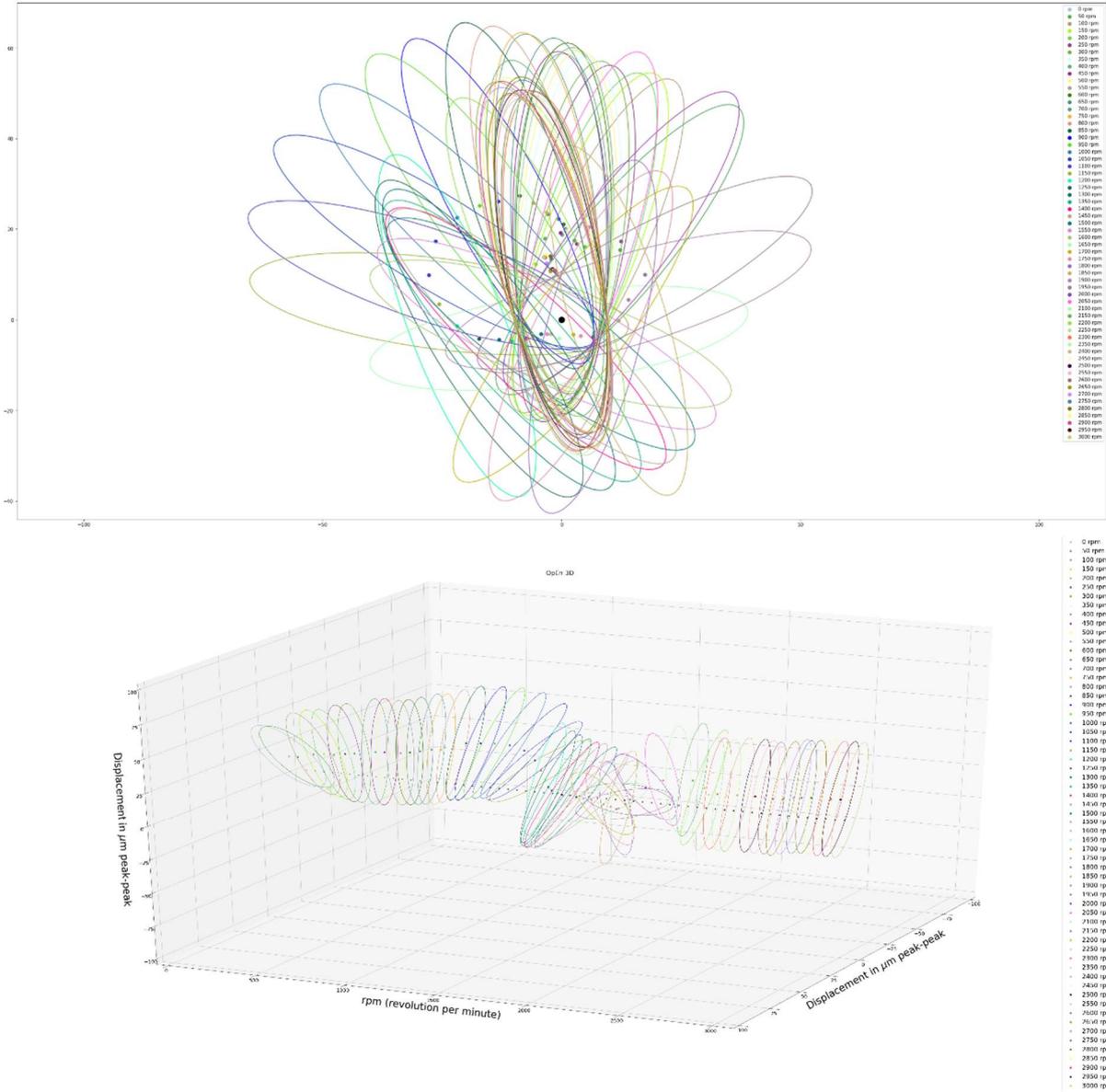


Figure 40. Example of OpEn 3D ellipsis set: top - mean baseline and ellipse at each speed point during transient; bottom - 3D view (including turbogenerator rotational speed) of the ellipse set for the whole transient state for one sensor only.

The severity parameter to evaluate the vibration vector analysis is the value of the shortest distance between actual data and the OpEn 3D of the ellipse acceptance region.

The acceptable values of the amplitude and phase of the vibration vector form an ellipse. The ellipse area defines the OpEn acceptance region and the edge of the ellipse - the limits of OpEn 3D.

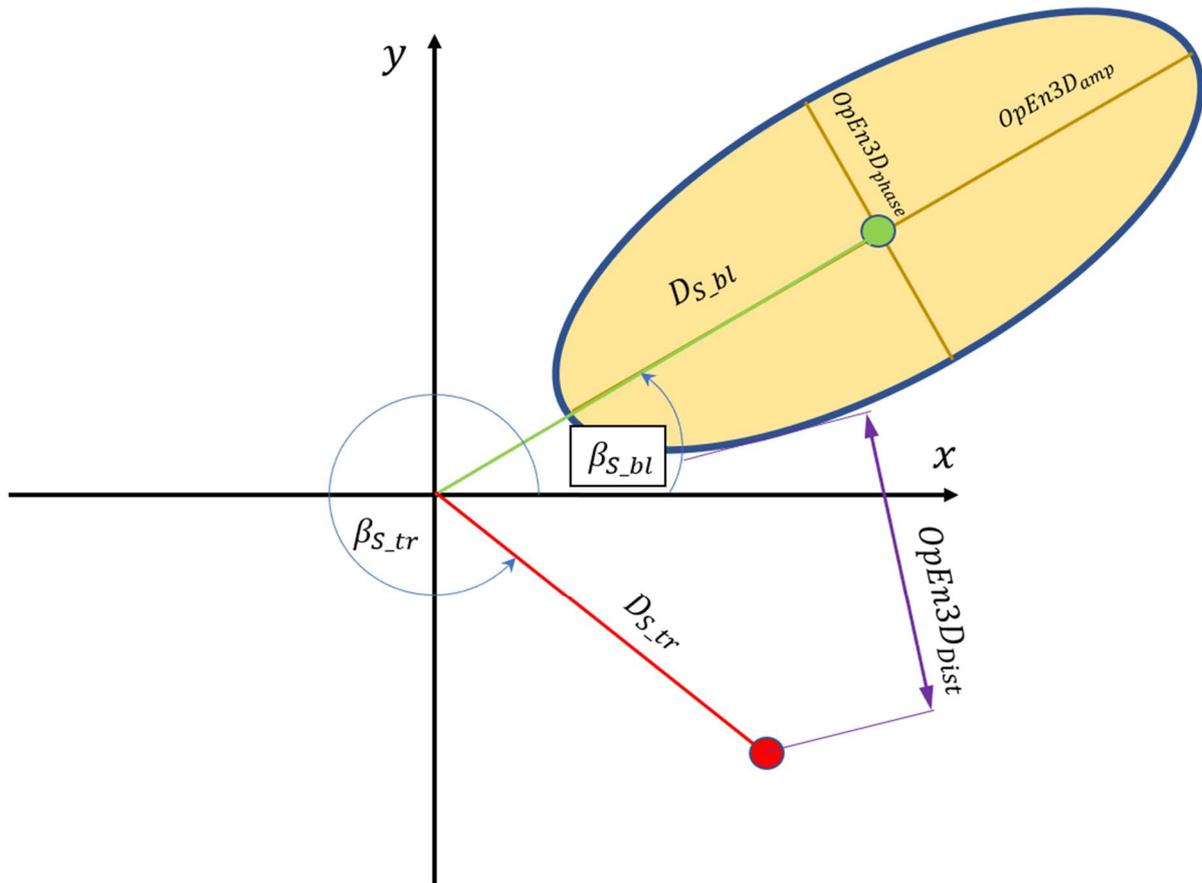


Figure 41. OpEn 3D acceptance region and severity parameter definition.

The $OpEn3D_{Dist}$ severity parameter to evaluate the vibration vector analysis, mark in the Figure 41 by the purple line, is the value of the shortest distance between actual data and the OpEn 3D of the ellipse acceptance region. The acceptance ellipse middle point, defined as the intersection point of the semi-axis of the ellipse, lies at the baseline point defined for each rotational speed value. This is the green point, i.e., the end of the vibration vector, which in the polar system has the coordinates (D_{S_bl}, β_{S_bl}) , respectively depicted in Figure 41. The acceptance region for vibration vectors proposed by the author was determined thanks to the author's many years of experience in analyzing transient states of large rotor machines. The ellipsis semi-axes are the acceptable amplitude, and the phase marked in Figure 41 as $OpEn3D_{amp}$ and $OpEn3D_{phase}$, respectively. The actual transient vector is described by (D_{S_tr}, β_{S_tr}) coordinates and marked in Figure 41 and Figure 42 in red. The new transient data is compared to the OpEn3D acceptance region. In addition, the actual vibration vector is compared with its corresponding ellipse for each rotational speed separately for the current transient. The set of acceptance ellipsis for the whole transient for one sensor is depicted Figure 40. Figure 42 depicts an example of applying this parameter. When transient data are within the OpEn 3D acceptance region,

the shaftline is considered to be in good condition (Figure 42 – top). The $OpEn3D_{Dist}$ distance from the OpEn 3D acceptance region is equal to 0. Therefore, no further actions are undertaken.

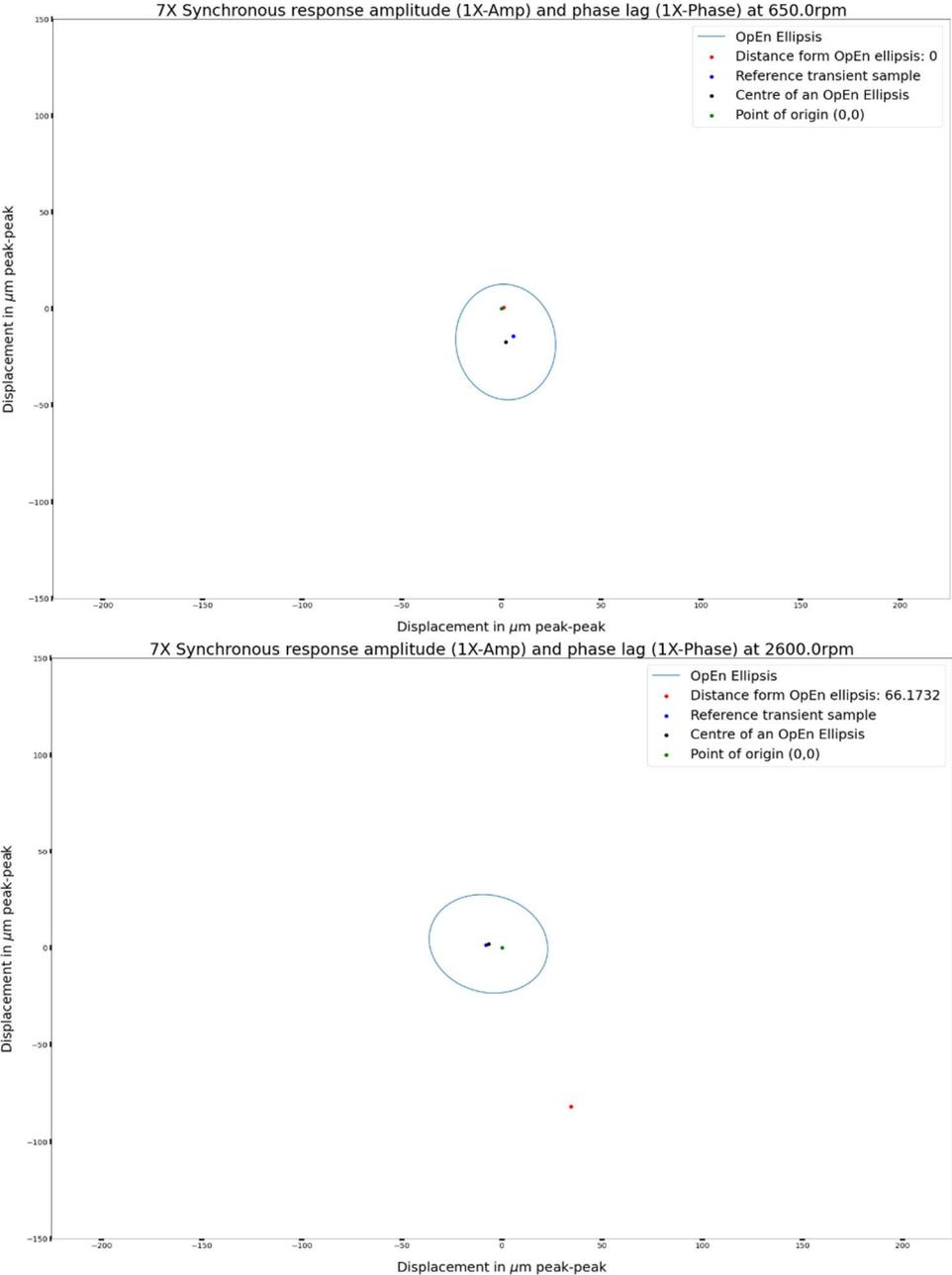


Figure 42. OpEn 3D ellipse severity parameter visualization: top – actual data inside of an ellipse; bottom – actual data outside of an ellipse.

On the other hand, if the data fell outside the OpEn 3D acceptance region, the distance of the vibration vector from the ellipse is calculated. Figure 42, bottom, presents an example of determining the severity parameter as the distance of the vibration vector from the ellipse. At 2600RPM, the vibration vector point is outside the acceptance region. Next, the severity parameter is calculated and defined. It is the closest distance from the point outside the ellipse to the ellipse.

5. Fault Identification method

The first part of the FDI algorithm is the fault detection method, which was proposed in the previous section. If a fault is detected, it is very useful to learn what was the reason for a deviation from the baseline. To answer this question, one needs to perform the fault identification. This operation constitutes the second part of the method proposed in this dissertation. This section presents the details of this method.

5.1. Parameter identification method

The vibration response of the unit remains the fundamental method to assess the technical state of turbo-sets. Sometimes the interval between shutdowns can last up to a year without intermission. During this period, the turbo-set is operated in varying conditions such as significant load change, different steam temperatures and pressures, and many more. These conditions can cause a large amount of stress, which eventually can lead to fatigue and, in extreme cases, to failure. The automated assessment of complex technical systems was the subject of numerous research.

After a series of total machine destruction cases, German insurance company Allianz carried out extensive research [41]. It concluded that without a proper turbomachinery diagnosis, the consequences could be fatal to the equipment and very dangerous to the people who operate them. Since this report, a significant amount of research and effort has been invested in inventing and implementing new and more precise methods of technical condition assessment. Demetel et al. in [27] highlight the fact that most industrial systems are non-linear and require appropriate analysis methods. Each such attempt must include a feature extractor and classifier. The authors have analyzed multiple generic methods to diagnose the pneumatic systems of the material handling systems, starting from dimension reduction to clustering for classification. Zagorowska et al. [36] presented an interesting approach and new insights to track the evolution of malfunction during steady-state operation with a novel approach to trend tracking technic. Głuch, in chapter 3 of [39], describes a state-of-the-art thermal and flow diagnostic of steam turbines in great detail. Duan et al. in [46] presented several attempts at tracking turbogenerator degradation with Deep Neural Networks. However, the features were calculated from turbo-set operation data, not during transient states. Akhtar et al. [40] perform an analysis of a complex case of Gas Turbine vibrations. They confirm that tedious analysis work and the availability of experts are required to detect and identify a large turbomachinery fault correctly. Sachin et al. [47] present an interesting approach to bearing diagnosis. They propose reducing the number of features. Paper claims that proper feature ordering and selection could significantly improve classification accuracy, especially for machines equipped with modern CMS, which acquire and calculate many features.

The above methods and research improve the quality of assessing the technical condition of machines. However, they do not consider the transient states of high inertia machines, which can last up to two hours, as depicted in Figure 19. Therefore, there is a lack of a method to help the maintenance personnel quickly assess the state of machinery during turbine cast-downs and start-ups, ideally in an automated way. Data from these states are as important as the steady-state ones, but the amount of

transient data is incomparably smaller than that of steady-state operation. This makes this data an invaluable source of information. The amount of information about the dynamic state obtained from them determines the need to introduce supervision, assessment, and diagnostic system during coast downs and start-ups of the above devices.

The main quality indicator of rotating machines is the bearing vibration. Therefore, there are many works dedicated to bearings' fault diagnosis. Wei et al. [48] use an adaptive approach to extract features from faulty bearings with success. Kun et al. [49] have also proposed an interesting approach to bearing faults classification. The paper presents the use of Ensemble Empirical Mode Decomposition (EEMD) and Singular Value Decomposition (SVD) to extract fault features, and then an advanced clustering method is used for fault pattern recognition. Wang et al. [29] also use ML technics and incorporate them into planetary gearbox malfunctions detection. [28]

The papers mentioned above studied only machines with rolling element bearings or planetary gears during their steady-state operation conditions. There is a lack of papers concerning assessing the behavior of fluid-film bearings in large rotating machinery during transient states. The availability of machines equipped with rolling bearings or planetary gears is significantly greater than that of large rotor machines. Therefore, examining small rotor devices is not associated with a large financial outlay and workload. There are many small rotor machines in the industry. Often, due to their redundancy, they can be freely tested without significantly increasing the company's production costs. The available data makes the ML idea and any AI-based algorithms justified in these cases. Moreover, large amount of data increases the probability of detecting damage and recognizing a fault pattern, and also gives a better training-to-testing data ratio.

Unfortunately, there is a lack of papers concerning assessing the behavior of fluid-film bearings in large rotating machinery during transient states. The author of this dissertation aims to reduce this gap by introducing a method of automatic anomaly detection during transient states.

5.2. Malfunction identification methods

Large turbogenerators exhibit a highly non-linear response during transient states due to various factors, such as passage through its resonances, lube oil damping, fluid forces (e.g., steam), rubbing, and other factors. The author proposes decomposing each transient response into a set of signals to deal with such a complex response. He utilizes the signal structure known "a priori", and such decomposed signals can be interpreted separately, thereby facilitating the analysis. Nonlinearity in transient response, even in respect of a single rotor supported by two bearings, can cause some difficulties regarding its description in a mathematical sense.

The transient state data points of a rotor have a highly nonlinear nature. The author proposed decomposing a transient curve into a few known and less complex signals to address a particular malfunction of complex behavior. The advantages of decomposition on nonlinear stochastic signals were thoroughly described by Chui and Mhaskar [50], Cicone [51]. Chui and Mhaskar, and Cicone researched the enhancements and improvements of the decomposition of nonstationary time signals. Chui and Mhaskar proposed a new mathematical theory behind their method that performs better

than the synchrosqueezing transform method. Cicone presented a detailed analysis of the methods for decomposing nonstationary signals. He also presented a detailed methodology for composing the signal into individual, simpler components. The methods used do not require "a priori" assumptions. He also analyzed the most popular decomposition methods, indicating their advantages and disadvantages.

Each decomposed signal represents a different rotor behavior during the transient. The author offers three types of functions to approximate the transient signal curve. Equations (21)–(23) represents the analytic form of the decompose type of the functions.

$$f_1 = a_1 e^{\frac{(x-b_1)^2}{c_1}} \quad (21)$$

$$f_2 = a_2 \left(\frac{x - x_0}{x_k - x_0} \right)^2 \quad (22)$$

$$f_3 = b_3 \quad (23)$$

Where:

- f_1 – Gauss function with a_1, b_1, c_1 parameters respectively,
- f_2 – parabola with a_2 parameter and b_2 as a bias term respectively,
- a_1 – amplitude of the Gauss function,
- b_1 – placement of the resonance peak along the rotational speed axis,
- c_1 – width of the resonance,
- a_2 – factor related to synchronous response,
- x_0 – starting point of centrifugal force response,
- x_k – rotational speed range (in given transient),
- b_3 – bias constant (electrical/mechanical runout).

Based on assumption above, numerous function combinations with different sets of parameters can be superposed to fit the measured data.

Hence, the complex function described by the equations (6) and (7) can be represented as a sum of the simple functions described by the equations (21)–(23), similarly to the procedure described by Chui [50] and Rao [52]:

$$\varphi_{approx} = \varphi_1 + \varphi_2 + \dots + \varphi_n = \sum_{i=1}^n \varphi_i \quad (24)$$

where:

- φ_i – decomposed function described by the combination of equations (21)–(23).

The MSE (Mean Square Error) and the MAE (Mean Absolute Error) are used as a measure of accuracy concerning the decomposition approximation. MSE_{tran} has been defined as a sum of the squares, whereas MAE_{tran} has been defined as a sum of the absolute values between real function φ_{real} and our approximation φ_{approx} within the space of all the samples throughout the transient in the

equations (7) and (8), respectively. Consequently, the MSE_{tran} and the MAE_{tran} are used to assess a fit of the decomposed parts to the actual data which was obtained during the measurements:

$$MSE_{tran} = \sum_{\omega=1}^N \left[\varphi_{real}(\omega) - \sum_{i=1}^n \varphi_i(\omega) \right]^2 \quad (25)$$

$$MAE_{tran} = \sum_{\omega=1}^N \left| \varphi_{real}(\omega) - \sum_{i=1}^n \varphi_i(\omega) \right| \quad (26)$$

where:

ω – revolution per minute (rotational speed).

The malfunction identification method proposed in the doctoral dissertation consists of a tournament between three scenarios involving different variants of decomposed functions approximating a given transient. The simplest scenario proposes an approximation with three partial functions, i.e., one Gaussian function with a parabola and bias. Six parameters describe the partial functions of this scenario. The most complex variant consists of five partial functions, i.e., three Gaussian functions, a parabola, and a bias function. Twelve parameters are used to describe the partial functions for this scenario.

The proposed method can yield up to 12 parameters describing a single transient with three resonance peaks, unbalance, and run-out. Each resonance is described by three parameters that can be monitored independently: the resonance peak, its placement, the rotational speed interval, and the width related to damping in a bearing. In addition, rotor unbalance response is monitored with two parameters of a single parabola. According to the tests carried out during the author's research, a set of the 12 parameters described above should suffice to correctly model and monitor the transients and their most essential parameters.

5.3. Multidimensional Data Driven Decomposition (MD3)

Multidimensional Data Driven Decomposition (MD3) is an extension of a concept presented by authors in [14]. This method consists of two main parts listed below:

1. Real data preparation and preprocessing,
2. Identification of decomposed function parameters.

The first step is required to transform very different data sets into unified vectors, which can be a subject of comparison. The content of data sets measured on the physical objects is often different. Each transient can vary depending on a large number of external factors, which are not recorded in the vibration response of the system. For instance, there happen to be processed depending conditions, e.g., lousy quality vacuum in the condenser, which can cause a machine to stop much quicker than during normal coast-down. In addition, there are transients when the Full Speed No Load (FSNL) state cannot be achieved during the start-up. There are no turbine-related issues (common ones are vacuum-related problems, lube oil-related problems, boiler-related issues, or others.). Even if the

whole RPM span transient is recorded, it is evident that the vibration data are not stored at the same rotational speed instances.

Each monitoring and acquisition system records data with different resolutions in terms of time and rotational speed intervals. It is a result of design of monitoring systems used in the field and numerous datasets have such a feature. This data already exists and it is not possible to repeat these measurements. Portable measurement system configuration has two different triggering options, according to the change of speed and time. Typical values are 20-60 seconds for time intervals and 5-50 RPM for the rotational speed change. Depending on the trip time instance, the measurement systems record the transient process at different points in time and speed. Therefore, direct comparison of the transient vibration parameters in an automated way is not possible.

As a first step, the vibration data need to be preprocessed to have the same set of RPM values. A cubic spline interpolation was introduced as the preprocessing procedure to solve this problem. It allows defining a set of equally spaced rotational speed values at which the vibration values shall be interpolated. Later, the fitness functions of the decomposed functions will be evaluated. Barszcz and Zabaryłło [13] describe the usage of cubic spline interpolation and its benefits in transient state analysis. Dyer and Dyer in [25] and Barszcz and Zabaryłło in [13] present the advantage of using the equally spaced knots for the polynomial spline function (i.e., equally spaced rotational speed increments during transient). De Boor, in his book [53] in chapter XIV shows that the advantage of cubic spline interpolation is to smooth the interpolation function in the points of interest. In our research, the points of interest at which the cubic spline is calculated are the rotational speed instances from the following $\{rpm_1, rpm_2, rpm_3, \dots, rpm_{max}\}$. Typically, the set consists of equidistant values, e.g., 200 RPM ending at 3,000 RPM (for European power plants) with a 50 RPM distance between points.

The second step of the procedure decomposes a preprocessed transient into essential components. Finally, the procedure relates a set of function parameters and coefficients to physical phenomena occurring during the coast-downs and start-ups across the shaft line when a fault is present. Thus, the input transients are decomposed into more straightforward base functions. These functions are used as a measure of particular malfunction. Based on experience and research, the authors took a set of three decomposition base functions into account:

1. One Gaussian function, one parabola, and one constant/bias function. It produces a set of six parameters. This scenario can identify one critical speed and unbalance.
2. Two Gaussian functions are considered, one parabola and one constant/bias function. It produces a set of nine parameters. This scenario can identify rotors with two critical speed zones and unbalance.
3. Three Gaussian functions are considered, one parabola and one constant/bias function. This scenario can identify rotors with up to three critical speed zones and unbalance.

Each scenario is qualified based on that same quality performance parameter, namely MSE. Mean Squared Error (MSE) measures the fitness function to be minimalized. Equation 1 presents the definition of MSE, as defined by Leon-Garcia in chapter 4 in [44].

Figure 43 presents the flowchart of the MD3 method divided into particular operations.

The first four blocks refer to step 1, as described at the beginning of the section. The MD3 method starts with detecting an anomaly in the OpEn procedure. After detecting a potential malfunction, the data in a set of features for individual rotational speed increments are passed for preprocessing. There are several steps in the newly received data preprocessing procedure. The first step is to sort the data samples according to the rotational speed value. This step is essential when there are different transient conditions configured. For example, during coast-down, the recording of the rotational speed would start at the highest one. The situation reverses when the start-up is recorded, and the rotational speed will start at 0 RPM. The procedure sorts the data in ascending order to rotational speed values to analyze the data systematically. Next, the samples with the same rotational speed tags are removed from the dataset. Further on, the speed range for the currently analyzed transient is determined, and the range is divided into equidistant points on the rotational speed axis. The Cubic Spline interpolation establishes equidistant points from the current transient as the last operation in this step.

The latter operations in Figure 43 belong to the second step. For each transient, three scenarios are evaluated based on the MSE quality index. First, the scenario with the best-decomposed functions fitting parameters, i.e., the smallest value of the MSE index, is chosen to represent the current transient state. These parameters can be used in malfunction identification, and they are stored for future reference.

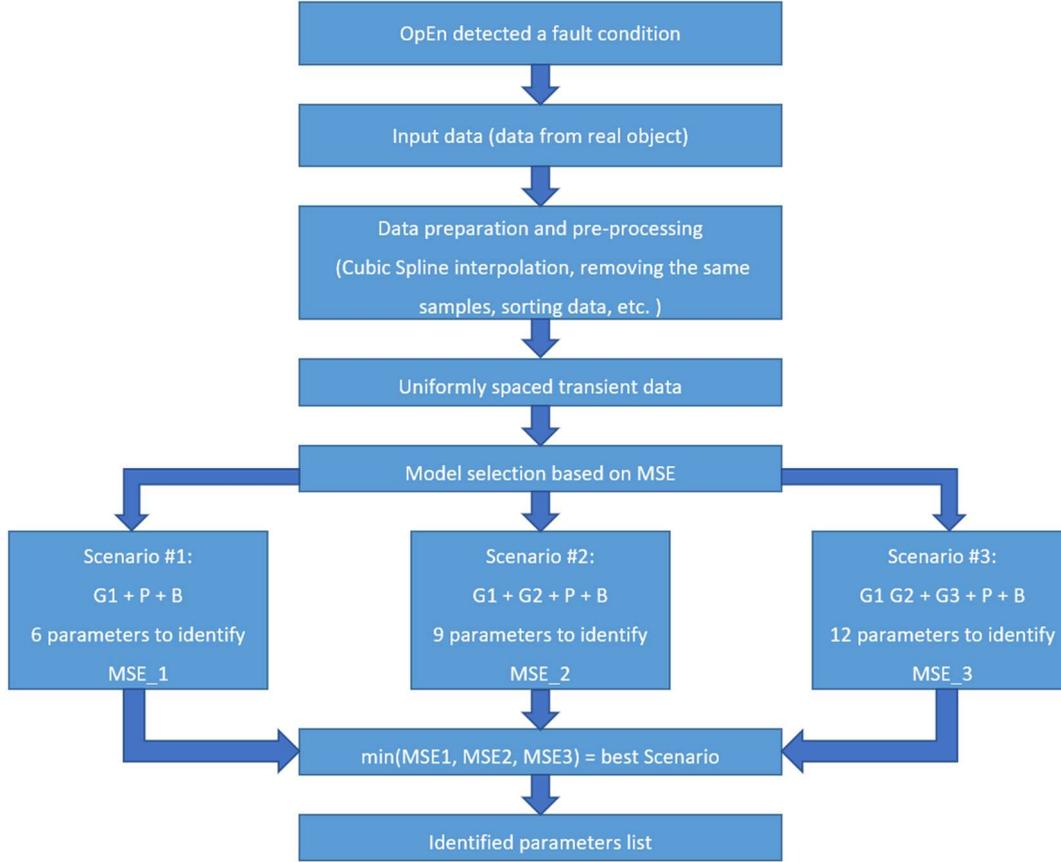


Figure 43. The MD3 method flowchart.

Estimating the values of the proposed functions is the heart of the method. The author uses Differential Evolution (DE) algorithm to determine these parameters. The algorithm finds the best fit of the assumed model vs. real-object data. Equation 2 presents how the decomposed functions are combined to form the final transient function. Finally, Equations 3÷5 present the analytical representation of individual decomposed functions.

$$\varphi_i(rpm) = \sum_{j=1}^n Gauss_{i,j}(rpm_k) + P_i(rpm_k) + B_i(rpm_k) \quad (27)$$

$$Gauss_{i,j}(rpm) = a_{g_j} e^{\left(-\frac{(rpm_k - p_{g_j})^2}{2w_{g_j}^2} \right)} \quad (28)$$

$$P_i(rpm) = a_{p_i} \left(\frac{rpm_k - x_p}{rpm_{max} - x_p} \right)^2 \quad (29)$$

$$B_i = const_b \quad (30)$$

Where:

rpm_k – particular rotational speed from equally spaced rotational speed increment set where $rpm_k \in \{rpm_1, rpm_2, rpm_3, \dots, rpm_{max}\}$;

rpm_{max} – maximum rotational speed in the transient set;
 φ_{real} – cubic spline interpolation of real data in ω instances;
 φ_i – cubic spline interpolation of real data in ω instances;
 n – a number of Gaussian functions chosen for the decomposition;
 $j \in \{1,2,3, \dots, n\}$ – particular Gaussian function in n set;
 $Gauss_{i,j}(rpm)$ – j -th Gaussian function;
 P_i – parabola function (2nd-degree polynomial);
 B_i – bias function with its parameter (constant not dependent on rpm);
 a_{g_j} – amplitude of j -th Gaussian function at the top of its critical speed (resonant speed);
 p_{g_j} – the peak of the j -th Gaussian function in terms of rotational speed;
 w_{g_j} – width of the resonant zone of the j -th Gaussian function;
 a_{p_i} – amplitude of i -th parabola function at the end of the recorded transient;
 x_p – point of start of the parabola in terms of rotational speed (rpm);
 $const_b$ – constant term taking into account initial vibration indication of the shaft.

The benefit of the method is physical interpretation of the aforementioned parameters. This is important advantage over methods, which yield parameters without clear connection to physical features of particular faults.

5.4. Identification parameters

Based on the assumption described in Section 5.2, the author proposes that several decomposed function combinations can be applied to fit the transient curve. In addition, different sets of functions and Identification scenarios can be applied to fit the measured data in the best way.

The complex transient function behavior described by the equations (4)-(7) can be represented as a sum of the simpler functions described by the equations (21)-(23) and (28)-(30), similarly to the procedure described by Chui and Mhaskar [50] and by Rao [52].

The first decomposed type of function is the Gaussian probability distribution function described by the equation (21) and depicted in Figure 44. It can approximate the resonance behavior of the rotor. Its parameters define amplitude at the resonance peak, placement along the rotational speed axis, and the resonance width, which in the example in the Figure 44 are 1850, 100 and 700, respectively. These parameters are essential when analyzing the rotor response in the critical interval. Changes can indicate a potential issue with the bearing parameters and a start or evolution of a bearing degradation process. Further, parameter change of resonance peak amplitude can indicate a potential imbalance problem. When combined with the change width parameter, it can indicate the deterioration of the bearing damping parameters. Finally, the placement of the resonance peak change can indicate rubs during the transient state or, in combination with the change in 2X amplitude response, can be a symptom of shaft crack propagation.

The second one is the parabola depicted in Figure 44 by red dotted line, and described by the equation (22). It can represent the centrifugal force related to the unbalance response. Changes in these function parameters indicate that some imbalance to the rotor was introduced. There can be several reasons for that. According to the author's experience, the most common change in imbalance

response is the circumferential mass reduction caused by a sudden event (falling off the blade's tip) or erosion-induced reduction of the blade mass. In both cases, centrifugal force changes, and the imbalance response of the rotor affects the parameter of the parabola function.

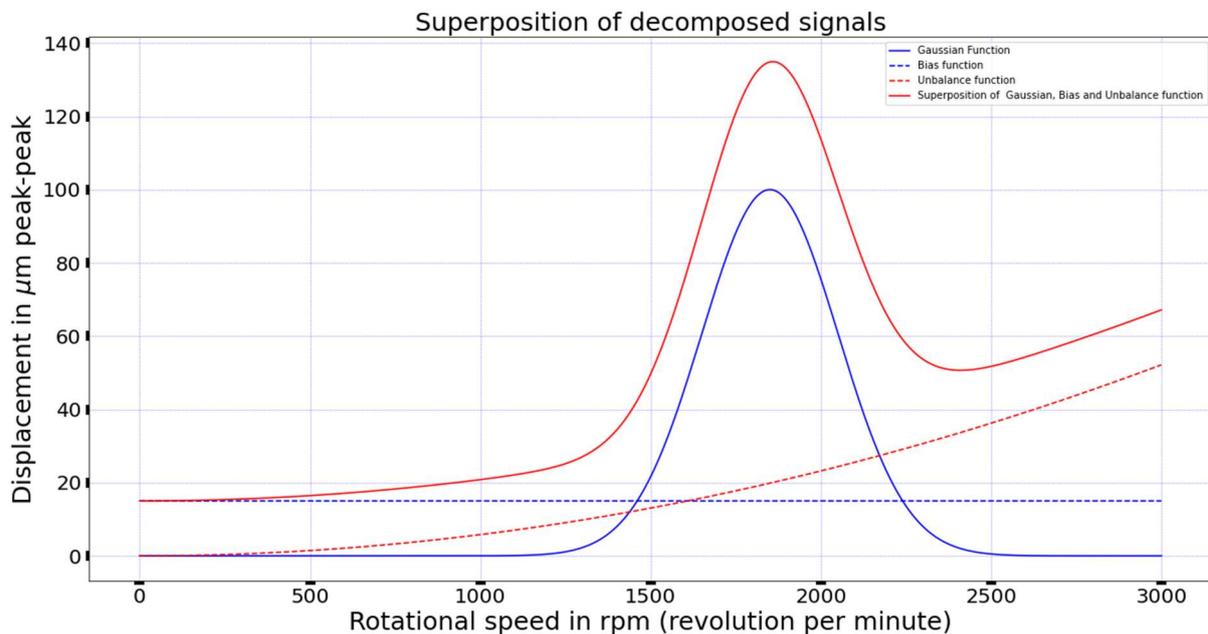


Figure 44. Example of the decomposed signals simulating runout, resonance and unbalance and their superposition.

The third one is the constant function described by the equation (23). It can represent the initial mechanical runout of the shaft. It can also account for the measurement noise.

5.5. Differential Evolution algorithms as a part of Genetic Algorithms family

The fundamental question is how to properly adjust the parameters of the functions proposed in the previous sections. The method should be based on the data available after each transient. After analysis of literature and initial tests, the Genetic Algorithms were selected for this task.

Genetic Algorithms (GA) are parallel mathematical algorithms that transform each population (i.e., individual parameters of mathematical objects) into a new set of parameters based on a fitness function. The fitness function is a way of evaluating the cost of an individual and population to adapt to the environment related to each population and its parameters. They are based on Darwin's theory of evolution, which stated that only the best-adapted individuals (the fittest ones) will survive to reproduce and create a new population that would be better adapted to the natural conditions. The DE algorithm is part of the Genetic Algorithms (GA) family. GA are based on the concept of population evolution in a natural habitat. The idea of finding the best solution to a given problem (goal function) was described by Koza and Poli in Chapter 5 [54]. Finding the solution starts with some initial set of solutions with different parameters (called population) and using the quality parameter (called fitness function) to determine the best solutions from the solutions' pool (the fittest individuals from the particular population, called parents). Then, another set of solutions based on the parents (called

children) is produced. Children inherit many properties from their parents, but they can also be subjected to modifications in their parameters (called mutation and genes' crossover). Then, the new solution set of the next evolution is ready to be evaluated. This can go on until the quality parameter is met or for an arbitrary number of evolutions. Throughout their book they describe the basics of GAs and deliver an extensive discussion on its advantages and limitations of GAs. The main advantage of the GA is its possibility to adapt to different problem-solving tasks.

GA algorithms are used extensively across many fields of science and engineering. For example, Roetzel et al., in chapter 6 in [55], described the heat exchanger networks design using GA with network parameters like total annual cost, target temperatures hot and cold temperatures with good results. Furthermore, Li et al. in [56] present a GA enhancement and reinforcement to feature an extraction and classification algorithm based on the neural network used to diagnose electrocardiogram signals. They argue that GA feature optimization and Back Propagation Neural Networks could be applied in cardiac arrhythmia automatic identification due to dimension reduction. It can yield a classification accuracy of 99.33%.

The Differential Evolution (DE) is an evolutionary algorithm, which constitutes part of the GA family. The DE is a stochastic search algorithm based on the population. What is more, the DE is based on Darwin's theory of evolution, where the strongest (i.e. fittest) and therefore the best individual has the best chance to survive and reproduce. The "fittest" parents from the population pass on their genes (qualities) to the next population, which has a greater chance of survival, accommodation, and reproduction, even higher than their parents have in a given environment. An example of such an approach was described by Storn and Price [57]. Further implementations and enhancements of the DE algorithm are presented in Qin et al. [58] and Das et al. [59].

Muratoglu [60],[61] presents an interesting application of the DE algorithm to optimize rotating machinery, namely hydro turbines. In the paper, five different primary hydrofoil families were optimized and scaled. The optimized hydrofoils were found to deliver successful performance for hydrokinetic turbines.

The aforementioned algorithm is used in the so-called "derivative-free optimization". The DE algorithm finds a minimum of a function $f(x): R^n \rightarrow R$, where it is hard to approximate the derivatives of an analytical function (which may be complex or non-derivative), or it is impossible to identify the analytical form of a target function to be optimized and therefore the derivatives cannot be computed easily.

The DE algorithm has the following steps (visual description is shown in Fig. 8):

- 1) Create a population with a N_p individuals. Each individual has a set of parameters x_i . Each parameter vector corresponds to an objective function. Select a target and base vector.
- 2) Randomly select two population members.
- 3) Compute a weighted difference vector from the previously picked two-parameter vectors.
- 4) Add a computed weighted difference vector to the base vector, thereby creating a mutant vector. Use the target vector and mutant vector for a crossover (trial vector).

- 5) Select from the two vectors, i.e., the target vector or trial vector (selection is based on the cost function).
- 6) Consequently, a new population is created.

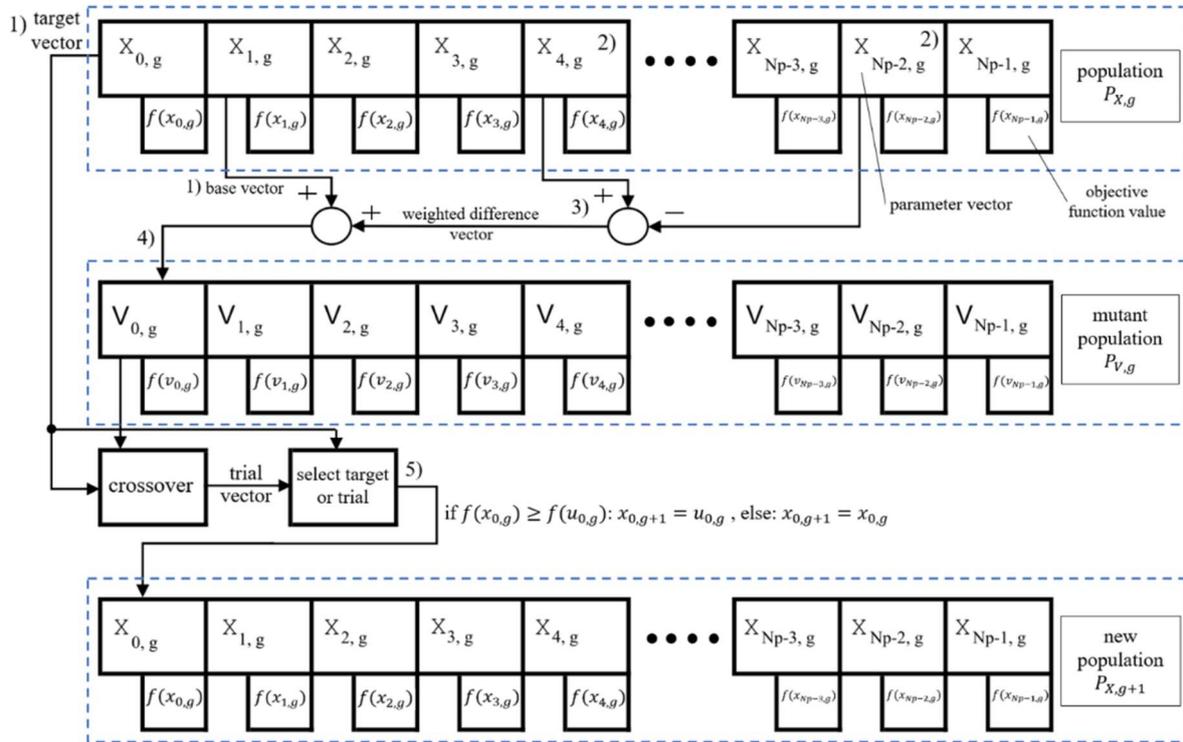


Figure 45. Differential Evolution (DE) algorithm scheme.

Where:

g – generation counter (increases every algorithm cycle)

$x_{(Np-i,g)}$ – parameter vector for the $Np - i, i \in \{0,1,2 \dots Np\}$ population member

$f(x_{(Np-i,g)})$ – objective function value for the parameter vector $x_{(Np-i,g)}$

$v_{(Np-i,g)}$ – parameter vector for the $Np - i, i \in \{0,1,2 \dots Np\}$ mutant population member

$f(v_{(Np-i,g)})$ – objective function value for the mutant parameter vector $v_{(Np-i,g)}$.

In the doctoral dissertation, the author used the DE algorithm to find the parameters of partial decomposed analytical functions expressed by equations (21)-(23), which, when summed, best reflect the currently studied transient response. The DE algorithm is the backbone of the MD3 method, described in more detail in section 5.3, to determine the parameters of the partial functions of the individual scenarios. The best scenario is determined based on the results of its calculations.

The MD3 method, and in particular the DE algorithm used to search for the parameters of decomposed functions, is the essence of the automatic identification part of the FDI system.

6. Architecture of the automated FDI system

This chapter presents the architecture of the proposed automated FDI system. Then, the flow of data necessary for detecting and identifying malfunctions will be described. Finally, the critical elements of the executive program are presented and explained.

6.1. System architecture

The architecture of monitoring and protection (M&P), and diagnostic systems (DS) may differ significantly depending on the production plant.

Figure 46 shows typical data flow in a power plant. Some plants may only have a M&P system, depicted in the Figure 46 by the red dashed rectangle. Others have expanded their systems architecture with a DS, presented by the black dashed rectangle in Figure 46.

In the case of power plants, the monitoring system collects, displays and saves data. The security system, which comes as a part of a monitoring system, watches over the device's safety. This layer operates without any human intervention and is able to automatically initiate the coast down procedure of the power generation unit. The diagnostic system is a separate layer, which analyzes the current state of the machine. It can perform advanced diagnostics procedures to facilitate an expert in diagnosing the machine's malfunctions.

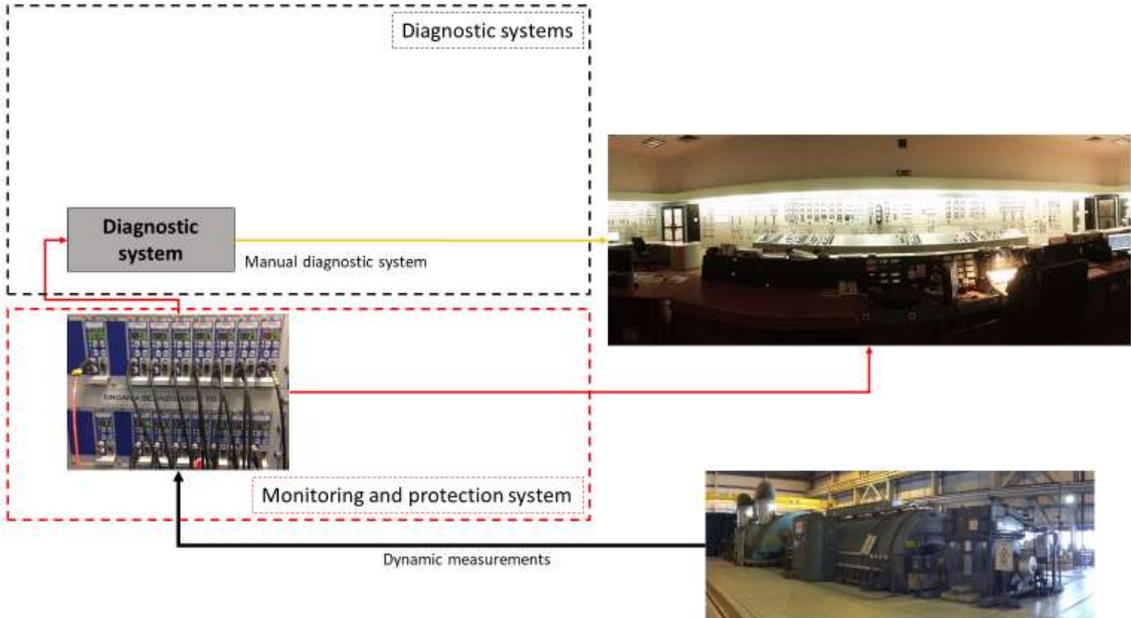


Figure 46. Basic monitoring and diagnostic system arrangement.

The data from the sensors go to the M&D system (black line in Figure 46). The monitoring system saves and displays the data collected from the object in the control room. The protection system uses the data collected by the monitoring system and compares them with the upfront threshold values. If the instantaneous values measured by the monitoring system exceed the threshold - the machine is turned off by the protection system. It is the most straightforward system, marked in Figure 46 with a red

dashed line. However, this approach does not guarantee conscious management of the machinery and the optimal organization of repairs. Having only the overall data and not processing the vibration data for diagnostic features, the personnel cannot determine the machine's technical condition correctly. This strategy often leads to unforeseen outages.

The DS is an extension added to the M&D system. It can extract diagnostic features necessary for the correct analysis and diagnostics of rotating machines. This system uses measurements provided by the M&D system. Raw data from the M&P system is passed to DS (red line in Figure 46), where appropriate tools are used to extract diagnostic features (yellow line in Figure 46). These characteristics can be computed in flight or a posteriori as the need arises. The diagnostic systems used in utility power plants do not work automatically. Unfortunately, this involves little checking of the technical condition of the device. Usually, an expert is only hired when the problem is so severe that the safe operation of the equipment is jeopardized. This approach leads to unnecessary and increased expenses and reduces the device's expected life.

The author during his research proposed, developed, and tested the automated FDI tool. The flowchart of the system is shown in Figure 47. It consists of two main parts - the OpEn module and the MD3 module. OpEn is a module for automatic failure detection, which uses the concept of the acceptance region wrapped around the modeled baseline.

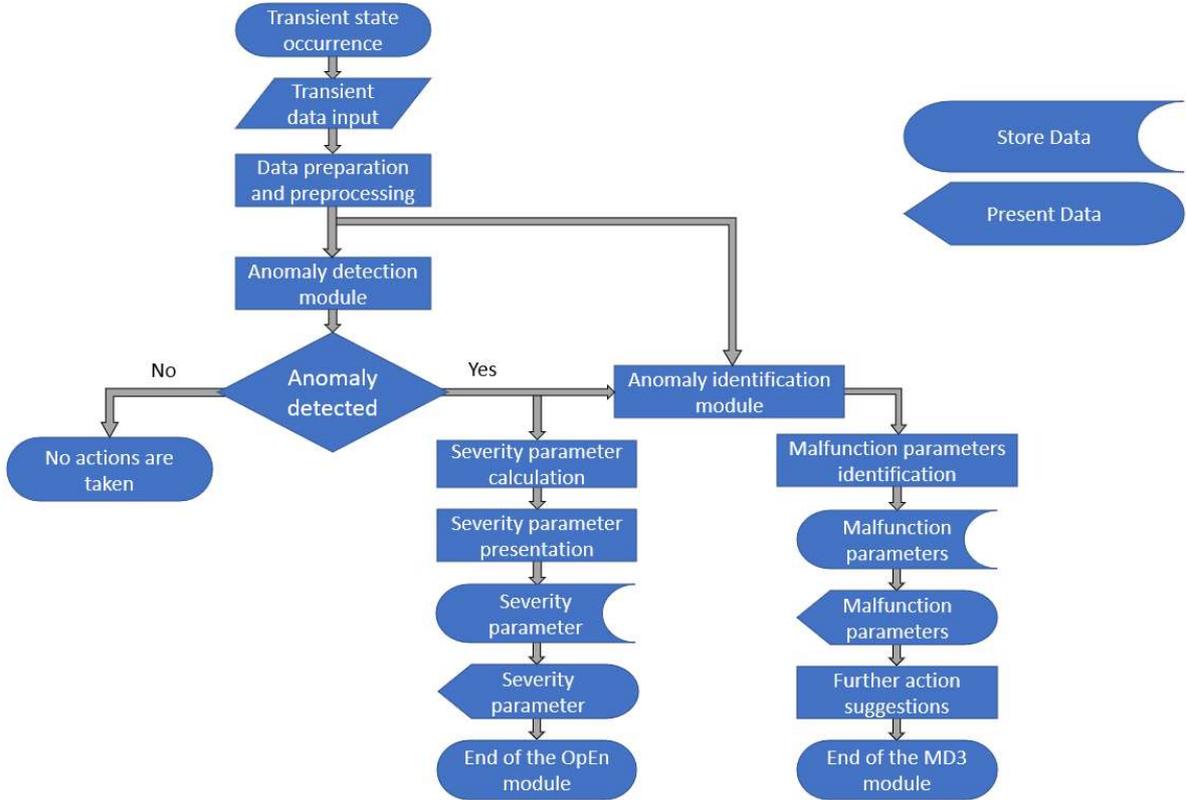


Figure 47. Flowchart of the proposed system.

The OpEn module runs in the background during each transition state. If the data in the transient does not exceed the acceptance area, the module does not perform actions visible to the operator. When transient data exceed acceptance values, the module calculates the severity parameters and passes them to the device operators. Then the transient data preprocessed in the OpEn module is passed to

the MD3 module. The MD3 module identifies basic malfunctions without expert intervention. The parameters of individual malfunctions are identified for the decomposed functions and optimized to fit the input transient best. Each of the decomposed functions of the MD3 module has its mechanical meaning.

Both the methods were implemented in the Python environment. The software tool was proposed, designed and validated. The complete system was tested on laboratory and real turbine data.

The automatic fault detection and identification (AFDI) system proposed by the author is part of the DS of power plant devices. Therefore, it can be a simple diagnostic and operational support system. Figure 48 depicts proposed structure and arrangement of M&D and DS equipped in AFD&I system in power plants.

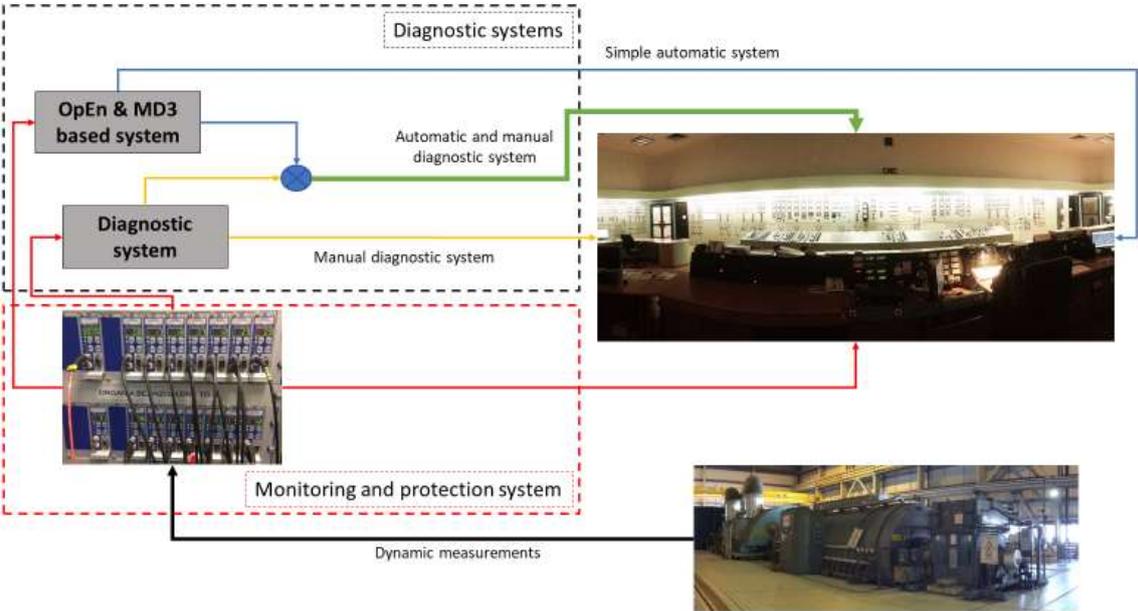


Figure 48. Monitoring-protection and diagnostic systems in power plants supported by the automated FDI system (OpEn & MD3 modules).

The AFD&I system can operate autonomously without the DS layer. In Figure 48, this is shown as a blue line. In this case, it will be the first line of supervision of the baseline condition of the machine. If properly implemented, it will detect the beginnings of the evolution of malfunctions, giving the owners time to prepare for possible repairs. For systems equipped with DS, the AFD&I system can become a support for diagnostic teams indicating the beginnings of the malfunctions evolution in an automated manner. This system will eliminate the need to constantly check the technical condition of all devices and limit the process of checking to those that actually show faults origins.

The system consists of two stages: first step is fault detection, described in Section 3 and the second step is fault identification, described in Section 4. The preparatory process includes activities such as loading reference transient data sets and processing them, leading to the determination of baseline measurements and OpEn values. The enforcement proceedings include activities related to data processing from a given transient state, an anomaly detection module, and a basic malfunction identification module. Loading data and CS interpolation modules are common to both paths.

In the following sections of the chapter, the individual modules of the proposed AFD&I system will be presented. The author used Python version 3.8 libraries and Google Collaboratory as the scripting environment to implement the system.

6.2. Algorithms for the detection and identification methods

In following paragraph, the main algorithms for the anomaly detection and malfunction identification are presented. First a pseudocode for the reference data import module is depicted. Depending on arbitrary chosen option for the OpEn method it returns matrices with baseline acceptable region values for the OpEn 2D and 3D case. Afterward, OpEn algorithms for current transient are laid out for 2D and 3D case, respectively. Next the MD3 method is explained and its pseudocode is presented. At the end of the chapter the python code for the most important function is given and explained.

6.2.1. Data import module

First pillar of the entire system is the data import module for the baseline values calculation presented in the Algorithm 1. This module retrieves the given transient data from the transient data sets given location. All set of transients are analyzed for one sensor individually. Each transient data set contains all vibration features provided by the measuring system – in this case the portable data acquisition unit. The module then cleans the data matrix by removing the rows with the same RPM values. Then it sorts the data in ascending order with respect to the RPM. The module is used for iterative data input into the algorithm.

Algorithm 1. Data Import and CS module for the baseline measurement

Input: sensor list, set of transient data in given location, OpEn 2D Upper Value, OpEn 2D Lower Value, OpEn 3D Amp Value, OpEn 3D Kph Value, mode (BULL/BEAR or both)

Result: CS interpolated transient data vibration values for all turboset probes in given location, OpEn 2D Baseline Upper and Lower Limits (OpEn 2D BULL) matrix, OpEn 3D Baseline Ellipsis Acceptance Region (OpEn 3D BEAR) matrix

Procedure:

for every probe in probe set:

for every transient in given transient location

 Open file

 Select vibration features applicable to given analysis and given probe

 Remove and clean duplicated data

 Sort data in ascending order (RPM-wise) all the vibration data and all of its features

 Probe transient vibration data matrix

 minimal and maximal RPM

 CS interpolation for all vibration feature

end (all transients for one probe)

 OpEn centerline matrix for one probe (arithmetic mean of non-zero values RPM-wise)

```

end (all transients for all probes in given location)

if mode = BULL
    OpEn centerline matrix + Upper value = OpEn 2D Upper Value matrix
    OpEn centerline matrix – Lower value = OpEn 2D Lower Value matrix
    from OpEn 2D Upper Value matrix and OpEn 2D Lower Value matrix create OpEn 2D BULL matrix
if mode = BEAR
    from OpEn 3D Amp Value, OpEn 3D Kph Value and OpEn centerline matrix create a OpEn 3D BEAR matrix
if mode = both
    Perform mode = BULL
    Perform module = BEAR
end

```

Module returns the acceptable region of OpEn matrices. Depending on the system capabilities it can produce OpEn 2D Baseline Upper and Lower Limit matrix (OpEn BULL) or OpEn 3D Baseline Ellipsis Acceptance Region matrix (OpEn BEAR), or both.

6.2.2. Detection method for the OpEn 2D and 2D case

Algorithm 2 and Algorithm 3 presents the process of collecting and processing the current transient to detect anomalies in the case of OpEn 2D and the OpEn 3D method, respectively. This is the second pillar of the system. In the beginning, each module takes matrices with reference data for itself. In the case of the OpEn 2D module, it is the OpEn 2D BULL matrix, visible in Algorithm 2, where each element contains two values of a given sensor at a given rotational speed. On the other hand, the OpEn 3D module uses the OpEn 3D BEAR matrix, shown in Algorithm 3, in which each matrix element is a set of ellipse coordinates for a given sensor at a given rotational speed. Then, each module opens the current file with the current transient from the indicated location. In the next step, vibration signal features are selected. In the case of OpEn 2D, only one vibration signal feature is selected, e.g., total vibration amplitude (Direct) or the amplitude of the first component (1X Amplitude). The OpEn 3D case analyzes the vibration vector as a whole, i.e., the combination of the amplitude and phase of the first component (1X Amplitude and 1X Phase).

Algorithm 2. Data Import and CS module for the current transient processing using OpEn 2D method.

```

Input      :      sensor list, transient data set, OpEn 2D BULL matrix
Result    :      Set of OpEn versus CS interpolated transient data values for all turboset probes in transient

```

```

Procedure :
    Upload OpEn BULL matrix for whole probe set
    Open actual transient data set file:
    Select vibration features applicable to given analysis

```

Remove and clean duplicated data

Sort data in ascending order (RPM-wise) all the vibration data and all of its features

Set common CS domain with OpEn 2D BULL matrix (equidesant spaced rotational speed increment)

for every probe in probe set:

 CS interpolation of given vibration feature

for each, common, rotational speed value in current transient

 Compare CS vector element with the respective OpEn 2D BULL matrix element

if OpEn Lower value \leq current data value **and** current data value \leq OpEn Upper value

 No actions are taken – behavior on this RPM is considerate as correct

Standby

else

 Set MD3 method triggered active (save probe name, RPM value)

 Apply RMSE procedure to the given transient data

 Apply Kurtosis procedure to the given transient data

 Apply MIN and MAX procedures to the given transient data

end

 Save the OpEn 2D severity parameters (MIN / MAX, RMSE, Kurtosis)

end

if MD3 trigger active == **True**

 Display the OpEn 2D severity parameters

 Go to MD3 method

else

Standby

The next step for both OpEn 2D and OpEn 3D, as in the case of Algorithm 1, is the cleaning and sorting of the data to perform the CS interpolation correctly. Then, the algorithm determines a mutual rotational speed domain for the actual transient data set and the baseline matrices. This operation is vital from the point of view of the principle of operation of the OpEn method. It enables a suitable comparison of the baseline values with the current transient. The mutual rotational speed comparison approach also makes it possible to compare baseline transients with those that are not fully registered – i.e., that do not start or end in a known machine state (turning or FSNL operation). The next step is to perform CS interpolation in order to obtain the signal features values in the predefined rotational speed domain with an interval identical to that in the OpEn BULL and OpEn BEAR matrix for the 2D and 3D case, respectively. Then, a nested loop follows. It iteratively takes each sensor and compares element by element of the current transient's value with the baseline values for all the elements in the common domain.

In the case of OpEn 2D, presented in Algorithm 2, the current transient value in a particular rotational speed instance is compared with the corresponding maximum and the minimum acceptable values. If the vector value goes beyond the assigned baseline interval (OpEn BULL) for any point in a given vector, the sensor data and its name is saved. The transient from the sensor and the severity parameters are calculated and saved. MD3 trigger is activated to switch to the MD3 method and identify the malfunction.

Algorithm 3. Data Import and CS module for the current transient processing using OpEn 3D method.

Input: CS data points matrix of probe set for the current transient, OpEn 3D BEAR matrix

Result: Anomaly detection and advisory for further actions or standby

Procedure :

Upload OpEn 3D baseline values for whole probe set

Open actual transient data set file

Select vibration features applicable to given analysis (1X Amplitude and 1X Phase)

Remove and clean duplicated data

Sort data in ascending order (RPM-wise) all the vibration data and all of its features

Set common CS domain with baseline measurements (equidistant spaced rotational speed increment)

CS interpolation of given vibration features

for every probe in probe set:

for each, common, rotational speed value in current transient

 Compare if the CS vector fell within the OpEn 3D BEAR)

if current data vector element fell within OpEn 3D BEAR element values

 No actions are taken – behavior in this RPM is considered as correct

else

 Distance from the ellipsis is calculated

 Set MD3 method triggered active (save probe name, RPM value)

 Save the severity parameters (distance from the ellipsis)

end

if MD3 trigger active == **True**

 Display and store distance from the Ellipsis

 Go to MD3 method

else

Standby

In the case of OpEn 3D, presented in Algorithm 3, the current transient vector (amplitude and phase) in a particular rotational speed instance is compared with the corresponding baseline ellipsis acceptable region (OpEn BEAR). If the vector falls out of the assigned baseline ellipsis for any point in a given vector, the sensor data and its name is saved. The transient from the sensor and the severity parameter as the distance from the ellipsis is calculated and saved. MD3 trigger is activated to which to the MD3 method and identify the malfunction.

At the end of the procedure, the MD3 trigger status is checked. If it is active, severity parameters are displayed and sensor names with values outside the baseline interval are passed to the MD3 method. If the MD3 trigger is inactive, the system remains in standby mode - no messages or values are displayed (the program remains in the background).

6.2.3. MD3 method algorithm

Third pillar is the malfunction identification module named MD3 method. Algorithm 4 pseudocode for the MD3 method. After the OpEn method has nominated the individual sensors for the identification procedure, the MD3 method is activated. In the beginning, it gets CS interpolation of previously processed transients from the indicated sensors. This shortens the time and saves the computational power required to complete the entire identification procedure. Then, three weight vectors for the different scenarios are created. The ranges of the weights of these vectors are predefined. The limit values search range of the weight vectors were determined in the accordance with author's knowledge and experience in the transient analysis.

Algorithm 4. MD3 method algorithm.

Input: CS data points matrix for probe names nominated by OpEn method in current transient (applicable vibration feature(s) for the analysis), number of evolutions (evol), population size (pop_size)

Result: Malfunction identification parameters

Procedure:

Upload CS data points matrix obtained in OpEn method

Generate a 6x1 weight vector with a set of random values (for scenario no. 1) – w1

Generate a 9x1 weight vector with a set of random values (for scenario no. 2) – w2

Generate a 12x1 weight vector with a set of random values (for scenario no. 3) – w3

for each probe indicated by MD3 trigger in OpEn method:

Establish the upper bound of the search for maximal number of identification parameters

Establish the lower bound of the search for maximal number of identification parameters

set number of evolutions

set number of population size

for every scenario[i]

Run DE algorithm for w[i] weights and with fitness_func[i] as a fitness function

Establish the best evolution number and individual based on RMSE norm

best solution = identification parameters = fittest individual

Add the best solutions to the scenario_weights[i] vector

end

Best scenario = min(RMSE(scenario_1), RMSE(scenario_2), RMSE(scenario_3))

display and save the best scenario number

display and save the best scenario weight vector parameters

display and save the best scenario RMSE value

plot a chart of the best scenario function versus CS of real data

end

In the further part, for each of the sensors indicated for identification, the algorithm determines the upper and lower limits of the search for identification parameters. Then, three scenarios are calculated using the established evolution quantity and population size parameters. Each scenario is based on parameter identification using the Differential Evolution (DE) algorithm. DE algorithm searches for weight vectors that best approximate a given objective function. The qualitative parameter describing the adjustment of the identified parameters to the objective function is the RMSE standard. The scenario with the smallest RESE value is nominated as the best representation of the malfunction transient. The best scenario weight vector parameters are saved. The CS transient of the nominated probe is plotted against the decomposed function described by the best scenario identification parameters.

6.2.4. Key Python functions

In the research during the doctoral dissertation, the author used Python 3.x and developed Google Colaboratory scripts for the research and data analysis. The latter allowed for the combination of functions written in the Python programming language and allowed access to data stored in the cloud. Table 5 lists the most important Python 3.x libraries that were used in the detection and identification system:

Table 5. Python libraries used in research.

```
import os
from google.colab import drive
from glob import glob
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from scipy.stats import norm, kurtosis
from scipy.interpolate import CubicSpline
import shapely
from shapely.geometry import Point, LineString, Polygon
from shapely.affinity import scale, rotate
```

```
import shapely.geometry as geom
from shapely.ops import nearest_points
```

Library `os`, `google.colab` and `glob` were used to handle the cloud resources such as transient data file loading and writing the baseline measurements. `Numpy` and `Pandas` libraries facilitated the faster data numerical operations. Large `SciPy` library and its sub-library such as `Interpolate.CubicSpline`, `Stat.Norm`, and `Stat.Kurtosis` and were used to perform cubic spline interpolation and calculate the severity parameters, respectively. Whole `Shapely` library together with its sub-libraries were used to process the `OpEn 3D` anomaly detection part of the system.

Table 6 presents the list of access paths to resources located in the cloud. First on the list is the access path to the baseline data calculated by the program and ready to be downloaded each time a new transient is analyzed. Then, the access path to the reference measurements is used to determine the baseline values. There are also various modes of malfunction coming from the machines studied by the author. Finally, multiple malfunction modes originate from the machines tested by the author with confirmed diagnoses. The last two paths on the list are data from the laboratory test rig. These data describe in a controlled manner simulation of the rotor unbalance. Validation of the results was conducted on the data described above was used both on the laboratory stand and on the data from real models tested by the author.

Table 6. Cloud resources used in the research (uploaded by the author from his portable data acquisition unit).

```
# google drive mount
drive.mount('/content/drive/')
# particular mode filepaths
bl_pth = "/content/drive/My Drive/Colab/Baseline_transient/"
ref_trans = "/content/drive/My Drive/Colab/Transient/Ref_trans/"
rub_trans = "/content/drive/My Drive/Colab/Transient/Rub_HP_Rear/"
pru_trans = "/content/drive/My Drive/Colab/Pru/"
mor_trans = "/content/drive/My Drive/Colab/Mor/"
koz_trans = "/content/drive/My Drive/Colab/Koz/"
rot_kit = "/content/drive/My Drive/Colab/Rot_Kit/"
rot_kit_2 = "/content/drive/My Drive/Colab/Rot_Kit_2/"
# dict combining both, names of modes and their acces paths to all
files
pth = {"base_1":bl_pth, "ref":ref_trans, "rub":rub_trans, "pru":pru_trans, "mor":
mor_trans, "koz": koz_trans, "rk": rot_kit, "rk2": rot_kit_2}
```

Table 7 lists the function that loads data from a given location. It is a universal module that selects signal features and cleans up and sorts data. The procedure is used when determining baseline measurements and for the `OpEn` method, both 2D and 3D.

Table 7. Function retriving and preprocessing the transient data set from given location.

```
def transient_data_load(f_path, probe_id):
    """
    Function retriving the data from location f_list
```

```

it preprocessed it in terms of removing doubled rotational speed
Sorts it by "Channel Name" (which is passed to it as probe_id)
and returns following dict with a name of transient and df with i
ts speed :
x - pre-processed rotational speed vector
y - 1XAmpl vector
kph - 1Xkeyphasor (Pase Lag vector of the first charmonic)
"""
use_data = pd.read_excel(f_path, header = 2)
use_data = use_data[["Channel Name", "Speed(P)", "1XAmplitude", "1X
Phase", "Direct"]].loc[use_data['Channel Name'] == "{}".format(probe
_id)]
use_data.drop_duplicates(subset ="Speed(P)", keep = False, inplac
e = True)
use_data.sort_values("Speed(P)", ascending = True, inplace = True
)
use_data.set_index("Speed(P)", inplace=True)
# use_data_name = all_files_pth["ref"][0].split(" ")[-
1].split(".")[0]
return use_data

```

An essential procedure from the point of view of comparing transients is the procedure presented in Table 8. The current rotational speed vector returns the ordered rotational speed vector. The output vector of the function is defined at identical instances of rotational speed and with the same equidistant interval. This procedure enables the correct domain determination and points of determining the amplitudes of CS interpolation for transients

Table 8. Lowest and highest RPM value in transient.

```

def x_start_stop(x, delta_RPM) :
    """
    Funkcja zwraca uporządkowany przedział max i min prędkości obroto
wej
dla danego przebiegu wejściowego x i różnicy obrotów delta_RPM
    """
    if x[0] - delta_RPM < 0:
        x_startowy = 0
    else:
        for i in range(delta_RPM):
            x_temp = x[0] + i
            if x_temp % delta_RPM == 0:
                x_startowy = x[0] + i
                break
        else: continue

    if x[-1] + delta_RPM >10000:
        x_koncowy = 10000

```

```

else:
    for i in range(delta_RPM):
        x_temp = x[-1] - i
        if x_temp % delta_RPM == 0:
            x_koncowy = x[-1] - i
            break
        else: continue
x0 = np.arange(x_startowy, x_koncowy+1, delta_RPM)
return x0

```

The procedure for creating a CS interpolation with data points set is presented in Table 9. The function uses the functions described in Table 7 and Table 8 to obtain data set points in specific rotational speed instances. In this case, the procedure returns two vectors consisting of CS interpolating the vibration amplitude and phase. This function is an example used in the OpEn 3D method. In this case, the RPM measurement is also accompanied by the amplitude and phase of the vibration signal.

Table 9. CS interpolation procedure.

```

def cs_amp_kph_df(trans_ptn, probe):

    # zaczytanie danych z konkretnego czujnika "probe_id"
    raw_trans_df = transient_data_load(trans_ptn, probe)
    # obroty w danym przebiegu
    x_raw = raw_trans_df.index
    # tworzenie operatora CS dla 1XAmplitude
    cs_y_oper = CubicSpline(x_raw, raw_trans_df["1XAmplitude"])
    # Dziedzina dla interpol CS
    x_cs_domain = x_start_stop(x_raw, 50)
    # interpolacja CS na dziedzinie x_cs_domain
    amp_cs = cs_y_oper(x_cs_domain)
    amp_cs_df = pd.DataFrame(data=amp_cs, index=x_cs_domain)
    amp_cs_df.columns = ["{}".format(probe)]
    # tworzenie operatora CS dla 1X Phase
    cs_kph_oper = CubicSpline(x_raw, raw_trans_df["1X Phase"])
    # interpolacja CS na dziedzinie x_cs_domain
    kph_cs = cs_kph_oper(x_cs_domain)
    kph_cs_df = pd.DataFrame(data=kph_cs, index=x_cs_domain)
    kph_cs_df.columns=["{}".format(probe)]

    return amp_cs_df, kph_cs_df

```

Searching and returning access paths to all files from a given location in the cloud function is presented in Table 10. The program goes through the selected folder with subfolders and returns a list of access paths as an iterable list. This procedure allows access to many files in an automated manner.

Table 10. Fncion returning list of all transient data set file names within the folder.

```

def file_root(cont_ptn):
    """

```

```

Function returns a list to all files within particular folder
"""
p_f = []
dir_files = []
# direct = []
for root, dirs, files in os.walk(r"{}".format(cont_ptn), topdown
= True):
    for name in files:
        p_f.append(os.path.join(root, name))
for i in range(len(p_f)):
    if p_f[i].split(".")[1] == "xlsx":
        dir_files.append(p_f[i])
    else:
        pass
return dir_files

```

Examples of the two partial functions for generating the decomposition function are shown in Table 11. They are used in the MD3 method described in section 5.2 by the formulas (21) and (22), respectively. The function called gauss is used to represent both the area of the critical revolutions of the machine and to model the non-linearity of the rotor response in situations of very large amplitudes. The function described as "parabola" is used to model the rotor responses related to unbalance. Due to the nature of the phenomenon, it was necessary to introduce the centrifugal force input threshold as an additional identification parameter.

Table 11. Parabola and Gaussian functions- the decomposed functions.

```

def gauss(x, amp, peak, wide):
    return amp*np.exp(-(x-peak)**2/(2*wide**2))

def parabola(x, a, x0):
    if x0<x:
        parabola = a*((x-x0)/(4200-x0))**2
    else: parabola = 0
    return parabola

```

Table 12 show the three scenarios for decomposed functions. The first function consists of only three partial functions: term bias described by equation (23), parabola described by equation (22), and one gaussian function (21). The function model described in this way has six parameters that the DE algorithm should identify. The second and third models are an extension of the most straightforward model, and they have two and three gauss functions, respectively. Nine and twelve parameters describe and identify the second and third MD3 method scenarios.

Table 12. Three models used in MD3 method "tournament".

```

def fmodel_1(x,w):
    y = []
    for i in range(len(x)):

```

```

    y.append(gauss(x[i], w[0], w[1], w[2]) + parabola(x[i],w[9],w[
10]) +w[11])
    return y

def fmodel_2(x,w):
    y = []
    for i in range(len(x)):
        y.append(gauss(x[i], w[0], w[1], w[2]) + gauss(x[i], w[3], w[4]
, w[5]) + parabola(x[i],w[9],w[10]) +w[11])
    return y

def fmodel_3(x,w):
    y = []
    for i in range(len(x)):
        y.append(gauss(x[i], w[0], w[1], w[2]) + gauss(x[i], w[3], w[4]
, w[5]) + gauss(x[i], w[6], w[7], w[8]) + parabola(x[i],w[9],w[10])
+w[11])
    return y

```

Table 13 shows the function's implementation for assessing the fitness of the particular model parameters to the current transient function. This function is the norm applied for every function from the MD3 method scenario. The outcome of this procedure is the fitness parameter which determines the best scenario function, i.e., parameters describe the current transient.

Table 13. The MD3 method the fitness function quality parameter.

```

def rmse(w):
    y_pred = fmodel(x, w)
    return np.sqrt(sum((y - y_pred)**2) / len(y))

```

The heuristic limits of the search for the parameter values of partial functions are presented in Table 14. These values are determined anew for each transient. However, all the scenarios used to identify the current failure have the same search values.

Table 14. Upper and lower search limits boundaries for the DE algorithm.

```

lb = [min(y), x[0], 10, min(y), 1/3*x[-1], 10, min(y), 2/3*x[-1],
10, 0, -x[-1], 0]
ub = [max(y), 1/2*x[-1], x[-1]/5, max(y), 3/4*x[-1], x[-
1]/5, max(y), x[-1], x[-1]/5, max(y), x[-1], max(y)/4]
bounds=[(lb[i], ub[i]) for i in range(len(lb))]

```

Table 15 shows the functions used in the system for creating baseline measurements from reference transients. The detailed process is described in section 4.2.

Table 15. Baseline measurements functions.

```

def probe_baseline(cold_stt_trans, probe_id, whole_RPM, prb):

```

```

x_cs_domain, y_cs, kph_cs, curr_trans_name, whole_ref_set_1xamp,
whole_ref_set_1xkph = trans_CS(cold_stt_trans, probe_id, whole_RPM,
prb)
x_stat = np.zeros(len(whole_RPM))
y_stat = np.zeros(len(whole_RPM))
whole_ref_set_x = np.zeros((len(cold_stt_trans),len(whole_RPM)))
whole_ref_set_y = np.zeros((len(cold_stt_trans),len(whole_RPM)))

for trans in range(len(cold_stt_trans)):
    # print("trans no = ",trans)
    for RPM in range(len(whole_RPM)):
        whole_ref_set_x[trans][RPM] = whole_ref_set_1xamp[trans][RPM]
* np.cos(np.deg2rad(whole_ref_set_1xkph[trans][RPM]))
        whole_ref_set_y[trans][RPM] = whole_ref_set_1xamp[trans][RPM]
* np.sin(np.deg2rad(whole_ref_set_1xkph[trans][RPM]))
        # print("RPM no = ",RPM)

x_stat = np.sum(whole_ref_set_x,axis=0)/np.count_nonzero(whole_ref_set_x,axis=0)
y_stat = np.sum(whole_ref_set_y,axis=0)/np.count_nonzero(whole_ref_set_y,axis=0)
kph_stat = np.rad2deg(np.arctan2(y_stat,x_stat))
r_stat = np.sqrt(x_stat**2 + y_stat**2)

for RPM in range(len(kph_stat)):
    if kph_stat[RPM] < 0:
        kph_stat[RPM] = kph_stat[RPM] + 360
return x_stat, y_stat, kph_stat, r_stat, whole_ref_set_1xamp, whole_ref_set_1xkph

def trans_CS(cold_stt_trans, probe_id, whole_RPM, probe):

    # tworzenie pustego zbioru dla wszystkich przebiegów z danego kata logu
    whole_ref_set_1xamp = np.zeros((len(whole_RPM),len(cold_stt_trans))).T
    whole_ref_set_1xkph = np.zeros((len(whole_RPM),len(cold_stt_trans))).T

    # wyliczam CS dla U3 - stan zimny oraz dla probe_id[2]
    # t_x_y = []
    x_raw = []
    y_raw = []
    kph_raw = []
    x_cs_domain = []
    # y_qbc_spln_oper = []
    y_cs = []
    # kph_qbc_spln_oper = []
    kph_cs = []

```

```

curr_trans_name = []
for i in range(len(cold_stt_trans)):
    # t_x_y.append(("trans_{}", x_raw_ {}, y_raw {}, kph_raw_ {}, x_cs_
domain_ {}, y_qbc_spln_oper_ {}, y_cs_ {}, kph_qbc_spln_oper_ {}, kph_c
s, curr_trans_name_ {}".format(i,i,i,i,i,i,i,i,i,i,cold_stt_trans[i]))
.split(", "))

    # dane z poszczególnych przebiegów
    # trans_ {}
    _, temp_x_raw, temp_y_raw, temp_kph_raw = transient_data_load(c
old_stt_trans[i], probe_id[probe])

    # x_cs_domain_ {}
    x_cs_domain.append(x_start_stop(temp_x_raw, 50))

    # tworzenie operatora cubic spline (CS) dla y_raw w punktach x_
raw
    # y_qbc_spln_oper_ {}
    cs_y_oper = CubicSpline(temp_x_raw, temp_y_raw)
    # tworzenie wartości funkcji CS dla punktów w miejscach x_cs_do
main_ (t_x_y[i][2]) i ampl. y_raw
    # y_cs_ {}
    y_cs_temp = cs_y_oper(x_cs_domain[i])
    y_cs_temp[0] = y_cs_temp[1]
    y_cs_temp[-1] = y_cs_temp[-2]
    y_cs.append(y_cs_temp)

    # tworzenie operatora cubic spline (CS) dla kph_raw w punktach
x_raw
    # kph_qbc_spln_oper_ {}
    cs_kph_oper = CubicSpline(temp_x_raw, temp_kph_raw)
    # tworzenie wartości funkcji CS dla punktów w miejscach x_cs_do
main_ (t_x_y[i][2])
    # kph_cs
    kph_cs_temp = cs_kph_oper(x_cs_domain[i])
    kph_cs_temp[0] = kph_cs_temp[1]
    kph_cs_temp[-1] = kph_cs_temp[-2]
    kph_cs.append(kph_cs_temp)

    # curr_trans_name_ {}
    curr_trans_name = "{}".format(cold_stt_trans[i])
    # pdb.set_trace()
    indx_RPM = np.where(whole_RPM == x_cs_domain[i][0])[0][0]

    # Przypisywanie wartości do poszczególnych indeksów
    # odpowiadającej konkretnej prędkości obrotowej
    for k in range(len(x_cs_domain[i])):
        whole_ref_set_lxamp[i][indx_RPM + k] = y_cs[i][k]

```

```

whole_ref_set_lxkph[i][indx_RPM + k] = kph_cs[i][k]
return x_cs_domain, y_cs, kph_cs, curr_trans_name, whole_ref_set_
lxamp, whole_ref_set_lxkph

```

Table 16 shows the calculation of baseline measurement values and the behavior of the calculation results at a specific location on the virtual disk for a single sensor. The author used the PyDrive, google.colab, and oauth2client.client libraries from Google Colaboratory resources to carry out these operations. Performing the procedure n times, where n corresponds to the number of sensors or sensors included in the data point set matrix, allows obtaining baseline data for the entire shaftline.

Table 16. Baseline measurements procedure: calculating and saving in the cloud.

```

probe_no_x_stat = []
probe_no_y_stat = []
probe_no_kph_stat = []
probe_no_r_stat = []
for prb in range(len(probe_id)):
    x_stat_temp, y_stat_temp, kph_stat_temp, r_stat_temp, whole_ref_s
et_lxamp, whole_ref_set_lxkph = probe_baseline(cold_stt_trans, prob
e_id, whole_RPM, prb)
    probe_no_x_stat.append(x_stat_temp)
    probe_no_y_stat.append(y_stat_temp)
    probe_no_kph_stat.append(kph_stat_temp)
    probe_no_r_stat.append(r_stat_temp)
    print("probe no: ", prb)

!pip install -U -q PyDrive
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from google.colab import files
from oauth2client.client import GoogleCredentials

auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)
stat_CS_r = pd.DataFrame(data=probe_no_r_stat, index=probe_id, colu
mns=whole_RPM)
stat_CS_kph = pd.DataFrame(data=probe_no_kph_stat, index=probe_id,
columns=whole_RPM)
stat_CS_x = pd.DataFrame(data=probe_no_x_stat, index=probe_id, colu
mns=whole_RPM)
stat_CS_y = pd.DataFrame(data=probe_no_y_stat, index=probe_id, colu
mns=whole_RPM)
stat_CS_r.to_csv('stat_CS_r.csv', sep='\t')
files.download('stat_CS_r.csv')
stat_CS_kph.to_csv('stat_CS_kph.csv', sep='\t')
files.download('stat_CS_kph.csv')

```

```

stat_CS_x.to_csv('stat_CS_x.csv', sep='\t')
files.download('stat_CS_x.csv')
stat_CS_y.to_csv('stat_CS_y.csv', sep='\t')
files.download('stat_CS_y.csv')

```

Table 17 presents the procedure for downloading baseline data from virtual resources and creating baseline ellipses, which they will use to detect anomalies for the OpEn 3D case. First, baseline data is read from the cloud and entered into an appropriately named matrix instance. Next, the vibration feature data is transformed into the form of a polar and Cartesian coordinate system. The name of the matrix's rows is multi-level, which easily defines the properties of a specific sensor.

Table 17. Baseline data load (for the OpEn 3D case) and creating baseline ellipsis coordinates.

```

# dict with all filepaths
# key - specific data
# value - root of all data in the folder
all_files_pth = {}
for key in pth:
    temp = {key : file_root(pth["{}".format(key)])}
    all_files_pth.update(temp)
# -----
# CS domain - in which we will compare our results
# (fix rotational speed increment - delta = 50RPM)
# and a all probes across the system
whole_RPM = np.arange(0,3001,50)
probe_id = ['1Y', '1X', '2Y', '2X', '3Y', '3X', '4Y', '4X', '5Y', '
5X', '6Y', '6X', '7Y', '7X']
# -----
# OpEn acceptance region
OpEn_amp, OpEn_kph = 30, 25
# -----
# dict contains all baseline measurements
baseline_data = {}
for i in range(len(all_files_pth["base_1"])):
    bl_data_temp, bl_name_temp = baseline_data_load(all_files_pth["ba
se_1"][i])
    bl_temp = {bl_name_temp.split("_")[-1]: bl_data_temp}
    baseline_data.update(bl_temp)
# -----
# df for all baseline data - name of df is its property
x_df = pd.DataFrame(data=baseline_data["x"], index=probe_id, column
s=whole_RPM)
y_df = pd.DataFrame(data=baseline_data["y"], index=probe_id, column
s=whole_RPM)

```

```

r_df = pd.DataFrame(data=baseline_data["r"], index=probe_id, columns=whole_RPM)
kph_ = pd.DataFrame(data=baseline_data["kph"], index=probe_id, columns=whole_RPM)
# -----
# -----
# df as a matrix of a vectors with ellipsis coordinates for each sensor and each RPM
fy = pd.DataFrame(data=None, index=probe_id, columns=whole_RPM)
fx = pd.DataFrame(data=None, index=probe_id, columns=whole_RPM)
# populating df with a ellipsis co-ordinates
for i in range(0,baseline_data["x"].shape[0]):
    for k in range(0,baseline_data["x"].shape[1]):
        x = baseline_data["x"][baseline_data["x"].columns[k]][baseline_data["x"].index[i]]
        y = baseline_data["y"][baseline_data["y"].columns[k]][baseline_data["y"].index[i]]
        f_x,f_y = x_y_ell(x,y,OpEn_amp,OpEn_kph)
        fx[fx.columns[k]][fy.index[i]] = f_x
        fy[fy.columns[k]][fy.index[i]] = f_y
        # fx[nazwa_czujnika][prędkość_obrotowa], czyli fx[0][0], pierwszy czujnik (1Y) dla pędk. obr = 0
        # fy[nazwa_czujnika][prędkość_obrotowa]
        #
# -----
# -----

```

The OpEn procedure for the 2D case is presented in Table 18:

1. The current transient data points set is retrieved from the file with the specified location.
2. Loaded data are preprocessed and prepared for CS interpolation.
3. CS interpolation follows. Data from every probe in the current transient are compared with their OpEn BULL values. If any point exceeds the OpEn BULL matrix values assigned to them - severity parameters are calculated, and MD3 mode is set to 1.
4. The corresponding graphs and severity parameters values are presented.

Table 18. OpEn procedure with severity parameter calculation and charts plotting (for the OpEn 2D case).

```

cold = np.arange(0,12)
hot = np.arange(12, len(ref_trans_U3))
cold_stt_trans = [ref_trans_U3[i] for i in cold]
hot_stt_trans = [ref_trans_U3[i] for i in hot]

# wyliczam CS dla U3 - stan zimny
t_x_y = []
for i in range(len(cold_stt_trans)):

```

```

t_x_y.append(("trans_{}", x_{}_y_{}_x_cube_{}_y_spline_temp_{}_
y_cube_{{}}, {{}}".format(i,i,i,i,i,i,cold_stt_trans[i])).split(",")
# dane z poszczególnych przebiegów
t_x_y[i][0], t_x_y[i][1], t_x_y[i][2] = transient_data_load(cold
stt_trans[i])
t_x_y[i][3] = x_start_stop(t_x_y[i][1], 50)
# tworzenie operatora cubic spline (CS)
# -----
-----

t_x_y[i][4] = CubicSpline(t_x_y[i][1], t_x_y[i][0]["1XAmplitude"]
)
# -----
-----

# tworzenie wartości funkcji CS dla punktów w miejscach x_cube
t_x_y[i][5] = t_x_y[i][4](t_x_y[i][3])
t_x_y[i][5][0] = t_x_y[i][5][1]
t_x_y[i][5][-1] = t_x_y[i][5][-2]

ub = 30
lb = 15
whole_RPM = np.arange(0,3001,50)
whole_ref_set = np.zeros((len(whole_RPM),len(cold_stt_trans))).T
for i in range(len(cold_stt_trans)):
    # wspólne obroty
    indx_RPM = np.where(whole_RPM == t_x_y[i][3][0])[0][0]
    for k in range(len(t_x_y[i][3])):
        whole_ref_set[i][indx_RPM + k] = t_x_y[i][5][k]
Whole_ref = pd.DataFrame(whole_ref_set.T)

# pozbycie się wartości brzegowych - gdzie interpolacja nie zachowu
je się stabilnie
Whole_ref.set_index(whole_RPM,inplace=True)

# Nazwanie kolumn wg ich nazwy pliku
cols = []
for i in range(len(cold_stt_trans)):
    cols.append(cold_stt_trans[i].split("/")[-1])
Whole_ref.columns = cols

RMSE = []
KURT = []
MIN_MAX_MAX = []
MIN_MAX_MIN = []

for iter in range(len(cold_stt_trans)):
    # x_Yref_Y_trans = y_costs(t_x_y[iter][3], t_x_y[iter][5], whole
RPM[1:-1], stat_ref)

```

```

RMSE.append([rmse(y_costs(t_x_y[iter][3], t_x_y[iter][5], whole_RPM,
stat_ref)[: ,2] - y_costs(t_x_y[iter][3], t_x_y[iter][5], whole_RPM,
stat_ref)[: ,1])])
KURT.append([kurt(y_costs(t_x_y[iter][3], t_x_y[iter][5], whole_RPM,
stat_ref)[: ,2] - y_costs(t_x_y[iter][3], t_x_y[iter][5], whole_RPM,
stat_ref)[: ,1])])
_, max_max, _, min_min = min_max(y_costs(t_x_y[iter][3], t_x_y[iter][5],
whole_RPM, stat_ref), ub, lb)
# ub , czyli ub% wyżej od amplitudy w danym punkcie delt_RPM,
# lb = lb% niżej od amplitudy w danym punkcie delt_RPM
MIN_MAX_MAX.append(max_max[0])
MIN_MAX_MIN.append(min_min[0])

RMSE = np.array(RMSE)
KURT = np.array(KURT)
MIN_MAX_MAX = np.array(MIN_MAX_MAX)
MIN_MAX_MIN = np.array(MIN_MAX_MIN)

MD3_mode = 1

fil_name = []
fil_name.append(cold_stt_trans[iter].split("/")[-1].split(".")[0].split(" ")[-1])

fig, axes = plt.subplots(figsize=(30,15))
for i in range(0, len(t_x_y)):
    axes.scatter(t_x_y[i][3], t_x_y[i][5], label = "{}".format(cold_stt_trans[i].split(" ")[-1].split(".")[0]))
    axes.plot(whole_RPM, stat_ref_ub, color='red', marker='o', linestyle='dashed', label="OoB_Upper")
    axes.plot(whole_RPM, stat_ref_lb, color='blue', marker='o', linestyle='dashed', label="OoB_Lower")
    axes.plot(whole_RPM, stat_ref, color='black', marker = "o", label = "centre of CS")
    axes.set_title("CS for synchronous response amplitude (the shaft relative vibration in bearing #1, 45° left from the vertical axis)", fontsize=25)
    axes.set_xlabel("RPM (revolution per minute)", fontsize=25) # Notice the use of set_ to begin methods
    axes.set_ylabel(r'Amplitude in {}m peak-peak [0,360]'.format(r"$\mu$"), fontsize=25)
    axes.set_ylim([-0.1, 200])
    axes.grid(color='b', alpha=0.5, linestyle='dashed', linewidth=0.5)
    axes.tick_params(labelcolor='k', labelsizes='large', width=10)
    axes.tick_params(axis='x', labelsizes=20)
    axes.tick_params(axis='y', labelsizes=20)
    plt.legend(fontsize=20)
plt.legend()

```

Table 19 shows the necessary procedures and functions for the OpEn 3D method. For example, the `cs_amp_kph_df` function returns the CS vibration feature interpolation of a synchronous component, and the `x_y_ell` function creates the coordinates of an ellipse rotated by a suitable angle, i.e., baseline ellipse coordinates.

Table 19. OpEn 3D method functions.

```
def cs_amp_kph_df(trans_pth, probe):

    # zaczytanie danych z konkretnego czujnika "probe_id"
    raw_trans_df = transient_data_load(trans_pth, probe)
    # obroty w danym przebiegu
    x_raw = raw_trans_df.index
    # tworzenie operatora CS dla 1XAmplitude
    cs_y_oper = CubicSpline(x_raw, raw_trans_df["1XAmplitude"])
    # Dziedzina dla interpol CS
    x_cs_domain = x_start_stop(x_raw, 50)
    # pdb.set_trace()
    # interpolacja CS na dziedzinie x_cs_domain
    amp_cs = cs_y_oper(x_cs_domain)
    amp_cs_df = pd.DataFrame(data=amp_cs, index=x_cs_domain)
    amp_cs_df.columns = ["{}".format(probe)]

    # amp_cs_df = amp_cs_df.T
    # tworzenie operatora CS dla 1X Phase
    cs_kph_oper = CubicSpline(x_raw, raw_trans_df["1X Phase"])
    # interpolacja CS na dziedzinie x_cs_domain
    kph_cs = cs_kph_oper(x_cs_domain)
    kph_cs_df = pd.DataFrame(data=kph_cs, index=x_cs_domain)
    kph_cs_df.columns=["{}".format(probe)]
    # pdb.set_trace()
    return amp_cs_df, kph_cs_df

def x_y_ell(x, y, OpEn_amp, OpEn_kph):
    # pdb.set_trace()
    circle = Point(x, y).buffer(1) # type(circle)=polygon
    ellipse = shapely.affinity.scale(circle, OpEn_amp, OpEn_kph)
    ellipse = rotate(ellipse, np.arctan2(y,x), origin=(x,y), use_radians=True)
    fx, fy = ellipse.exterior.xy
    return fx, fy
```

The entire process of the OpEn 3D method is presented in Table 20. First, the procedure retrieves the previously calculated baseline ellipse coordinates and CS interpolated points from the current transient. Then, the position of the point relative to the baseline ellipsis is determined for the given sensor in the particular RPM instance. The loop is repeated for all common RPM values contained in the current transient and for all measurements' sensors. Finally, the severity parameters are computed

if any points go beyond the OpEn 3D BEAR. In the end, both the severity parameters and the corresponding drawings are presented.

Table 20. OpEn procedure with severity parameter calculation and charts plotting (for the OpEn 3D case).

```

dist_list = []
RPM_list = []
probe = 12
amp_trans = []

for i in range(len(amp_all_probes_koz_df_T.columns)):

    RPM = trans_x_all.xls(f"{probe_id[probe]}", level="Probe ID").columns[i]
    # print("RPM: ",RPM)
    y_RPM_bl = y_df.loc[f"{probe_id[probe]}"][RPM]
    x_RPM_bl = x_df.loc[f"{probe_id[probe]}"][RPM]

    x_tr = koz_x_all.xls(f"{probe_id[probe]}", level="Probe ID")[RPM][0]
    y_tr = koz_y_all.xls(f"{probe_id[probe]}", level="Probe ID")[RPM][0]
    amp_trans.append(np.sqrt(x_tr**2 + y_tr**2))
    trans_piont = geom.Point([x_tr,y_tr])

    circle = Point(x_RPM_bl, y_RPM_bl).buffer(1) # type(circle)=polygon
    ellipse = shapely.affinity.scale(circle, OpEn_amp, OpEn_kph)
    ellipse = rotate(ellipse, np.arctan2(y_RPM_bl,x_RPM_bl), origin=(x_RPM_bl,y_RPM_bl), use_radians=True)
    fx, fy = ellipse.exterior.xy

    if ellipse.contains(trans_piont):
        dist = 0
        cont = {f"@{RPM} point within ell":dist}
        dist_list.append(0)
        RPM_list.append(RPM)

    else:
        dist = trans_piont.distance(ellipse)
        dist_list.append(dist)
        RPM_list.append(RPM)
    # ax.scatter(x_tr,y_tr, RPM, zdir="y", c="r")

fig,ax = plt.subplots(figsize = (30,30))
ax.set_title(f"{probe_id[probe]} Synchronous response amplitude (1X-Amp) and phase lag (1X-Phase) at {RPM}RPM",fontsize=30)
# ax.set_ylabel("RPM (revolution per minute)",fontsize=25, labelpad=30) # Notice the use of set_ to begin methods

```

```

ax.set_xlabel(r'Displacement in  $\mu\text{m}$  peak-
peak',fontsize=25, labelpad=20)
ax.set_ylabel(r'Displacement in  $\mu\text{m}$  peak-
peak',fontsize=25, labelpad=20)
ax.tick_params(labelcolor='k', labelsizе='x-large', width=10)
plt.rc('xtick',labelsizе=20)
plt.rc('ytick',labelsizе=20)
ax.set_xlim([-300,300])
ax.set_ylim([-300,300])
ax.scatter(koz_x_all.xs(f"{probe_id[probe]}", level="Probe ID") [R
PM] [0], koz_y_all.xs(f"{probe_id[probe]}", level="Probe ID") [RPM] [0
],c="r",
            label=f"Distance form OpEn ellipsis: {round(dist_list[
-1], 4)}")
ax.scatter(trans_x_all.xs(f"{probe_id[probe]}", level="Probe ID")
[RPM] [0], trans_y_all.xs(f"{probe_id[probe]}", level="Probe ID") [RPM]
[0],c="b",
            label=f"Reference transient sample")
ax.scatter(x_df.loc[f"{probe_id[probe]}"] [RPM], y_df.loc[f"{probe
_id[probe]}"] [RPM], c="k", label="Centre of an OpEn Ellipsis")
ax.scatter(0,0, label="Point of origin (0,0)", c="g")
ax.plot (fx,fy, label=f"OpEn Ellipsis")
plt.legend(fontsize=25)

```

The author proposed the MD3 method as a procedure implemented to find the best set of parameters for identifying a given malfunction. The sample of the code implementing the strategy detailed in section 5.2 and 5.4 is presented in Table 21. The procedure uses the data point sets previously processed in the OpEn method. Utilizing The data interpolated beforehand reduces the time and computational requirements of the proposed method of identifying and selecting identified parameters and thus the entire system. Table 21 shows an example of applying the MD3 method to simulated data from the test rig. The technique is capable of accepting different types and arrangements of sensors. The previous OpEn method nominates sensors taken for further analysis. First, data points set with rotor unbalance are read after removal of the malfunction. The vector of identification parameters is created as a vector of weights for the objective function. Then the vectors with an unbalanced rotor and after balancing are determined. A mutual speed domain is defined for both transients. The next step is to determine the extreme values of the search for the objective function identification parameters. After that, the procedure of finding parameters of the decomposed functions is performed for three separate scenarios. Finally, the best parameter set is selected to approximate the function decomposed to the transient based on the RMSE standard described in Table 13. The results of the MD3 method are displayed and saved. First, the number of the best scenario and its RMSE is given. Then, the set of parameters that identify a given scenario is used to determine the type of failure.

Table 21. MD3 method procedure example.

```
# dict with all filepaths
```

```

# key - specific data
# value - root of all data in the folder
all_files_pth = {}
for key in pth:
    temp = {key : file_root(pth["{}".format(key)])}
    all_files_pth.update(temp)

# -----
# CS domain - which the results will be compared
# (fix rotational speed increment - delta = 50RPM)
# and a all probes across the system
whole_RPM = np.arange(0,3001,50)
probe_id = ['1Y', '1X', '2Y', '2X', '3Y', '3X', '4Y', '4X', '5Y', '5X', '6Y', '6X', '7Y', '7X']
probe_id_koz = ['6Y', '6X', '7Y', '7X']
probe_id_rk = ['1Y', '1X', '2Y', '2X']
trans_pth = sorted(all_files_pth["rk2"])
all_amp_rk2 = pd.DataFrame(data=None)
all_kph_rk2 = pd.DataFrame(data=None)
amp_all_probes_rk2_df = pd.DataFrame(data=None)
kph_all_probes_rk2_df = pd.DataFrame(data=None)
file_name = []
probe_name = []
for tr_no in range(len(all_files_pth["rk2"])):

    # tworzenie df ze wszystkimi czujnikami z danego przebiegu
    for i in range(len(probe_id_rk)):
        amp_cs_df_rk2, kph_cs_df_rk2 = cs_amp_kph_df(trans_pth[tr_no],
        probe_id_rk[i])
        amp_all_probes_rk2_df = pd.concat([amp_all_probes_rk2_df, amp_cs_df_rk2], axis=1)
        kph_all_probes_rk2_df = pd.concat([kph_all_probes_rk2_df, kph_cs_df_rk2], axis=1)
        file_name.append(all_files_pth["rk2"][tr_no].split("/")[-2:][-1].split(".")[0])
        probe_name.append(probe_id_rk[i])
        # pdb.set_trace()

    amp_all_probes_rk2_df_T = amp_all_probes_rk2_df.T
    kph_all_probes_rk2_df_T = kph_all_probes_rk2_df.T

hier_index = list(zip(file_name, probe_name))
hier_index = pd.MultiIndex.from_tuples(hier_index)
# nadawanie dwuwymiarowego indeksu dla przejrzystości danych
amp_all_probes_rk2_df_T = amp_all_probes_rk2_df_T.set_index(hier_index)
amp_all_probes_rk2_df_T.index.names = ["Transient", "Probe ID"]
kph_all_probes_rk2_df_T = kph_all_probes_rk2_df_T.set_index(hier_index)

```

```

kph_all_probes_rk2_df_T.index.names = ["Transient", "Probe ID"]
amp_all_probes_rk2_df_T.loc["RK_bal"][amp_all_probes_rk2_df_T.columns[-
1]] = amp_all_probes_rk2_df_T.loc["RK_bal"][amp_all_probes_rk2_df_T
.columns[-2]]
# # tworzenie wsp. kartezyjskich
x_rk2_all = amp_all_probes_rk2_df_T[amp_all_probes_rk2_df_T.columns
[:-
1]]*np.cos(np.deg2rad(kph_all_probes_rk2_df_T[kph_all_probes_rk2_df
_T.columns[:-1]]))
y_rk2_all = amp_all_probes_rk2_df_T[amp_all_probes_rk2_df_T.columns
[:-
1]]*np.sin(np.deg2rad(kph_all_probes_rk2_df_T[kph_all_probes_rk2_df
_T.columns[:-1]]))

# # lista wszystkich transientów z folderu, które mam zebrane w df
unq_list = pd.DataFrame(list(zip(*kph_all_probes_rk2_df_T.index))).
loc[0].unique()

# dostęp do danych kolumn po indeksach dwupoziomowych > df.xs("nazw
a indeksu z poziomu II" , level="nazwa kolumny indeksu poziomu II")
# czyli np kph_all_probes_df_T

wek = []
for i in range(0,12):
    wek.append(f"w[{i}]")
weights_tabular = pd.DataFrame(data=None, columns=wek)

y_imb = np.array(amp_all_probes_rk2_df_T.xs("1X", level="Probe ID")
.loc["RK_imb"], dtype=float)
y_bal = np.array(amp_all_probes_rk2_df_T.xs("1X", level="Probe ID")
.loc["RK_bal"], dtype=float)
# y = np.array(amp_all_probes_rk2_df_T.xs("1Y", level="Probe ID").l
oc["RK_imb"], dtype=float)
# w = [amp_g1, peak_g1, wide_g1, amp_g2, peak_g2, wide_g2, amp_g3,
peak_g3, wide_g3, par_amp, par_start, bias_term]

x = np.array(amp_all_probes_rk2_df_T.columns, dtype=float)
y = y_imb

lb = [min(y),          x[0],          10, min(y),          1/3*x[-
1],          10, min(y), 2/3*x[-1],          10,          0, -x[-1],          0]
ub = [max(y),          1/2*x[-1], x[-1]/5, max(y),          3/4*x[-1], x[-
1]/5, max(y),          x[-1], x[-1]/5, max(y),          x[-1], max(y)/4]
bounds=[(lb[i], ub[i]) for i in range(len(lb))]

evol = 750
p_size = 40

```

```

result_1 = list(de2(rmse_1, bounds, mut=.8, crossp=.7, popsize=p_size, its=evol))
pop_1, fit_1, idx_1 = result_1[-1]
w_1 = pop_1[idx_1]
result_2 = list(de2(rmse_2, bounds, mut=.8, crossp=.7, popsize=p_size, its=evol))
pop_2, fit_2, idx_2 = result_2[-1]
w_2 = pop_2[idx_2]
result_3 = list(de2(rmse_3, bounds, mut=.8, crossp=.7, popsize=p_size, its=evol))
pop_3, fit_3, idx_3 = result_3[-1]
w_3 = pop_3[idx_3]

scenario = [rmse_1(pop_1[idx_1]), rmse_2(pop_2[idx_2]), rmse_3(pop_3[idx_3])]
mse_imbal = np.argmin(scenario)
print(f"The best fit has Scenarion #{mse_imbal+1} with the MSE: {scenario[mse_imbal]}")
imbal = pd.DataFrame(data = [w_1, w_2, w_3])

```

The tables above show examples of the author's functions, procedures, and methods in his research. They form parts of an automated, functional AFDI system for detecting and identifying faults in machines covered by the author's research.

7. Validation of the anomaly detection method

The author based the validation of the results of the OpEn 2D and OpEn 3D methods on data from the tests of large rotor machines. These are machines with active power exceeding 200MW and 560MW.

The kinematic diagram of the 200MW machine is presented in Figure 17. All relative vibration sensors installed on the device were used in the research. The turbine in such an arrangement has seven bearings. Each bearing is equipped with two relative vibration sensors. The sensors are oriented at an angle of 90 degrees – perpendicular to each other. Such a setup gives 14 sensors to analyze for each transient.

All the devices used in the tests are equipped with flexible rotors, which means that when changing the rotational speed, they pass through at least one resonance, called the area of critical revolutions of the machine.

7.1. Baseline measurements for the validation method

The author used data from various machines to validate the OpEn 3D method. In addition, transient data were acquired from diagnostic tests of different machines power output rated machines, i.e., turbo generators in professional power plants in Poland and abroad. Finally, the baseline data has been obtained following the methodology presented in section 4.3.

Analyzing such a large amount of various types data is complicated and sometimes impossible for an analyst within a reasonable timeframe. Figure 49 shows the data set from one run after the CS interpolation step, i.e., with correct data ready for further analysis. The figure is a collective chart of one of the reference transients. As one can see, the data from different directions and sensors, even assuming the correctness of dynamic behavior, differ significantly from each other.

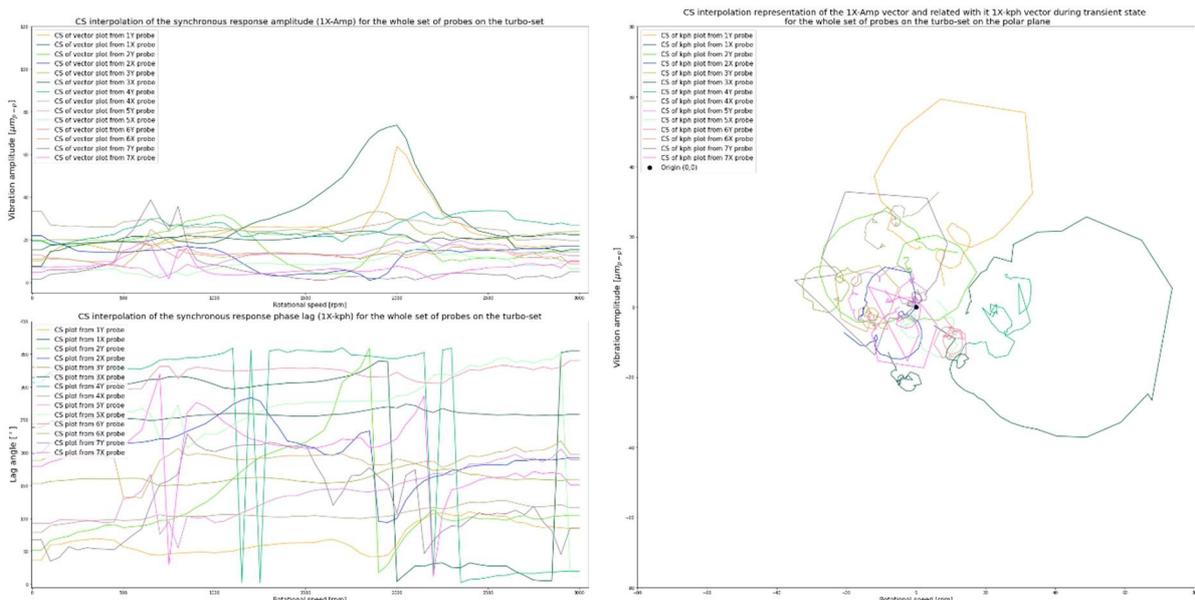


Figure 49. Example of reference one transient data for whole turboset.

The above figure consists of the data that make up the vibration vector in the amplitude and phase of the first component. The upper left graph shows the vibration vector amplitudes only. The lower left diagram shows the vibration amplitudes' phase lag values. In contrast, the chart on the right represents the vibration vector in polar coordinates. It reveals the validity of the analysis of the vibration vector as a whole and not each component separately. In particular, the phase angle shown in the lower left graph cannot be reliably analyzed as a value on its own due to the unstable behavior of its value around 360-0 degrees.

Baseline measurements data are in the form of matrices with the dimensions of 14x60 for machines whose transient state rotational speed is 0-3000 RPM. The methods for obtaining values for the individual matrices are described in sections 4.4 and 4.7. The row of each matrix is the values for the respective sensors. The columns are the appropriate values assigned to specific rotational speed values.

Table 22. Baseline measurement matrix with an synchronous amplitude in polar coordinate system.

	250	300	350	400	450	...	2850	2900	2950	3000
1Y	18.0797	17.6416	16.7544	15.7551	14.9392	...	17.2457	16.9995	17.1407	17.1407
1X	17.3126	17.9066	18.3764	18.6278	19.2790	...	15.5695	14.4821	15.2638	15.2638
2Y	17.1000	17.7225	18.4773	18.8098	19.2274	...	14.6703	14.2029	14.3156	14.3156
2X	14.6687	14.3613	14.1318	14.0141	14.0216	...	16.5651	16.8505	16.8406	16.8406
3Y	19.8054	19.6251	19.7208	19.5423	19.5133	...	23.2796	23.6045	23.7757	23.7757
3X	18.1819	18.4902	18.6752	18.8036	19.0409	...	22.4120	22.8693	22.3083	22.3083
4Y	19.5390	20.4219	20.8601	21.8092	22.4026	...	27.1439	27.8642	26.9442	26.9442
4X	26.2989	26.1141	25.9100	26.0061	25.9027	...	20.4005	19.9586	19.5699	19.5699
5Y	7.4255	7.3889	8.1528	8.1592	8.3797	...	14.3790	13.1150	12.4154	12.4154
5X	5.4076	5.4469	5.8714	6.0663	6.3985	...	7.9204	9.3487	6.3604	6.3604
6Y	10.1077	9.8364	9.6704	9.8126	9.6594	...	11.0051	12.1029	8.5884	8.5884
6X	10.2239	9.7091	8.8620	8.4680	7.6782	...	11.9129	10.3102	9.6962	9.6962
7Y	4.8272	5.8035	6.8282	7.9428	9.4456	...	1.9068	2.6601	4.9510	4.9510
7X	6.2322	6.6912	7.3006	8.1711	9.1624	...	8.3239	6.7158	10.1279	10.1279

Due to the size of these matrices, Table 22-Table 25 only present the slices presented to show the structure and the values of the results for the exemplary rotational speed values. Entire baseline measurements prepared for the purposes of this dissertation will be provided as an appendix to the thesis.

Table 23. Baseline measurement matrix with a phase angle of the first component in polar coordinate system.

	250	300	350	400	450	...	2850	2900	2950	3000
1Y	66.7684	67.6967	69.5585	68.2684	67.0447	...	87.3691	86.0002	84.9439	84.9439
1X	301.5792	303.5645	304.7411	305.5701	307.7440	...	5.1136	352.2297	354.7945	354.7945
2Y	79.3813	81.4762	84.4833	89.7442	90.7489	...	102.8261	104.0489	105.1258	105.1258
2X	219.9748	218.8219	216.9984	217.4119	216.7296	...	188.9264	190.4522	192.5758	192.5758
3Y	160.0722	160.3929	160.1435	160.1268	160.0894	...	158.9433	157.5124	159.4700	159.4700
3X	251.0256	251.6832	252.3456	252.5319	252.5334	...	257.5204	256.9892	258.5358	258.5358
4Y	327.4881	329.5929	331.4779	332.3640	333.5153	...	18.3314	19.5947	20.0900	20.0900
4X	94.9360	96.3703	97.0370	97.9308	98.4228	...	119.3251	121.0227	116.5172	116.5172
5Y	97.4220	97.2957	95.3839	93.7577	94.8259	...	192.6392	205.6180	189.5794	189.5794

5X	243.3398	242.1795	240.0160	237.8734	236.2938	...	342.1443	351.2257	18.1290	18.1290
6Y	325.6774	323.5088	320.4379	317.7451	314.2050	...	327.5943	330.2474	340.1089	340.1089
6X	197.5565	199.2203	197.3539	194.3232	191.5923	...	209.3198	218.2199	197.9550	197.9550
7Y	52.2139	55.6161	54.0261	53.5374	52.7177	...	81.4602	45.2078	85.6463	85.6463
7X	191.9125	194.5826	194.3495	195.3103	195.3209	...	169.9922	167.6255	151.2256	151.2256

Table 22 and Table 23 show an example of the baseline amplitude and the date phase, respectively. The baseline amplitudes are shown in Table 22. The data in this table can also be used as an OpEn centerline for the OpEn 2D method.

Table 24. Baseline measurement matrix with vibration vector in Cartesian coordinate system: x-coordinate.

	250	300	350	400	450	...	2850	2900	2950	3000
1Y	7.1315	6.6952	5.8515	5.8335	5.8265	...	0.7916	1.1858	1.5106	1.5106
1X	9.0662	9.9001	10.4722	10.8357	11.8014	...	15.5075	14.3491	15.2008	15.2008
2Y	3.1511	2.6268	1.7763	0.0840	-0.2513	...	-3.2567	-3.4477	-3.7355	-3.7355
2X	-11.2410	-11.1888	-11.2864	-11.1312	-11.2379	...	-16.3645	-16.5709	-16.4365	-16.4365
3Y	-18.6195	-18.4872	-18.5483	-18.3785	-18.3469	...	-21.7251	-21.8097	-22.2656	-22.2656
3X	-5.9118	-5.8109	-5.6637	-5.6444	-5.7151	...	-4.8431	-5.1487	-4.4339	-4.4339
4Y	16.4768	17.6129	18.3284	19.3210	20.0515	...	25.7664	26.2505	25.3048	25.3048
4X	-2.2628	-2.8974	-3.1743	-3.5882	-3.7941	...	-9.9914	-10.2862	-8.7373	-8.7373
5Y	-0.9592	-0.9383	-0.7650	-0.5347	-0.7050	...	-14.0306	-11.8257	-12.2422	-12.2422
5X	-2.4264	-2.5421	-2.9343	-3.2260	-3.5507	...	7.5389	9.2393	6.0447	6.0447
6Y	8.3477	7.9079	7.4552	7.2629	6.7348	...	9.2913	10.5075	8.0760	8.0760
6X	-9.7477	-9.1679	-8.4586	-8.2047	-7.5216	...	-10.3868	-8.1001	-9.2240	-9.2240
7Y	2.9577	3.2775	4.0110	4.7204	5.7216	...	0.2832	1.8741	0.3758	0.3758
7X	-6.0979	-6.4756	-7.0729	-7.8811	-8.8368	...	-8.1973	-6.5598	-8.8773	-8.8773

The matrix of the centers of the ellipses values in the (r, θ) , i.e., polar system is presented in Table 22 and Table 23. These are polar coordinates for the individual sensors for rows and the rotational speed increments values for columns, respectively.

Table 25. Baseline measurement matrix with vibration vector in Cartesian coordinate system: y-coordinate.

	250	300	350	400	450	...	2850	2900	2950	3000
1Y	16.6138	16.3218	15.6994	14.6354	13.7562	...	17.2275	16.9581	17.0740	17.0740
1X	-14.7489	-14.9209	-15.1006	-15.1519	-15.2449	...	1.3877	-1.9580	-1.3849	-1.3849
2Y	16.8072	17.5268	18.3917	18.8096	19.2258	...	14.3042	13.7781	13.8196	13.8196
2X	-9.4239	-9.0031	-8.5044	-8.5141	-8.3855	...	-2.5704	-3.0569	-3.6667	-3.6667
3Y	6.7504	6.5856	6.6985	6.6432	6.6453	...	8.3642	9.0283	8.3381	8.3381
3X	-17.1940	-17.5533	-17.7957	-17.9364	-18.1630	...	-21.8825	-22.2822	-21.8632	-21.8632
4Y	-10.5017	-10.3363	-9.9607	-10.1162	-9.9906	...	8.5371	9.3446	9.2552	9.2552
4X	26.2014	25.9529	25.7148	25.7573	25.6233	...	17.7863	17.1038	17.5111	17.5111
5Y	7.3633	7.3290	8.1168	8.1417	8.3500	...	-3.1463	-5.6705	-2.0661	-2.0661
5X	-4.8327	-4.8173	-5.0856	-5.1374	-5.3229	...	-2.4286	-1.4261	1.9791	1.9791
6Y	-5.6993	-5.8497	-6.1592	-6.5983	-6.9244	...	-5.8978	-6.0061	-2.9221	-2.9221
6X	-3.0840	-3.1962	-2.6433	-2.0949	-1.5429	...	-5.8335	-6.3787	-2.9890	-2.9890
7Y	3.8149	4.7895	5.5260	6.3880	7.5155	...	1.8857	1.8878	4.9367	4.9367
7X	-1.2864	-1.6847	-1.8094	-2.1576	-2.4209	...	1.4466	1.4392	4.8752	4.8752

Table 24 and Table 25 depict the coordinates of the centers of the ellipses for Cartesian coordinates. For practical and implementation reasons, the amplitudes and phases of the polar coordinate system have been converted to the (x, y) , i.e., Cartesian system. In the OpEn 3D method, the coordinate system is used to determine and assemble individual ellipse centers.

7.2. Single channel: 1X amplitude validation (OpEn 2D case)

The author validated the results of the OpEn 2D method on the data point sets obtained during the diagnostic tests of a 200MW class turbine set. During diagnostic tests, the turbine generator set experienced HP-IP cylinder rotor excessive misalignment. Data were collected for an invalid dynamic state of the machine. Then two series of improving the alignment of HP-IP cylinder rotors were undertaken. After each test, the maintenance department started the turbine set to obtain diagnostic data and provide further recommendations and instructions. After the final alignment attempt, the machine was allowed to start up fully. The turbine set reached FSNL, and its dynamic state allowed it to carry out further tests, synchronization, and load up to the nominal power, i.e., + 200MW.

This unit has seven journal bearings and one thrust bearing (combined journal and thrust bearing) placed in bearing pedestal no. 2. Schematic picture of this turbo-set is presented in Figure 17. Normally, these machines are equipped with eddy current relative shaft-to-rotor vibration sensors. Typically, all journal bearings in this type of turbine are equipped with such sensors. Every bearing has two sensors, oriented perpendicularly to each other. The most common set-up of eddy-current sensors is presented in Figure 7. Signal from these sensors is proportional to the shaft displacement with respect to the bearing housing.

The case study presents the data measured at bearing no. 1. Measurements were carried out during incremental improvement of the HP-IP part alignment. Data were collected during ten transient states, both startups, and cast-downs. Figure 50 presents all the transients on a single plot.

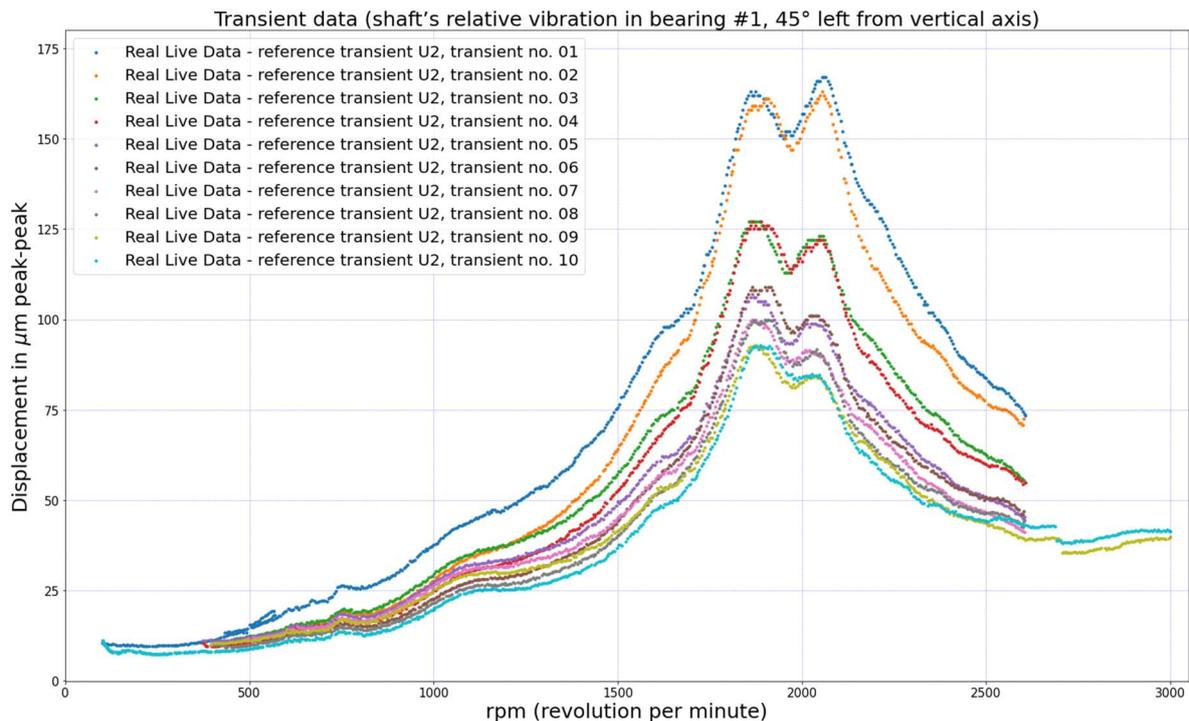


Figure 50. Transient data during measurements course. Transients 01-04 were non-satisfactory. Transients 05-10 were satisfactory. Transient 09 was selected as the reference.

In most cases, the coast-down transient is better suited for analysis than the run-up because the turbo-set does not experience additional excitation forces during this process. In such a case, the machine coast down is driven only by the inertia of the shaft. The author did not experience noticeable deviations between startups and coast downs during the analyzed measurements. That was a prerequisite for the inclusion of startups into our analysis as well.

There were no signs of any other malfunction apart from misalignment, for example, rubs which can produce a different response of a rotor system during startups and coast-down, as described in various examples, e.g. [14], [16], [21]. We classify the transients in the following way:

- first two pairs (transient no. 01÷04 in Figure 50) of transients are "non-satisfactory" in terms of vibration response,
- following three sets of pairs (transients no. 05÷10 in Figure 50) are "satisfactory" in proper alignment of the HP-IP coupling.

The OpEn centerline was calculated as presented in Section 4.3. Upper and lower bounds were set at $24\mu\text{m}_{pp}$ and $13\mu\text{m}_{pp}$, respectively, as explained in Section 4.5. During the first set of transients, the synchronous response exceeded the *OpEn BULL* in the [1500,2600] rotational speed interval. Transient no. 1 and 2 in Figure 51 depicts this scenario.

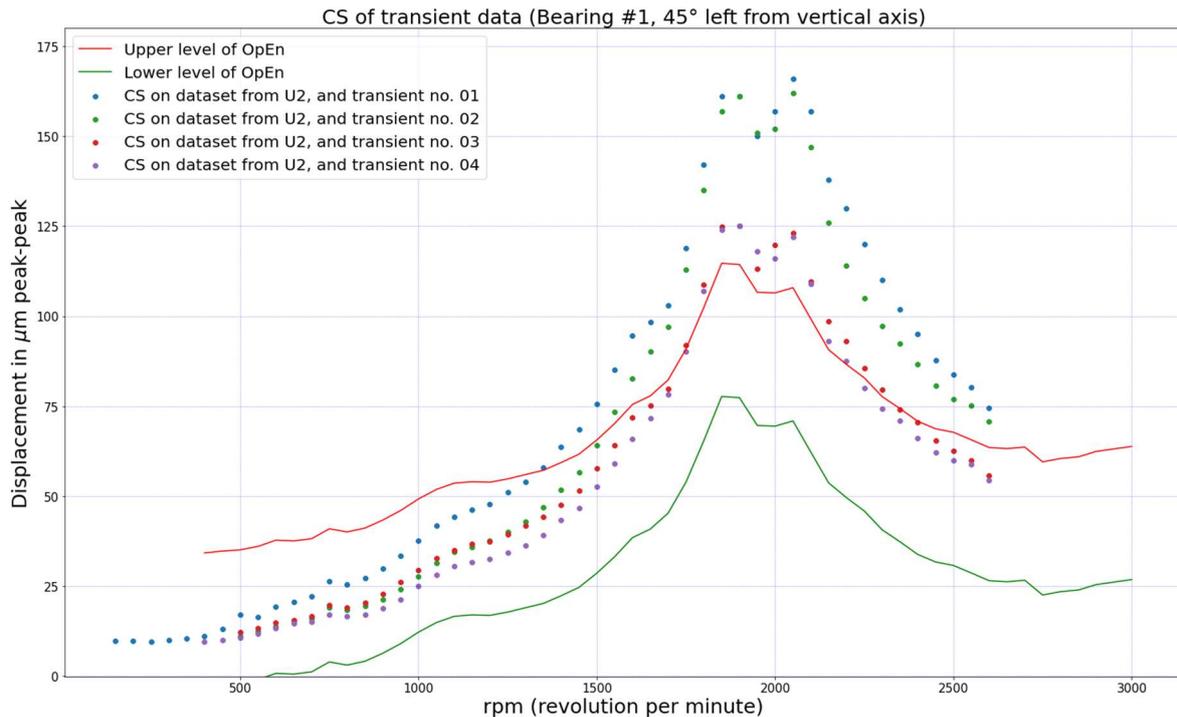


Figure 51. Example of a transients with misalignment (the initial state – transient no.1 and no.2 and after first improvement – transient no.3 and no.4).

After the first alignment improvement, the majority response of the rotor system fell into the OpEn. From the startup of the turbo-set up to approx. 1750RPM and above 2450RPM, all amplitudes were inside OpEn. Still, the system response values between approx. 1700÷2450RPM had higher values than the OpEn upper bound, which can be seen in Figure 51, transient no. 3 and 4.

The second improvement of the HP-IP cylinder alignment resulted in the proper response of the system. Figure 52 presents the dynamic data for the described situation.

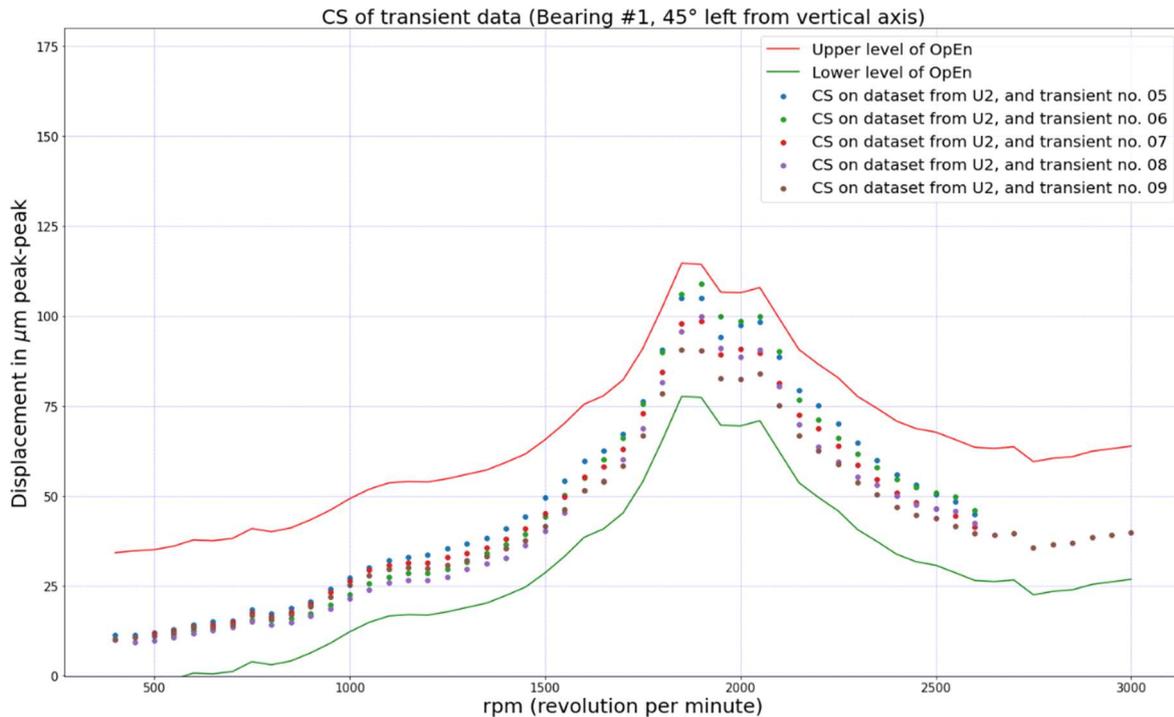


Figure 52. Example of acceptable dynamic behavior.

The figures presented only the qualitative results. To be able to automate the assessment process, the parameters proposed in Section 4.6 were applied and presented in Table 26. The transient no. 09 named "U2_09" was assumed to be the reference one. Hence, the RMSE and Kurtosis value in the one before last column in Table 26, also named "U2_09," is 0. It is worth underlying that RMSE and Kurtosis values for the last measured transient state named "U2_10" were the lowest ones even though it contained samples from the whole rotational speed span (which was >100RPM up to 3000RPM).

Table 26 summarizes the performance of the proposed distance criteria. After the second alignment improvement, RMSE of the further transient in the studied case does not exceed the value of 10, as shown in Table 26, and since then, all amplitudes of synchronous response fell between OpEn upper and lower values.

Table 26. Comparison of the OpEn 2D method selection criteria.

	U2_0 1	U2_0 2	U2_0 3	U2_0 4	U2_0 5	U2_0 6	U2_0 7	U2_0 8	U2_0 9	U2_1 0
RMSE	42.58	37.27	20.22	17.73	8.00	7.83	4.03	3.44	0.00	3.42
Kurtosis	1.84	1.94	1.77	1.95	1.77	2.27	2.10	2.90	0.00	2.05
MAX_Oo_OpEn	72.10	68.10	29.16	28.10	0.00	0.00	0.00	0.00	0.00	0.00
MIN_Oo_OpEn	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 53 depicts the visual evolution of the RMSE parameter. The RMSE is an error at each transient during the measurement course. However, after the machine's fourth transient (second HP-IP coupling improvement), the dynamic response is much closer to the reference transient than before.

RMSE is a suitable parameter as it is sensitive to the distance from the healthy state. As we show in the case study above, it is sensitive to the misalignment level. There is a value above which misalignment is beyond an acceptable level. In the studied example, the value of 10 can be a good condition indicator (still, for this particular sensor and this type of malfunction).

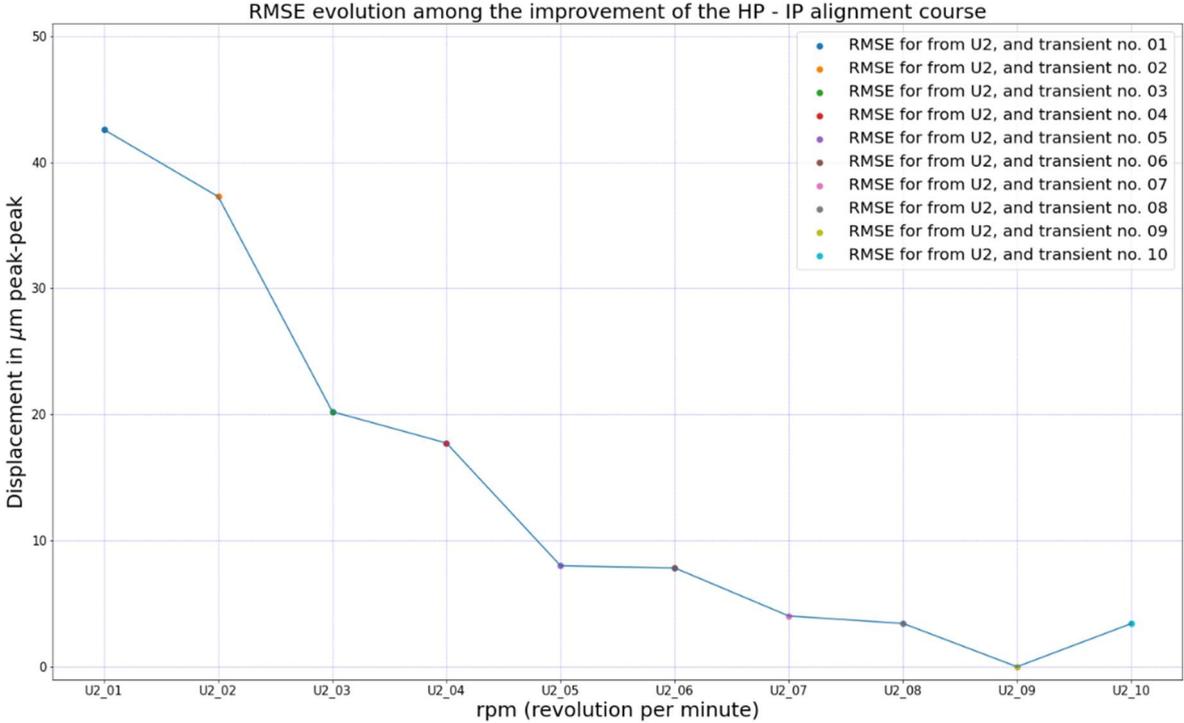


Figure 53. The evolution of the Root Mean Square Error (RMSE) vs HP-IP alignment incremental improvement.

The kurtosis parameter is between 2.90 and 1.77. The values do not show the relationship to the level of misalignment. Thus, the Kurtosis parameter is not helpful in this case study. In our investigation, Kurtosis measures how the new transient is similar to its reference one in shape. It can signal if some samples were far off the reference transient during a particular transient. This parameter may play a significant role in finding anomalies such as oil whirl or whip. The transient of a machine that experiences such phenomena can be extremely different from the reference one. Amplitudes generated during instabilities are often close to bearing clearances, which can harm turbo-set equipment such as the bearing itself, its oil seals, steam seals on the rotor and inside of a turbine casing, and hydrogen oil seals (on the generator), and others. Furthermore, the rotational speed intervals in which hydrodynamic instabilities can occur might be narrow compared to the whole rotational speed range. Thus, in such cases, RMSE as a single assessment parameter of the transient cannot suffice because even if the signal amplitude is much greater in a short interval, the number of samples in the transient as a whole will diminish it.

Setting up the Kurtosis parameter will be a subject of further studies. Author will study the effect of setting up RMSE and Kurtosis parameters on different signal components in different arrangements, for example, RMSE on synchronous response and phase angle and Kurtosis on direct (or sub-synchronous) response.

MAX_Oo_OpEn well describes misalignment in the studied example. This indicator, though, detects if, at any given moment during a transient state, the vibration exceeds the *OpEn_Upper* value. This

parameter detects if any samples exceed the upper bound, and in such a case, it returns the distance value and the relevant rotational speed. This parameter presents information about "the worst" sample. This parameter can signal abnormal machine behavior during transient, for instance, hydrodynamic instability. Thus, the *MAX_Oo_OpEn* is well suited for novelty detection purposes.

As shown in Table 26, no transient exceeded the *OpEn_Lower* value during the presented case study, so the *MIN_Oo_OpEn* parameter cannot be evaluated. Above-described situation can imply two things:

- OpEn lower value can be set to too a low value which can cause the false positive error (lack of detection in the early stage of malfunction evolution);
- misalignment is present in a shaft train, there will be no samples with amplitudes lower than expected.

These two scenarios will be the subject of our further studies.

7.3. 1X amplitude and phase validation (OpEn 3D case)

To validate the OpEn 3D method, the author analyzed machines with different malfunctions. The author will use the example of the rotor unbalance of the + 560MW turbine generator set for a detailed presentation to highlight the results of validating the OpEn 3D method. This data was additionally used to verify the MD3 method to validate the complete detection and identification process. The layout of the tested device is shown in Figure 18. The author had to make some changes in the sensor description to compare the data from the tested device with the correct values of the ellipses of the OpEn 3D method. In addition, the way of naming the data from the machine had to change. The sensors monitoring the dynamic state of the generator rotor in the 560MW unit are on bearings number eight and number nine. In the baseline matrices, the sensors responsible for the dynamic generator response are marked with number six and number seven.

To prove the validity of the thesis, the author decided to assign data from bearings number eight and number nine coming from the real object as bearing number six and number seven, respectively, for the OpEn 3D algorithm to compare the data from the generator to the ellipses assigned to the generator.

Based on the constant speed data, operational personnel reported high vibration levels in bearing number nine. Vibration measurements were carried out to verify the cause of the high vibration. Data were recorded during transient operation (coast-down) of the unit. The portable data acquisition interface unit was connected to eddy-current type vibration displacement sensors at all nine bearings in both directions.

Transient data was recorded, and unbalance of the generator rotor free end (near the bearing number nine) was diagnosed. After the balancing operation, the data was measured once more during the run-up. The balancing operation was qualified as satisfactory. The turbogenerator was considered eligible for long-term operation with no restrictions in terms of dynamic condition to run within a full range of operation (referred to as the class A).

The author used data from the generator sensors described above only to analyze and present the results. The data was used to validate the method.

Table 27 summarizes the results for the first two transients participating in the research and the OpEn 3D method. The table layout is as follows: the values in the first column of the matrix correspond to the individual rotational speed values for the successive instances determined by the CS interpolation domain. Then the column values assigned to "Transient no. 01" and "Transient no. 02" are the values of the distance from the baseline ellipse for the individual sensors, described 6X-7Y, in specific rotational speed instances during the duration of the transient state.

Transient no.1 summarizes the data collected immediately before the balancing activities. One can see that the most significant values of the distance from the ellipse come from the 7Y direction. For example, the distance value at 3000RPM is over $240\mu\text{m}_{pp}$. This is a very high value because the bearing clearances for the generator bearings can be $400\mu\text{m}$, and the clearances on the new hydrogen seals can be around $300\mu\text{m}$.

Table 27. OpEn 3D distance matrix: left – transient before balancing attempt; right – after trial mass balancing.

Rotational speed[rpm]	Transient no.01				Transient no.02			
	6Y	6X	7Y	7X	6Y	6X	7Y	7X
0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0
250	0	0	0	0	0	0	0	0
300	0	0	0	0	0	0	0	0
350	0	0	0	0	0	0	0	0
400	0	0	0	0	0	0	0	0
450	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0
550	0	0	0	0	2.840	0	0	0
600	0	0	4.576	0	5.980	0	11.247	0
650	0	0	2.559	0	3.534	0	13.445	0
700	0	0	15.879	0	0	0	29.818	0
750	0	0	6.344	0	0	0	17.951	0
800	1.379	0	0	0	0	0	5.032	2.764
850	5.853	0	0	5.006	0	0	0	5.930
900	1.612	0	0	0	0	0	0	0
950	9.494	0	0	0	6.102	0	0	0
1000	8.809	0	0	0	6.397	0	0	0
1050	7.334	0	0	0	6.396	0	0	0
1100	9.394	0	0	0	4.621	0	0	0
1150	4.617	0	0	0	0.197	0	0	0
1200	6.616	0	0	0	3.753	0	0	0
1250	5.202	0	0	0	0	0	0	0
1300	0	0	0	0	0	0	0	0
1350	0	0	0	0	0	0	0	0
1400	0	0	0	0	0	0	0	0
1450	0	0	0	0	0	0	0	0
1500	0	0	0	0	0	0	0	0
1550	0	0	1.716	0	0	0	0	0
1600	1.371	0	1.704	0	0	0	0	0
1650	4.857	0	11.696	0	0	0	0	0
1700	9.650	0	15.990	0	0	0	0	0
1750	10.643	0	24.584	0	0	0	0	0
1800	11.539	0	36.928	0	0	0	0	0
1850	8.953	0	38.346	0	0	0	0	0
1900	7.221	0	46.684	0	0.920	0	4.370	0
1950	10.049	0	51.388	0	3.704	0	9.588	0
2000	11.785	0	58.633	0	5.075	0	13.607	0
2050	5.582	0	58.056	0	0	0	13.801	0
2100	8.338	0	65.568	0	0	0	14.157	0
2150	7.160	0	69.574	0	0	0	10.142	0
2200	6.850	0	72.192	0	0	0	13.422	0
2250	10.662	0	74.539	1.838	0	0	19.053	0
2300	11.332	0	75.357	6.580	1.183	0	23.129	0
2350	9.319	0	79.502	7.704	0	0	25.360	0
2400	4.955	0	86.786	11.471	0	0	27.545	0
2450	8.037	0	93.911	14.183	0	0	27.505	0
2500	8.824	0	104.024	16.749	0	0	25.201	0
2550	8.578	0	108.776	18.645	0	0	24.218	0
2600	10.433	0	117.949	22.628	0	0	23.463	0
2650	13.785	0	120.163	25.486	0	0	20.540	0
2700	13.271	0	137.601	30.525	0	0	28.117	0
2750	14.580	0	155.175	36.467	0	0	38.077	0
2800	14.859	0	173.610	41.778	0	0	44.992	0
2850	11.248	0	190.245	49.343	0	0	56.392	0.921
2900	5.366	0	208.620	58.103	0	0	68.135	4.208
2950	0	0	220.438	68.400	0	0	77.424	5.474
3000	0	0	240.213	72.901	0	0	0	0

Figure 54 shows a graphical representation of the OpEn 3D method sample values. The figure presents a baseline ellipse with the origin of the coordinate system depicted as a black point. On the other hand, the red one is related to the vibration vector of the first synchronous component, i.e., the combination of the amplitude and phase measured during the transient. Additionally, each ellipse in the *OpEn BULL* matrix is rotated by a specific value determined by the vibration vector parameters described in Section 4.8. The figure shows six moments for which the distance from the ellipse is calculated. The graph in the upper left corner shows the last rotational speed for which the vibration vector is still inside the ellipse. The figure on the top right shows a slight exceedance of the acceptance ellipse limit already for the rotational speed value of 1550. The figures in the middle and lower parts show the evolution of the distance of the vibration vector from the ellipse boundary.

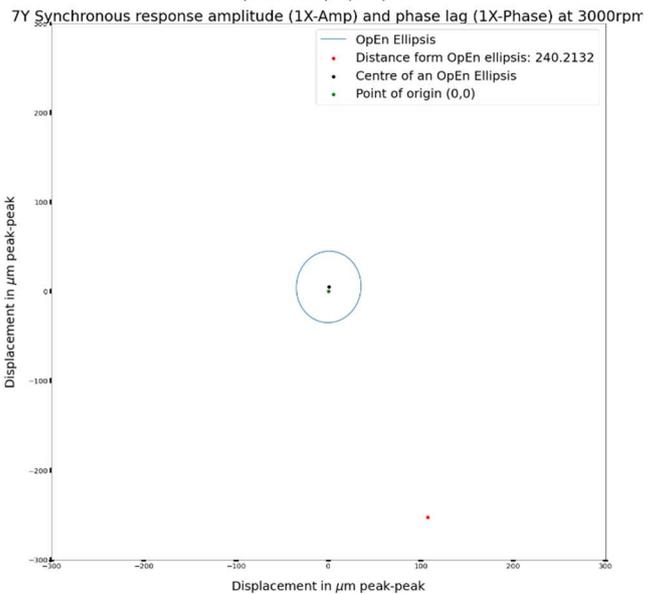
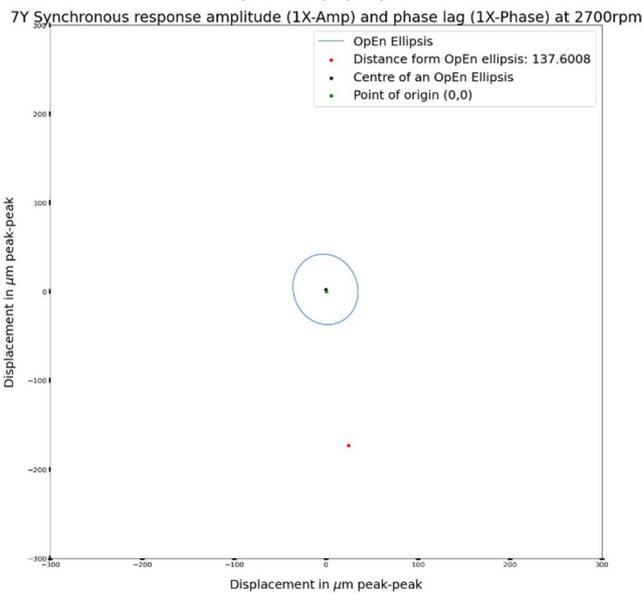
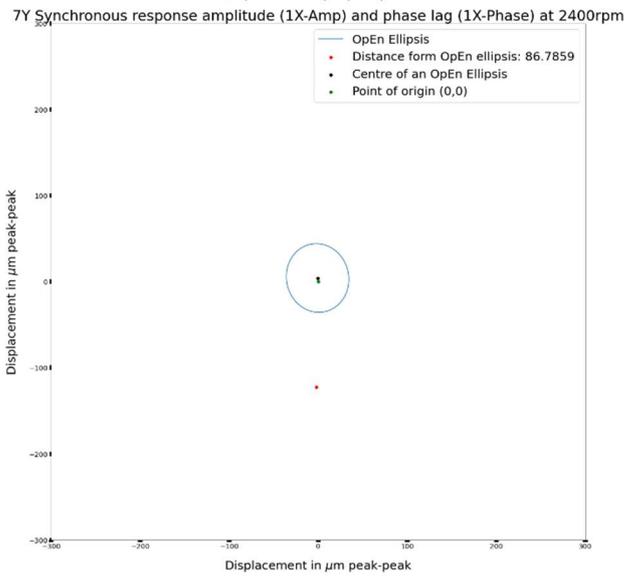
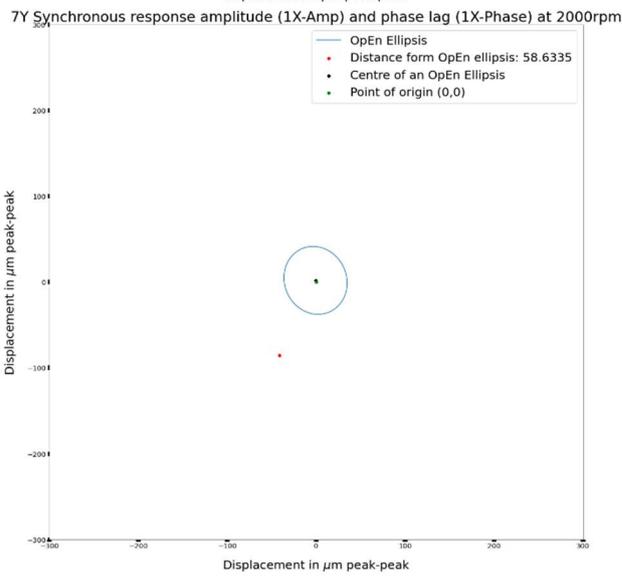
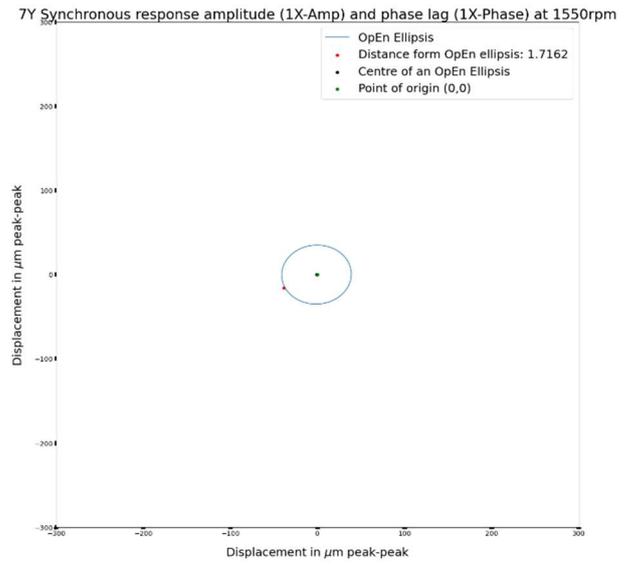
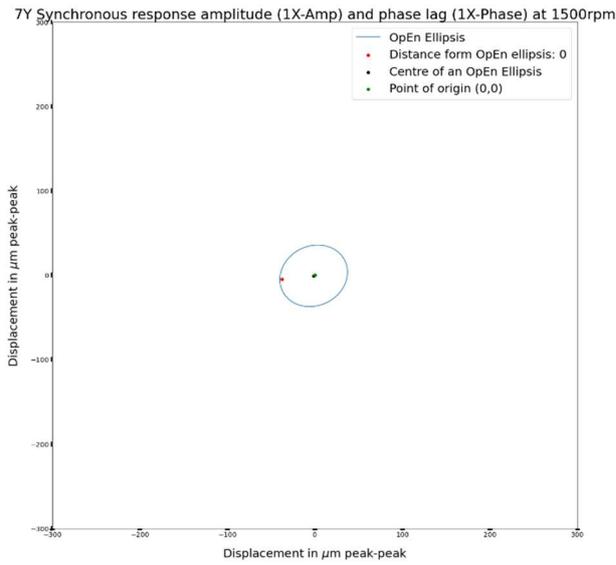


Figure 54. Graphical representation of the OpEn 3D distance matrix: 7Y sensor before balancing attempts.

The columns in Table 27 entitled Transient no 2 shows data after the first generator rotor balancing attempt. The weight was added on the NDE side to the rotor disc. The trial mass was approx. 560g on a radius of approx. 450mm. After the workshop personnel installed the mass on the test angle, the maintenance staff started the machine to measure the dynamic state. The measurements revealed a significant improvement in the generator rotor dynamic parameters. Table 27 shows that for rated speed (FSNL), the vibration amplitude on the NDE bearing in the Y direction (sensor described as 7Y) decreased by $170\mu\text{m}_{pp}$.

Unfortunately, the initial vibrations were so large that even such a significant improvement did not give an entirely satisfactory result. Therefore, the balancing cycle had to be repeated. During trim balance, the research team decided not to change the balance mass but only the angular orientation of the balancing mass. Following trim balance, the vibration response of the unit reviled a very close distance to the acceptance regions for both bearings and all directions of vibration measurement.

Table 28 collects the following two transients during trim balance and final transient, after which the turboset was considered acceptable for long-term operation without any restrictions. As a result, the power plant authorities decided to discontinue the corrective actions and leave the turbine set at the disposal of electricity production.

Table 28. OpEn 3D distance matrix: left – trim balance attempt; right – start-up for the long-term operation.

Rotational speed[rpm]	Transient no.03				Transient no.04			
	6Y	6X	7Y	7X	6Y	6X	7Y	7X
0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0
250	0	0	0	0	0	0	0	0
300	0	0	0	0	0	0	0	0
350	0	0	0	0	0	0	0	0
400	0	0	0	0	0	0	0	0
450	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0
550	0	0	0	0	0	0	0	0
600	0	0	6.772	0	0	0	16.799	0
650	0	0	32.911	0	0	0	22.625	0
700	0	0	67.868	0	0	0	36.461	0
750	0	0	52.766	0	0	0	24.009	0
800	0	0	32.472	9.018	0	0	12.665	2.492
850	0	0	0	9.447	0	0	0	7.009
900	0	0	0	0	0	0	0	0
950	0	0	0	0	1.288	0	0	0
1000	0	0	0	0	4.031	0	0	0
1050	0	0	0	0	7.342	0	0	0
1100	0	0	0	0	8.003	0	0	0
1150	0	0	0	0	5.058	0	0	0
1200	0	0	0	0	6.886	0	0	0
1250	0	0	0	0	3.648	0	0	0
1300	0	0	0	0	0	0	0	0
1350	0	0	0	0	0	0	0	0
1400	0	0	0	0	0	0	0	0
1450	0	0	0	0	0	0	0	0
1500	0	0	0	0	5.038	0	0	0
1550	0	0	0	0	8.719	0	0	0
1600	0	0	0	0	8.806	0	0	0
1650	0	0	0	0	11.878	0	0	0
1700	0.571	0	0	0	16.708	0	0	0
1750	0.648	0	1.464	0	19.700	0	0	0
1800	0.983	0	0	0	19.488	0	0	0
1850	5.294	0	0	0	21.926	0	0	0
1900	4.129	0	0	0	22.949	0	0	0
1950	3.313	0	0	0	23.364	0	0	0
2000	4.251	0	0	0	25.740	0	0	0
2050	2.901	0	0	0	23.263	0	0	0
2100	1.523	0	0	0	18.009	0	0	0
2150	1.793	0	0	0	18.521	0	0	0
2200	3.975	0	0	0	20.139	0	0	0
2250	6.271	0	0	0	23.330	0	0	0
2300	7.941	0	0	0	21.875	0	0	0
2350	4.161	0	0	0	18.450	0	0	0
2400	0	0	0	0	14.700	0	0	0
2450	0	0	0	0	12.760	0	0	0
2500	0	0	0	0	13.864	0	0	0
2550	0	0	6.000351	0	12.103	0	0	0
2600	0	0	5.98333	0	12.129	0	0	0
2650	0	0	13.58429	0	7.999	0	0	0
2700	0	0	13.77618	0	5.208	0	0	0
2750	0	0	17.45706	0	4.950	0	0	0
2800	0	0	20.09014	0	0	0	1.617	0
2850	0	0	18.23544	0	0	0	3.673	0
2900	0	0	19.97202	0	1.171	0	6.873	0
2950	0	0	24.858	0	0	0	15.557	0
3000	0	0	23.89176	0	0	0	18.827	0

Figure 55 presents the evolution of the distance of the vibration vector from the ellipse acceptance region for the 7Y sensor. The ellipses were given for the same rotational speed values in the case of data from before and after the corrective actions.

An evident improvement in the position of the vibration vector can be seen comparing Figure 54 and Figure 55 concerning the baseline region.

It is also worth mentioning that the author used the reference machines data sets to define the acceptance regions in the form of ellipses sets for the validation process. The data used to determine the baseline come from new machines or machines after repairs and factory acceptance installed correctly on site.

The data presented in this section comes from a machine operating for a long and indefinite period. The author did not have data from the transitional states immediately after its launch. Therefore, it is impossible to refer to the starting vibration level of the machine in this way.

However, after corrective actions, one can conclude that the machine's behavior over the entire measured rotational speed range is similar to the reference behavior defined by the acceptance regions. Almost all distances throughout the transient state are close to zero, as shown in table 23 in Transient no. 4 columns.

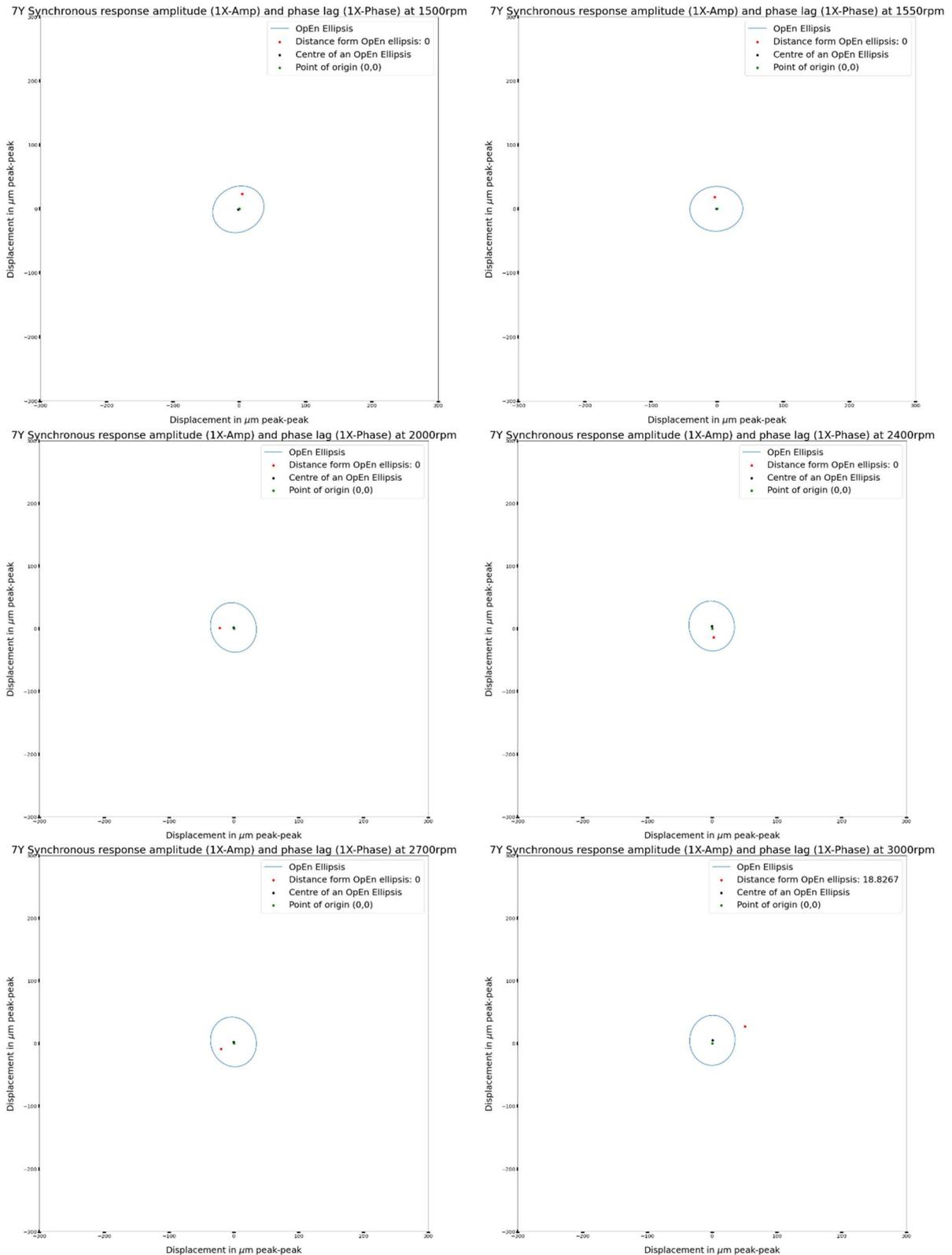


Figure 55. Graphical representation of the OpEn 3D distance matrix: 7Y sensor before balancing attempts

Figure 56 summarizes the change in the distance of the vibration vector from the acceptance region for a repaired NDE bearing. One can see a significant improvement already in Transient no. 2. The distances of the vibration vectors in both directions of this bearing have significantly decreased their values. The above proves a significant improvement in the dynamic condition of the tested device. The

graph reveals that the distance for Transient no. 4 turned out to be greater than zero. This means that the vibration vector for the 7Y direction goes beyond the acceptance region.

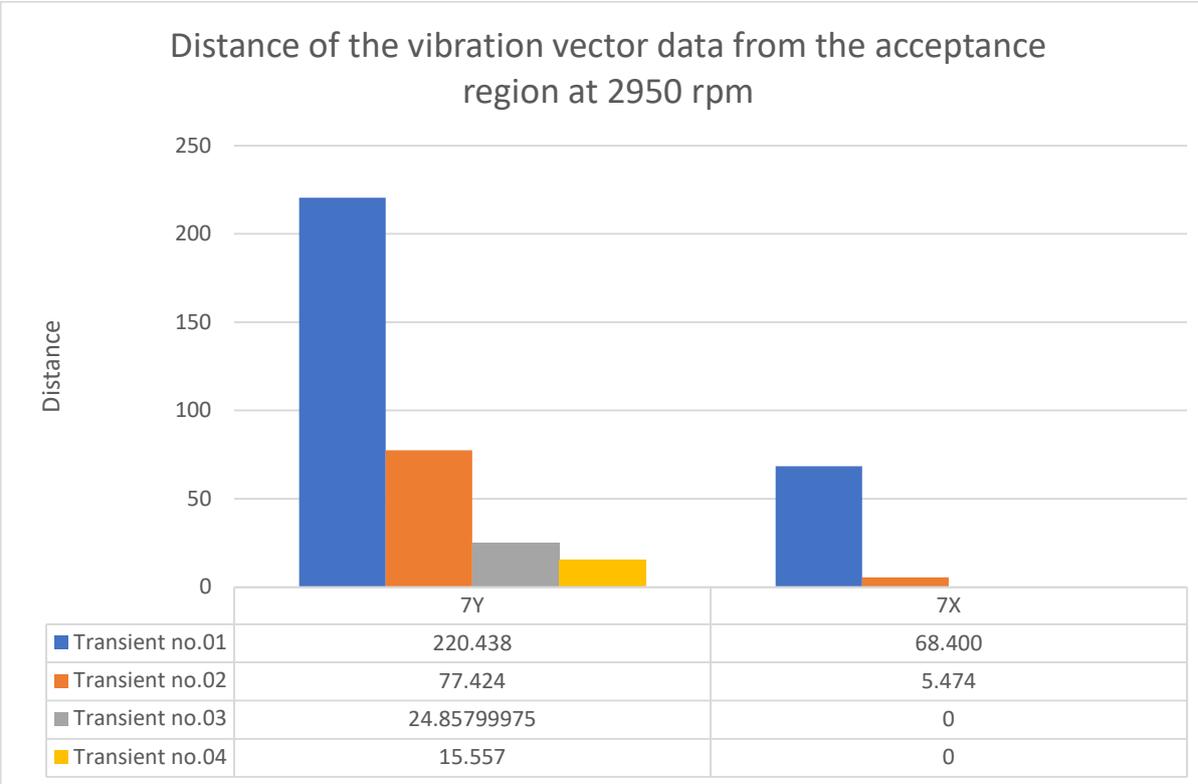


Figure 56. 7Y vibration vector data distance from the acceptance region evolution during corective actions.

Nevertheless, the value of 16 is minimal compared to 220. In addition, by examining Figure 55, one can conclude that the position of the vibration vector is very close to the ellipse corresponding to 3000 RPM. The center of the ellipse, which is marked with the green point in the graphs, is located on the opposite side of the origin of the coordinate system, shown in black. Therefore, one can assume that for a different value of residual unbalance, or after another balancing attempt, the vibration vector would be inside the ellipse.

he summary of activities aimed at improving the dynamic state of the turbine set under study is shown in Figure 57. The data in the graph present all transients who participated in the tests on the research object. All subsequent transients start at 1500 RPM. It allows for a fair comparison of the fit of the data.

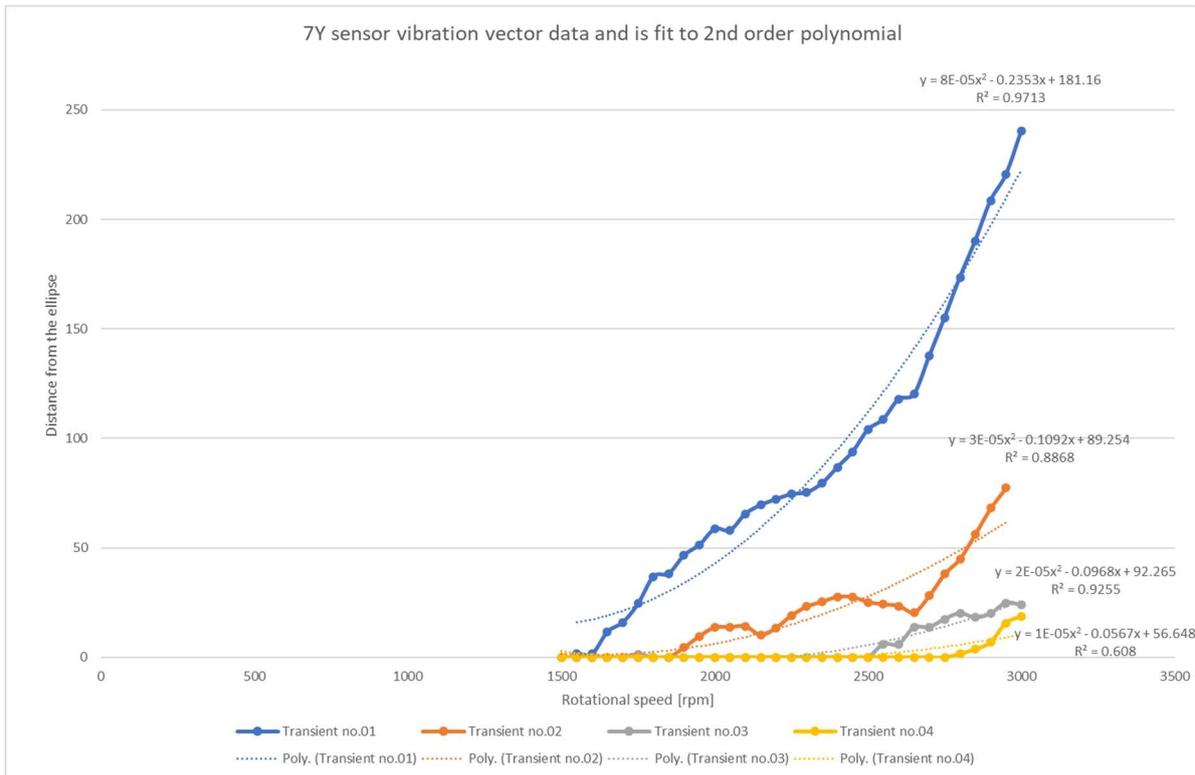


Figure 57. Polynomial data fit to the vibration data vector during consecutive transient.

The author defined a trend line as a second-order polynomial for the data from each transient. As described in section 12, such a polynomial can correctly nail the unbalance. For each of the trend lines, the correctness parameter r^2 was given. As seen in the discussed figure, the data from the transient no. 1 is characterized by the highest value of the parameter fitting to the second-order polynomial. 1 in the figure shown in blue. The adjustment value is 0.97. It proves the correctness of the hypothesis about excessive unbalance on the generator rotor on the NDE side.

After the research team completed the corrective actions, the trend line fit to the data dropped significantly, and the adjustment parameter was 0.61. It demonstrates a significant reduction in the centrifugal force from the unbalance. The unbalance malfunction was primarily removed from the tested object thanks to the corrective actions.

8. Validation of the identification method

The author carried out the MD3 method validation process in two stages. In the first stage, the method was implemented on data from the test rig. As a result, the stand was prepared with focus on simulating the malfunction of the rotor unbalance. In the next stage, the author used the MD3 method to identify the generator rotor unbalance parameters nominated by the OpEn 3D method described in the previous chapter.

8.1. Validation of model data

The authors validate the model on a Rotor Kit. It is a simplified model of a rotating machine with a flexible rotor. The model is presented in Figure 4, and it is a variation of the simplified Jeffcott rotor model well described by, e.g., Kiciński [9], Muszyńska [8], and Ehrich [7].

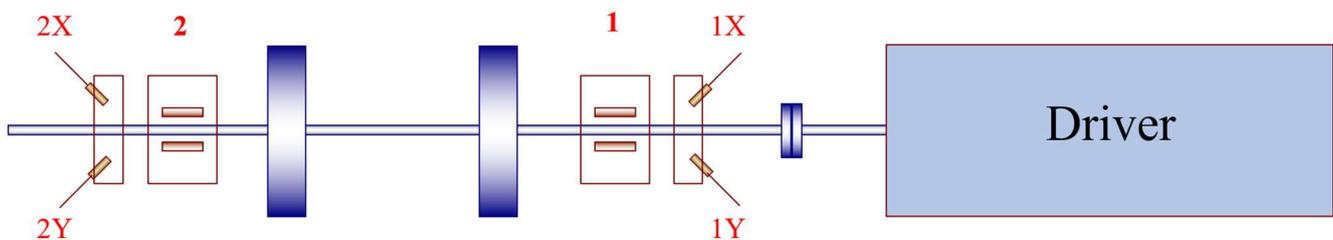


Figure 58. Scheme of the test rig used for validation purpose.

The model schematic, depicted in Figure 58, consists of two spaced masses, a variable speed-controlled driver, and brass-bushing bearings. The bearings are described by number 1 and 2 in the Figure 58, respectively. The sensors at each bearing are oriented by the convention driver-to-driven. The Y direction means that the sensor is oriented 45° in left from the vertical axis. The X direction means that the sensor is oriented 45° in right from the vertical axis, and the 90° from the Y sensor. Figure 9 presents detailed schematics of the sensors arrangement. The validation method uses two sensors on either side of the rotor.

Figure 5 presents the picture of the verification model on the test stand. To validate the identification of at least the first bending mode Rotor Kit has to be rotated with a velocity of over 4000 rpm. Then, the model for the unbalance response is validated by mass addition on both disks at the same angular orientation.



Figure 59. The Rotor Kit test rig used for model validation.

The data from two experiments were recorded. First, an imbalance mass was added to the rotor as described above. This trial is considered as the presented system imbalance response. The unbalance mass was removed during the second trial, and the transient data set was recorded. The vibration levels throughout the whole transient were at a low level and it was considered malfunction-free.

Several transient runs were performed and recorded. The resulting data showed convergence and repeatability of the test rig setup. Figure 6 depicts examples of the transient response of the data prepared for identification.

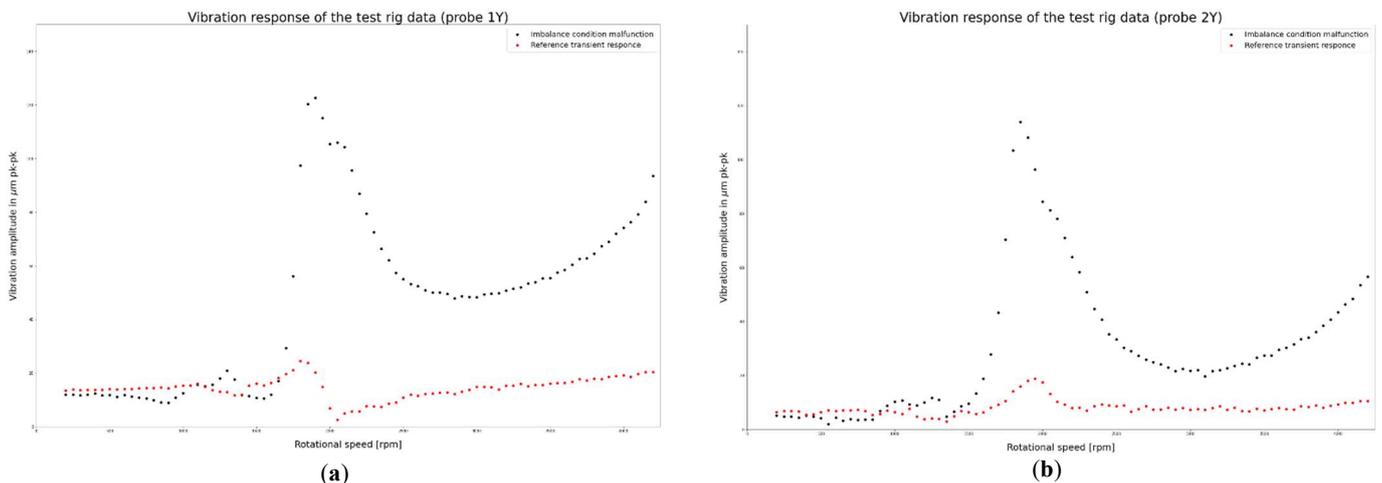


Figure 60. The transient vibration response of the test rig during with and without unbalance, respectively: (a) data from the bearing number 1; (b) data from the bearing number 2.

Sensors with the same angular orientation were taken into account to analyze the validation data. Each of the sensors is mounted on the same side of the rotor. The data (with and without unbalance) were recorded and processed by the MD3 identification method. Figure 61 presents the curve shapes plotted as lines based on scenario 1÷3 against the real-object transient data curve (plotted as a scatter plot). Based on the MD3 method, scenario 2 was selected as the best approximation of the sensor 1Y

data, and scenario 3 was the best one to fit the transient data from sensor number 2Y. Table 29 presents a summary of RMSE values for all three scenarios for the case of an unbalanced rotor.

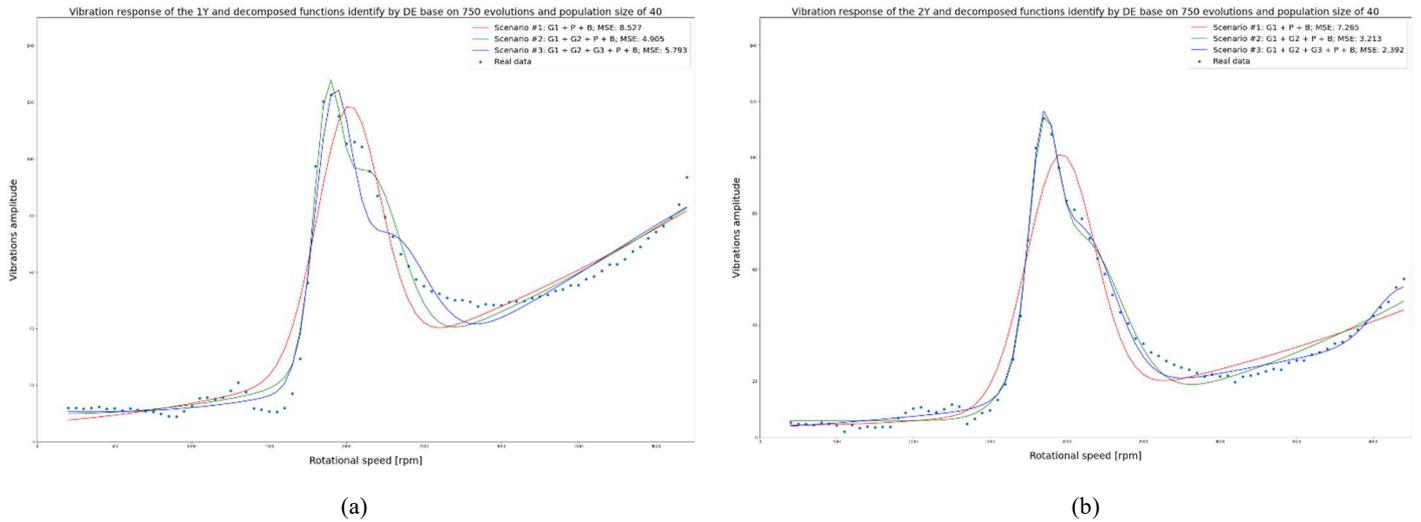


Figure 61. Outcome of the MD3 method based on three scenarios proposed in Chapter 5 on a test rig with imbalance malfunction simulation: (a) Data obtained from the 1Y sensor, driven side of the rotor; (b) Data obtained from the 2Y sensor, non-driven side of the rotor.

Table 29. The MSE values are based on the scenario and sensor location for the imbalance simulation on the test rig.

RMSE	Sensor 1Y	Sensor 2Y
Scenario 1	8.527	7.265
Scenario 2	4.905	3.213
Scenario 3	5.793	2.992

The scenario number one shows the worst fit to the transient data. The RMSE indexes for both sensors 1Y and 2Y have the highest value. Such a situation is most likely due to a split resonance (i.e. two resonances close to each other) measured in both bearings. The split occurs in the target function between 1500-2500rpm. Unable to adjust to the two resonances close together, the scenario chose an "in-between" resonance—such a compromise results from an increased mis-match between the scenario functions and the measured transient function.

Scenario 2 and Scenario 3 for sensor 1Y and sensor 2Y, respectively, were selected as the best sets of decomposition function parameters. A summary of all the identified parameters by the DE algorithm is presented in Table 30. In addition, parameters for the best scenario are highlighted.

Table 30. Parameters of the decomposed functions identified by the DE algorithm and chosen by the MD3 method based on the imbalance data and the MSE as the quality index.

Scenario/sensor	a_{g_1}	rpm_{g_1}	w_{g_1}	a_{g_2}	rpm_{g_2}	w_{g_2}	a_{g_3}	rpm_{g_3}	w_{g_3}	a_p	x_p	$const_b$
1/1Y	91.27	2014.41	204.93	-	-	-	-	-	-	75.45	-462.77	6.19
2/1Y	71.36	2100.00	209.76	63.31	1866.85	75.07	-	-	-	73.46	-46.59	8.78
3/1Y	75.67	1891.59	97.01	61.81	2157.43	241.15	13.36	4200.00	10.00	75.05	575.15	9.60
1/2Y	89.12	1962.59	222.43	-	-	-	-	-	-	41.02	266.84	4.41
2/2Y	64.23	1848.03	91.10	62.21	2082.89	268.12	-	-	-	42.59	1347.73	6.03

In the next step, the unbalance weights were removed and the transient data was recorded for the analysis with the same set of sensors as earlier. Figure 62 depicts malfunction-free transients for the sensors 1Y and 2Y, respectively. Due to small amplitude values during these runs, the results of the DE algorithm, i.e., decomposition function parameters and hence the RMSE quality index, are similar in values. The values of the RMSE index concerning scenarios are presented in Table 31.

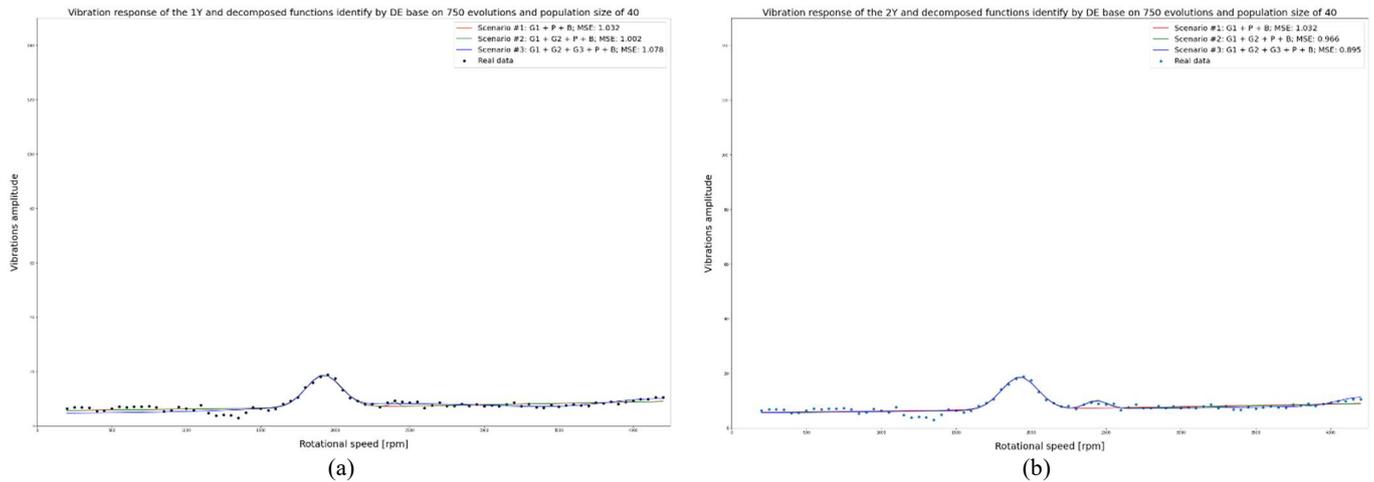


Figure 62. MD3 method outcome based on three scenarios proposed in Section 2 on a test rig without malfunction: (a) Data obtained from the 1Y sensor, driver side of the rotor; (b) Data obtained from the 2Y sensor, non-driven side of the rotor.

Based on the RMSE quality index, the MD3 method shows that the best fit of the decomposed functions for the reference transients provides scenario 2 and scenario 3 for the sensors 1Y and 2Y, respectively. However, it is visible that the values are very similar for all the scenarios. In such a case, a simpler model should be chosen if in doubt.

Table 31. The RMSE values, based on the scenario and sensor location, for the reference transient simulation on the test rig.

MSE	Sensor 1Y	Sensor 2Y
Scenario 1	1.032	1.032
Scenario 2	1.002	0.966
Scenario 3	1.078	0.895

Table 32 summarizes all the decomposed function coefficients nominated by the MD3 method. This table highlights the best solution for sensor data 1Y and 2Y in bold font.

Table 32. Parameters of the decomposed functions identified by the DE algorithm and chosen by the MD3 method based on the reference data and the RMSE as the quality index.

Scenario/sensor	a_{g_1}	rpm_{g_1}	w_{g_1}	a_{g_2}	rpm_{g_2}	w_{g_2}	a_{g_3}	rpm_{g_3}	w_{g_3}	a_p	x_p	$const_b$
1/1Y	11.82	1924.58	118.94	-	-	-	-	-	-	4.35	-3634.30	4.72
2/1Y	11.85	1924.53	119.41	18.87	2378.38	10.00	-	-	-	4.32	-3523.06	4.72
3/1Y	11.44	1920.28	113.26	3.44	2519.68	840.00	4.82	4200.00	374.83	18.87	4200.00	4.72
1/2Y	11.82	1924.58	118.94	-	-	-	-	-	-	4.35	-3634.30	4.72
2/2Y	11.96	1924.34	121.07	3.04	2434.71	66.31	-	-	-	4.24	-3132.82	4.72
3/2Y	11.96	1924.34	121.07	3.04	2434.71	66.31	3.43	4200.00	10.00	4.24	-3132.82	4.72

Figure 63 shows the values of the coefficients responsible for the identification of imbalance. In imbalance rotor case all scenarios, including the simplest scenario one, can correctly detect rotor unbalance coefficient. All scenarios have similar values for both sensors 1Y and 2Y, described in the figure as $1Y_{imbal}$ and $2Y_{imbal}$ respectively.

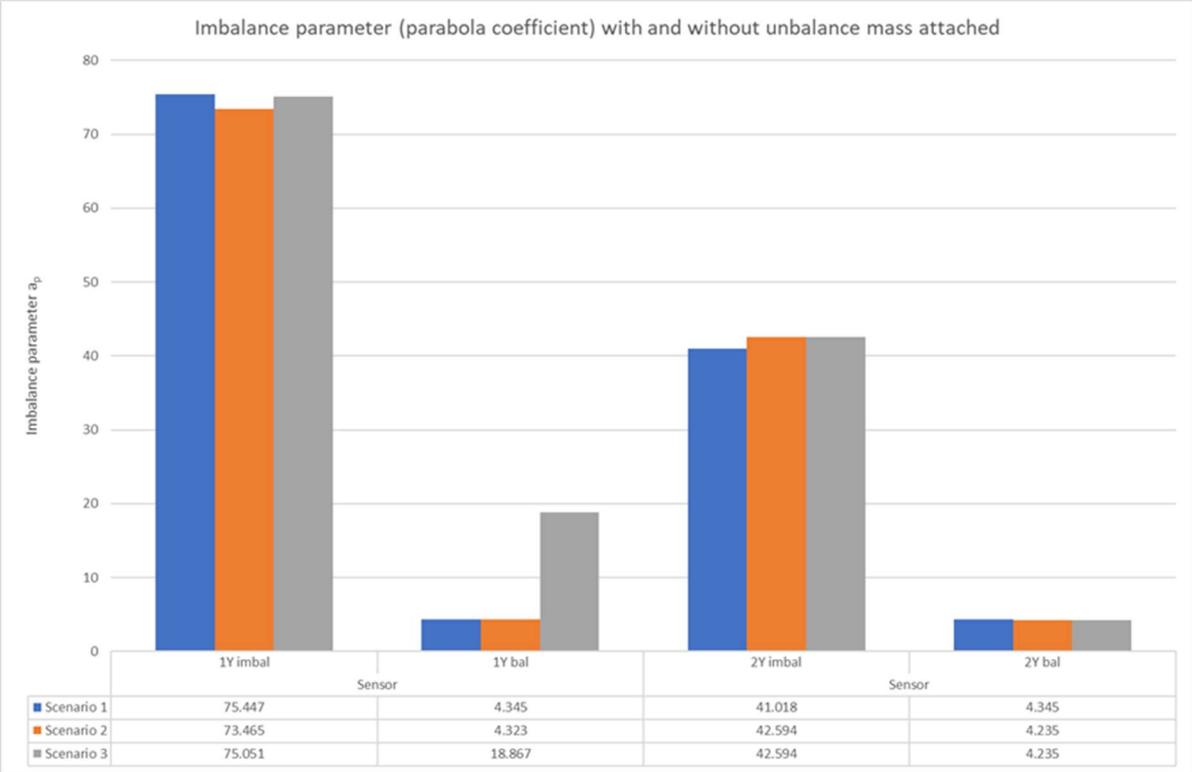


Figure 63. Imbalance coefficients for sensor 1Y and 2Y.

For a balanced rotor case, the unbalance coefficient in the scenario three for sensor 1Y has a higher value than the other scenarios. Although the RMSE index is smaller, and the imbalance coefficient is close to the actual value, the time and computing power needed to execute and evaluate this case may not be rationally justified. For such a simple case, i.e. only one resonance speed interval and no excessive imbalance, matching the function with the coefficients from scenario two should be sufficient and satisfactory.

The research and tests conducted on the test rig confirm the correctness of the assumptions of the MD3 method. The method can effectively identify at least one critical speed range and the failure in the form of rotor imbalance. The model has been positively verified. Moreover, the decomposed function parameters produced by the method reflect the actual mechanical values of a given object. Thus, it can be used to track changes that turbo-set undergoes during each transient condition

8.2. Validation of MD3 method on real turbine

The Multidimensional Data Driven Decomposition method was applied to the data from a real turbogenerator. The authors use the data measured on a 560MW steam unit in this case study. Figure 18 depicts a shaft-bearing line schematic representation. Based on the constant speed data, operational personnel reported high vibration levels in bearing number nine. Vibration measurements

were carried out to verify the cause of the high vibration. Data were recorded during transient operation (coast-down) of the unit. The portable data acquisition interface unit was connected to eddy-current type vibration displacement sensors at all nine bearings in both directions. Figure 9 shows the schematic and real-object sensor arrangement inside of bearing housing.

Transient data was recorded, and unbalance of the generator rotor free end (near the bearing number nine) was diagnosed. After the balancing operation, the data was measured once more during the run-up. The balancing operation was qualified as satisfactory. The turbogenerator was considered eligible for long-term operation with no restrictions in terms of dynamic condition to run within a full range of operation (referred to as the class A).

After the first measurement, the data were processed with the OpEn fault detection method. It detected a high level of synchronous response on bearing 9 in the Y direction. At the same time, it did not return any increased values of vibration amplitudes on bearing 8 in any direction. Lack of indication would typically eliminate bearing 8 data for the MD3 method. However, for this case study, these data were taken into account for comparison.

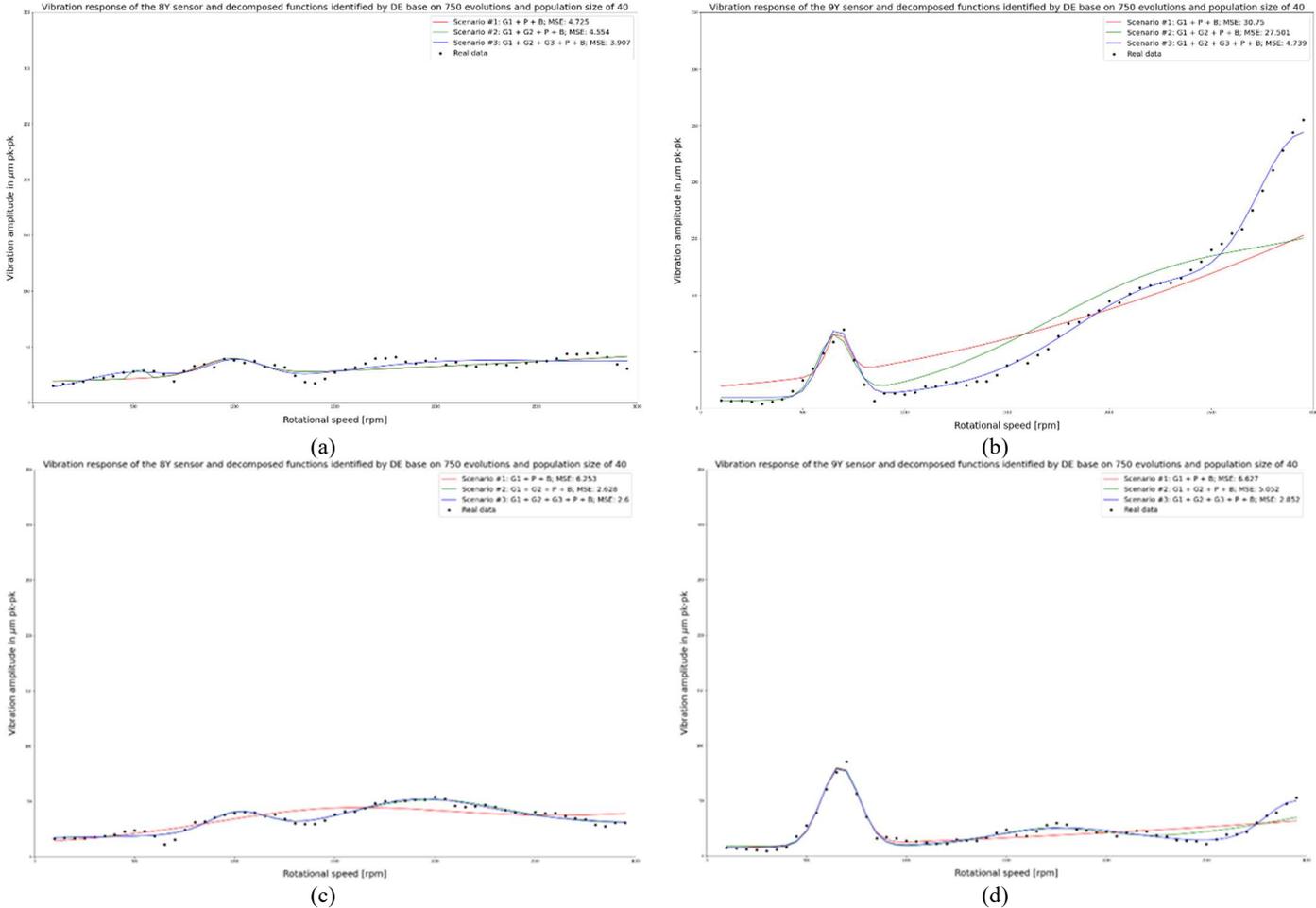


Figure 64. Decomposed function identified by the DE algorithm to fit data from a real object before and after balancing and the corresponding MSE indexes: (a) data from the bearing 8 and the sensor 8Y before balancing; (b) data from the bearing 9 and the sensor 9Y before balancing; (c) data from the bearing 8 and the sensor 8Y after balancing; (d) data from the bearing 9 and the sensor 9Y after balancing.

A high level of vibration amplitude on the one end of the rotor and a normal level of vibration response on the second end can be a symptom of the generator rotor unbalance in the vicinity of bearing 9. This hypothesis was later confirmed during corrective actions. Figure 64 presents all three scenarios identified by the DE algorithm according to the MD3 method. The RMSE index, as the decision criterion, selected scenario 3 as the best fit with the RMSE value of 4.74 (sensors 9Y). Scenario 3 was also the best fit for the transient without imbalance response approximation with the MSE index equal to 2.85.

Table 33

Table 33. The RMSE values based on the scenario and sensor location of the generator rotor with and without imbalance malfunction. summarize RMSE indexes for bearing 8 and bearing 9.

Table 33. The RMSE values based on the scenario and sensor location of the generator rotor with and without imbalance malfunction.

MSE	Sensor 8Y imbalance response	Sensor 8Y balance response	Sensor 9Y imbalance response	Sensor 9Y balance response
Scenario 1	4.725	6.253	30.75	6.627
Scenario 2	4.554	2.628	27.501	5.052
Scenario 3	3.907	2.6	4.739	2.852

Results returned for bearing 8 were different from those from bearing 9. As vibration levels were low for both the unbalanced and the balanced state, the functions identified by the DE algorithm had very similar decomposed function parameters. Thus the RMSE criterion in both cases had a low value. It also confirms that the algorithm was successful for bearing 8.

Table 8 summarizes all the decomposed function parameters depending on the scenario. For example, the functions which approximate the imbalance condition were high-lighted in scenario 3. Note that all scenarios satisfactory identified the first critical speed zone, which can be seen in Figure 64 (b).

Table 34. Decomposed function parameters identified by DE algorithm. Function parameters in imbalance and healthy condition. The MD3 scenario selection number is highlighted in bold font.

Scenario/ sensor-condition	a_{g1}	rpm_{g1}	w_{g1}	a_{g2}	rpm_{g2}	w_{g2}	a_{g3}	rpm_{g3}	w_{g3}	a_p	x_p	$const_b$
1/9Y-unbalanced	37.213	669.4	51.229	-	-	-	-	-	-	255.000	-492.1	15.481
2/9Y- unbalanced	53.401	655.9	80.750	47.058	2212.5	590.000	-	-	-	255.000	130.9	6.279
3/9Y- unbalanced	59.614	670.0	78.497	51.173	2212.5	403.799	128.925	2942.7	214.478	215.647	462.1	9.686
1/9Y-balanced	71.034	668.9	90.854	-	-	-	-	-	-	45.665	-1442.8	3.991
2/9Y-balanced	72.041	670.6	93.733	15.793	1706.9	271.219	-	-	-	85.197	1416.0	8.881
3/9Y-balanced	71.689	669.6	94.397	14.655	1779.0	329.603	36.022	2950.0	152.466	12.627	-2737.6	5.421
1/8Y-unbalanced	15.130	978.5	142.121	-	-	-	-	-	-	44.383	-2950.0	11.096
2/8Y- unbalanced	15.130	525.9	19.937	15.130	983.3	139.615	-	-	-	44.383	-2950.0	11.096
3/8Y- unbalanced	15.130	489.8	219.351	20.471	999.9	151.230	15.130	1995.7	589.995	44.383	-161.9	10.702
1/8Y-balanced	20.788	1475.0	590.000	-	-	-	-	-	-	50.268	-2950.0	3.867
2/8Y-balanced	18.769	1014.8	146.283	26.970	1947.1	407.326	-	-	-	24.457	-2605.7	13.500
3/8Y-balanced	17.810	1015.6	147.178	11.130	1903.7	272.631	17.601	1966.7	506.730	22.515	-2678.9	13.500

Scenario 1 has only one critical speed zone, parabola, and bias term in the model. Therefore, it cannot achieve a good fit to the real-object data, red line in Figure 64 (b). This model can find the peak of the

first critical speed rpm_{g_1} zone, but the amplitude a_{g_1} and width of the peak w_{g_1} are affected by parabola function (a_p, x_p) . Therefore, the parabola part of decomposed function cannot achieve a good approximation of such an excessive imbalance condition.

The second scenario can correctly replicate the first critical speed zone in all its features. However, due to the complexity of response in the rotational speed near FSNL, its performance was also not satisfactory, the green line in Figure 64 (b). The second critical speed zone and significant unbalance force made scenario two not sufficiently accurate.

Scenario 3 best approximates the real-object transient data, the blue line in Figure 64 (b). Thanks to three Gaussian functions in its model, it could replicate two critical speed zones and use the third one to enhance the model's performance to approximate additional nonlinearity introduced by the imbalance at the highest rotational speed values. This approximation of the unbalance condition resulted in the RMSE index being almost six times smaller than scenario 2 and seven times smaller than scenario 1. The particular RMSE index values for the imbalance condition are presented in Table 31 in a column titled "Sensor 9Y imbalance response".

Real-object data collected during the second measurement course (after balancing) revealed exciting results. For this case, each of the scenarios was a decent approximation of the healthy state of the machine. Figure 64 (d) shows that each scenario detected and identified the first critical rotational speed zone in all of its parameters consistently and in a convergent way. Furthermore, identified values of all decomposed function parameters for all scenarios concerning the first critical speed interval are almost identical.

Table 33 presents this in row 4÷6 and column 1÷3. Additionally, scenarios number two and three had better identified transient response between 1400÷2200rpm, and scenario three was superior to others in replicating the system response above 2500rpm. Also, in this case, scenario three was the best approximation of real-object transient data acquired from sensor number 9Y without an imbalance condition.

With this in mind and using the author's experience, a reasonable range of parameters of the decomposed functions can be determined for this type of failure. Table 35 presents the values selected as a range of search for decomposed function parameters.

Table 35. Healthy state operation decomposed function parameters by the DE algorithm.

Sensor	a_{g_1}	rpm_{g_1}	w_{g_1}	a_{g_2}	rpm_{g_2}	w_{g_2}	a_{g_3}	rpm_{g_3}	w_{g_3}	a_p	x_p	$const_b$
9Y	<80	670±30	<120	30	1800±50	<400	20	2950±50	<200	50	-	20

These values can be used as guideline parameters. The proper definition of this range can significantly reduce the time required by the DE algorithm to reach optimum.

9. Conclusions

9.1. Final remarks

Large turbogenerators are the heart of the power generation industry. They are designed and built for long-term operation with as few shut-down processes as possible. Sometimes, the turbo-set can be operated for months or even a year without a coast down. On the other hand, such a transient situation carries essential diagnostic information about a machine's condition. During such events, much information regarding the machine's condition is gone if not monitored and appropriately analyzed. Automation should be applied to facilitate the analysis of these valuable data. Up to now it was not possible, due to high cost of equipment and human expert.

The data-driven methods (OpEn and MD3) developed in this dissertation for the analysis and automatic diagnostics of failures are driven by the type and nature of data obtained during large turbomachinery measurements. Therefore, the methods proposed by the author in the doctoral dissertation are a compromise of the amount of available data and the accuracy/repeatability of the results. The dissertation is also a result of over 13 years of industrial practice combined with experience and expert knowledge in the field of signal processing, rotor dynamics and large turbomachinery.

The proposed Operational Envelope (OpEn) method can help the maintenance staff in machine operation and overhaul planning. *OpEn* is a novelty detection method that can be applied to the data taken during the transient state of a machine. Together with the OpEn algorithm, the author proposed a set of parameters that can be used to diagnose the transient automatically. Furthermore, those parameters can be used with other process data for better and more in-depth diagnostic purposes.

Two parameters called RMSE and "Max Out of OpEn" were shown as helpful in the automated detection of malfunctions. The other two may also be useful in the detection of other malfunctions. The *OpEn* 2D and *OpEn* 3D are an automated fault detection method for transient states. The 2D case analyzes only a single feature from a single sensor. The 3D case conjuncts two vibration signal features, i.e., synchronous amplitude and its phase. Novelty detection method proposed in the dissertation can be used to detect faults over different speed spans, different amplitudes during transient states, and different sets of sensors. All these factors make this method very flexible and a powerful tool in predictive maintenance schemes for many power facilities.

The Multidimensional Data Driven Decomposition (MD3) method proposed in this thesis is designed to identify machinery faults automatically. The author's novel approach to decompose the transient into several predefined signals, enables the analysis of individual dynamics system parameters becomes easier to evaluate and assess even to unqualified personnel. The decomposed transient components are responsible for particular failure modes and, as a consequence, not only can different malfunctions be detected, but they can also be identified. These parameters can be used to track and trend the evolution of the system dynamic response parameters without the engagement of the diagnostic teams. The MD3 method can assess data during each transient in contrast to portable equipment measurement that can miss the unplanned and sudden shut-downs and start-ups. The cornerstone of the method is to decompose a transient into a set of base functions. Such functions

have a simple form (Gaussian, parabolic or constant bias). Each such function has a mechanical meaning and can be used to diagnose and analyze transient responses collected during coast-downs and start-ups. The innovative MD3 method proposed in the article can increase the safety of the device and reduce the costs of electricity generation.

To tackle the problem of different content of transient data sets a set of models is used to fit the data. The best scenario selection strategy uses the MSE criterion to evaluate the three available models of decomposed function sets. The selection strategy is the ablation study of the MD3 method. This allows the MD3 method additionally increase the reliability of the method and reduce the risk of overfitting the model. Finally, the best model, which scenario has the lowest value of the MSE index, is used for the technical state assessment.

The Differential Evolution algorithm performance in terms of the time-to-transient fit ratio for all scenarios is investigated and presented. Input parameters of the DE for all scenarios are set up to:

Number of evolutions: 750;

Number of population: 40;

Crossover rate: 0.7;

Mutation rate: 0.8.

Both sections, Validation of Model Data and Case Study, confirm that the method can accurately pinpoint the type and magnitude of a particular fault. Based on the case study, the parameter responsible for the imbalance response was the a_p the coefficient in the decomposed function. In the real-object data case study, the MD3 method selected scenario three as the one with the best fitting capabilities for replicating the system's transient response. Often in Machine Learning research the ablation procedure is used to avoid the model overfitting. In our case, we achieve this goal by estimating parameters of several models of different complexity. Thus, we additionally increase the reliability of the method and reduce the risk of the model overfitting. Moreover, in the case study section, the authors provided a set of parameters to assess the technical condition of the rotor of a high-power generator. The parameters can be used as baseline parameters references to assess potential damage during transient states if the vibrations fell out of the acceptance region.

The Multidimensional Data Driven Decomposition (MD3) is an extension of the Data Driven Decomposition Method (D3), previously proposed by the authors in [14]. The multidimensional (multi-sensor) approach produces much better results than the analysis performed only with a single sensor (D3).

9.2. Author's contribution

During the process of the research, Author performed a list of tasks. Some were learning of new technologies (e.g. Python), the others were literature studies, to become familiar with the current state of the art. However, the majority of tasks were Author's original contribution to the field of signal

processing, focused on Fault Detection and Identification. These tasks were:

- Development and proposal of system architecture.
- Development of data preprocessing methods.
- Implementation the complete system in the Python environment.
- Analysis of the 200MW,360MW, 470MW, 560MW type turbines technical documentation (gathering of relevant knowledge about the dynamics of these types of machines, and assumptions' preparation for determining baseline measurements).
- Preparation of the simplified 1DOF model of turbogenerator shaftline dynamics [62].
- Analysis of the correct dynamic state databases and selecting reference data for baseline evaluation.
- Analysis of over 250 transient measurements databases.
- Selecting and ranking transients according to the methodology criterion (correct for baseline measurements; correct for the study: i.e., data contains potential malfunctions information; unusable data: from data points not covered by the test, bad quality data).
- Invention, creation, and development of the fault detection (OpEn) method.
- Establishing the upper and lower values for the OpEn 2D case and the ellipsis axis values for the OpEn 3D case.
- Establishing the severity parameters for the OpEn 2D and 3D cases.
- Establishing the upper and lower values for the OpEn 2D case and the ellipsis axis values for the OpEn 3D case.
- Invention, creation, and development of the fault identification (MD3) method.
- Proposition of the decomposed functions algebraic representations (Gaussian, parabola, bias/const).
- Defining the fitness functions for the Differential Evolution algorithm.
- Estimation of parameters of decomposed functions (adopting Python's DE algorithm code to find the decomposed function parameters).
- Planning, preparation and execution of experiment on a test rig.
- Validation of proposed methods on laboratory data.
- Validation of proposed methods on real object data.

Summing up, the Author developed complete set of methods, using data-driven approach, to automatically analyze the transient signals from large turbo sets. This allows to create the complete automated fault detection and identification system of large turbomachinery using Machine Learning approach. Such an achievement was the goal of this thesis, i.e. one can state that the goal of the dissertation was achieved.

The results of this dissertation can be used in FDI systems in commercial and industrial power plants as an autonomous diagnostic system. It can also be an extension and support for the existing diagnostic system, adding an element of automation to the diagnostic processes of the most critical machines.

9.3. Further research

During research, as in every scientific activity, several new challenges were identified. First, it was discovered that the method becomes unfeasible when more than 4÷6 transient responses are considered at once. The above findings led the authors to conclude that the MD3 analysis should be performed at particular rotor parts but not on the whole turbogenerator shaftline. Therefore, improving the method's performance and extending its multidimensionality capabilities should be the subject of further research.

Due to the size and complexity of the problem and the availability of data from real turbo sets, the author could validate the entire system for the unbalanced state only. The OpEn detection method has proven successful for other types of failure. However, due to the above, the author could not propose an algebraic representation of the decomposed functions and find the values of their boundary parameters for the MD3 method. It will be another direction of the author's research.

The author also plans to create methods that take into account other types of machine malfunctions. For example, the analysis of additional signal features, the overall vibration level can be used to detect sub-synchronous vibrations. These vibrations do not depend directly on the rotation of the turbine set, so they are not included in the harmonic analysis. However, the dominant value of the sub-synchronous components combined with the low values of the synchronous components may indicate the development of oil whirl in the bearing. Due to very high amplitude levels, these incidences can damage the machine's components.

During further research the author will research and validate the MD3 method for the rotor-to-stator rubs detection and assessment. He will also use a set of different signal features to detect other malfunctions. Additionally, the author plan to incorporate different DE strategies. It will involve different mutation and crossover rate definition proposed Ahmad et al. [63].

10. References

1. Wu, X.; Shen, J.; Li, Y.; Lee, K.Y. Steam Power Plant Configuration, Design, and Control. *WIREs Energy and Environment* **2015**, *4*, 537–563, doi:<https://doi.org/10.1002/wene.161>.
2. *Renewables 2020 Global Status Report*; REN21; (Paris: REN21 Secretariat), 2020; p. 40;.
3. Eisenmann, R.C., Jr. *Machinery Malfunction Diagnosis and Correction: Vibration Analysis and Troubleshooting for Process Industries*; 1st ed.; Prentice Hall: UpperSaddle River, US, 1997; Vol. 1;.
4. Bently, D.E.; Hatch, C.T. *Fundamentals of Rotating Machinery Diagnostics*; Grissom, B., Ed.; ASME Press: Minden, US, 2002; Vol. 1; ISBN 978-0-9714081-0-4.
5. Vo, D.-Q.; Ton-That, H.L. Free Vibration of Simply Supported Steel I-Girders with Trapezoidal Web Corrugations. *Reports in Mechanical Engineering* **2020**, *1*, 141–150, doi:10.31181/rme200101141v.
6. Vance, J.M.; Zeidan, F.Y.; Murphy, B.T. *Machinery Vibration and Rotordynamics* | Wiley; 2nd ed.; Wiley & Sons: New Jersey, USA, 2010; ISBN 978-0-471-46213-2.
7. Ehrich, F.F. *Handbook of Rotordynamics*; 1; 1st ed.; McGraw-Hill: New York, 1992; ISBN 0-07-019330-4.
8. Muszynska, A. *Rotordynamics*; CRC Press: Boca Raton, 2005; ISBN 978-0-429-13356-5.
9. Kiciński, J. *Rotor Dynamics*; 2nd ed.; Institute of Fluid-Flow Machinery Polish Academy of Science: Gdańsk, 2006; ISBN 83-7204-542-9.
10. Kiciński, J. *Dynamika Wirników i Łożysk Ślizgowych. Polish (Dynamics of Shafts and Hydrodynamic Bearings)*; Burka, E.B., Ed.; Szewalski Institute of Fluid-Flow Machinery.; Institute of Fluid-Flow Machinery Polish Academy of Science: Gdańsk, 2005; ISBN 83-88237-06-3.
11. Adams, M.L. *Rotating Machinery Vibration: From Analysis to Troubleshooting, Second Edition*; Silva, C.W. de, Ed.; 2nd ed.; CRC Press: Boca Raton, 2009; ISBN 978-0-429-13154-7.
12. Brito, L.C.; Susto, G.A.; Brito, J.N.; Duarte, M.A.V. An Explainable Artificial Intelligence Approach for Unsupervised Fault Detection and Diagnosis in Rotating Machinery. *Mechanical Systems and Signal Processing* **2022**, *163*, 108105, doi:10.1016/j.ymssp.2021.108105.
13. Barszcz, T.; Zabaryłło, M. Fault Detection Method Based on an Automated Operating Envelope during Transient States for the Large Turbomachinery. *J. vibroeng.* **2022**, *24*, 75–90, doi:10.21595/jve.2021.22165.
14. Barszcz, T.; Zabaryłło, M. Automatic Identification of Malfunctions of Large Turbomachinery During Transient States with Genetic Algorithm Optimization. *machine learning, fault detection, transient, turbine generator, genetic algorithm* **2022**, Vol. 29 (2022), 175–190, doi:DOI: 10.24425/mms.2022.138551.
15. Bielecki, A.; Bielecka, M.; Jabłoński, A.; Staszewski, W. Simple Method of Failure Detection of Rotary Machines. *Diagnostyka* **2021**, *22*, 17–22, doi:10.29354/diag/142862.
16. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. *IEEE Transactions on Industrial Electronics* **2016**, *63*, 3137–3147, doi:10.1109/TIE.2016.2519325.
17. Wang, H.; Sun, W. Motor Bearing Fault Diagnosis Based on Wavelet Packet Analysis and Sparse Filtering.; IEEE Xplore: IEEE 4th International Electrical and Energy Conference (CIEEC), 2021; pp. 1–7.
18. Simcenter Anovis | Siemens Software Available online: <https://www.plm.automation.siemens.com/global/pl/products/simcenter/simcenter-anovis.html> (accessed on 16 May 2022).
19. Industrial, S. Stork Industrial Equipment Supplier Available online: <https://storkindustrial.com/meggitt> (accessed on 16 May 2022).
20. ADRE Machine Data Acquisition and Diagnostics System | Bently Nevada Available online: <https://www.bakerhughes.com/bently-nevada/monitoring-systems/portable-vibration-testers-diagnostic-systems/adre> (accessed on 16 May 2022).

21. Noise and Vibration Dynamic Signal Analyzers Available online: <https://www.oros.com/solutions/instruments-accessories/> (accessed on 16 May 2022).
22. AMS 2600 Machinery Health Expert | Emerson GB Available online: <https://www.emerson.com/en-gb/catalog/automation/ams-a2600t8-en-gb> (accessed on 16 May 2022).
23. DAQ System - Analyzer Systems | Brüel & Kjær Available online: <https://www.bksv.com/en/instruments/daq-data-acquisition/analyzer-system> (accessed on 16 May 2022).
24. Schumaker, L. *Spline Functions: Basic Theory*; Cambridge Mathematical Library; 3rd ed.; Cambridge University Press: Cambridge, 2007; ISBN 978-0-521-70512-7.
25. Dyer, S.A.; Dyer, J.S. Cubic-Spline Interpolation. *IEEE Instrumentation Measurement Magazine* **2001**, *4*, 44–46, doi:<https://doi.org/10.1109/5289.911175>.
26. Gerald, C.F.; Wheatley, P.O. *Applied Numerical Analysis*; 7th ed.; Pearson Education: California Polytechnic State University; ISBN 0-321-13304-8.
27. Demetgul, M.; Yildiz, K.; Taskin, S.; Tansel, I.N.; Yazicioglu, O. Fault Diagnosis on Material Handling System Using Feature Selection and Data Mining Techniques. *Measurement* **2014**, *55*, 15–24, doi:<https://doi.org/10.1016/j.measurement.2014.04.037>.
28. Dworakowski, Z.; Dziejach, K.; Jabłoński, A. A Novelty Detection Approach to Monitoring of Epicyclic Gearbox Health. *Metrology and Measurement Systems* **2018**, *25*, 459–473, doi:<https://doi.org/10.24425/123896>.
29. Wang, Z.; Huang, H.; Wang, Y. Fault Diagnosis of Planetary Gearbox Using Multi-Criteria Feature Selection and Heterogeneous Ensemble Learning Classification. *Measurement* **2021**, *173*, 108654, doi:[10.1016/j.measurement.2020.108654](https://doi.org/10.1016/j.measurement.2020.108654).
30. Lis, A.; Dworakowski, Z.; Czubak, P. An Anomaly Detection Method for Rotating Machinery Monitoring Based on the Most Representative Data. *Journal of Vibroengineering* **2021**, *23*, 861–876, doi:[10.21595/jve.2021.21622](https://doi.org/10.21595/jve.2021.21622).
31. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Universal Domain Adaptation in Fault Diagnostics With Hybrid Weighted Deep Adversarial Learning. *IEEE Transactions on Industrial Informatics* **2021**, *17*, 7957–7967, doi:[10.1109/TII.2021.3064377](https://doi.org/10.1109/TII.2021.3064377).
32. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Open-Set Domain Adaptation in Machinery Fault Diagnostics Using Instance-Level Weighted Adversarial Learning. *IEEE Transactions on Industrial Informatics* **2021**, *17*, 7445–7455, doi:[10.1109/TII.2021.3054651](https://doi.org/10.1109/TII.2021.3054651).
33. Jabłoński, A.; Barszcz, T. Vulnerabilities and Fruits of Smart Monitoring. In *Smart Monitoring of Rotating Machinery for Industry 4.0*; Chaari, F., Chiementin, X., Zimroz, R., Bolaers, F., Haddar, M., Eds.; Applied Condition Monitoring; Springer International Publishing: Cham, 2022; pp. 1–9 ISBN 978-3-030-79519-1.
34. Capelli, L.; Massaccesi, G.; Molano, J.C.C.; Campo, F.; Borghi, D.; Rubini, R.; Cocconcelli, M. A Structured Approach to Machine Learning Condition Monitoring. In *Smart Monitoring of Rotating Machinery for Industry 4.0*; Chaari, F., Chiementin, X., Zimroz, R., Bolaers, F., Haddar, M., Eds.; Applied Condition Monitoring; Springer International Publishing: Cham, 2022; pp. 33–54 ISBN 978-3-030-79519-1.
35. Banaszkiwicz, M. Concept of Advanced Lifetime Monitoring System for Steam Turbines. *Diagnostyka* **2021**, *22*, 23–29, doi:[10.29354/diag/132548](https://doi.org/10.29354/diag/132548).
36. Zagorowska, M.; Ditlefsen, A.-M.; Thornhill, N.F.; Skourup, C. Turbomachinery Degradation Monitoring Using Adaptive Trend Analysis. *IFAC-PapersOnLine* **2019**, *52*, 679–684, doi:[10.1016/j.ifacol.2019.06.141](https://doi.org/10.1016/j.ifacol.2019.06.141).
37. Hanachi, H.; Mechefske, C.; Liu, J.; Banerjee, A.; Chen, Y. Enhancement of Prognostic Models for Short-Term Degradation of Gas Turbines. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM); June 2017; pp. 66–69.
38. Zohair, D.A.; Hafaifa, A.; Abdelhamid, I.; Abdellah, K. Gas Turbine Reliability Estimation to Reduce the Risk of Failure Occurrence with a Comparative Study between the Two-Parameter Weibull

- Distribution and a New Modified Weibull Distribution. *Diagnostyka* **2022**, *23*, 1–18, doi:10.29354/diag/146240.
39. *Diagnostics of New-Generation Thermal Power Plants*; Chmielniak, T., Trela, M., Eds.; 1st ed.; The Szwedzki Institute of Fluid-Flow Machinery: Gdańsk, 2008; ISBN 978-83-88237-31-7.
 40. Akhtar, M.; Kamran, M.S.; Hayat, N.; Rehman, A.U.; Khan, A.A. High-Vibration Diagnosis of Gas Turbines: An Experimental Investigation. *Journal of Vibration and Control* **2021**, *27*, 3–17, doi:10.1177/1077546320923917.
 41. *Handbook of Loss Prevention | SpringerLink*; Allianz-Versicherungs-AG, Ed.; 1; 1st ed.; Springer-Verlag Berlin Heidelberg: Berlin, 1978; Vol. 1; ISBN 978-3-540-07822-7.
 42. Bornassi, S.; Berruti, T.M.; Firrone, C.M.; Battiato, G. Vibration Parameters Identification of Turbomachinery Rotor Blades under Transient Condition Using Blade Tip-Timing Measurements. *Measurement* **2021**, *183*, 109861, doi:10.1016/j.measurement.2021.109861.
 43. Bachschmid, N.; Pennacchi, P.; Tanzi, E. Cracks in Rotating Shafts. In *Cracked Rotors: A Survey on Static and Dynamic Behaviour Including Modelling and Diagnosis*; Bachschmid, N., Pennacchi, P., Tanzi, E., Eds.; Springer: Berlin, Heidelberg, 2010; pp. 1–15 ISBN 978-3-642-01485-7.
 44. Leon-Garcia, A. *Probability and Random Processes for Electrical Engineering*; 2nd ed.; Addison-Wesley, 1994; ISBN 978-0-201-50037-0.
 45. Hajnaye, A.; Shirazi, K.H.; Aghaamiri, R. Vibration Measurement for Crack and Rub Detection in Rotors. *Metrology and Measurement Systems* **2020**, 65–80, doi:https://doi.org/10.24425/mms.2020.131719.
 46. Duan, R.; Zhou, J.; Liu, J.; Xu, Y. A Performance Degradation Prediction Approach for Turbo-Generator Bearing Considering Complex Working Conditions Based on Clustering Indicator and Self-Optimized Deep Learning Model. *Meas. Sci. Technol.* **2021**, *32*, 065103, doi:10.1088/1361-6501/abd366.
 47. Patel, S.P.; Upadhyay, S.H. Euclidean Distance Based Feature Ranking and Subset Selection for Bearing Fault Diagnosis. *Expert Systems with Applications* **2020**, *154*, 113400, doi:10.1016/j.eswa.2020.113400.
 48. Wei, Z.; Wang, Y.; He, S.; Bao, J. A Novel Intelligent Method for Bearing Fault Diagnosis Based on Affinity Propagation Clustering and Adaptive Feature Selection. *Knowledge-Based Systems* **2017**, *116*, 1–12, doi:10.1016/j.knosys.2016.10.022.
 49. Yu, K.; Lin, T.R.; Tan, J.W. A Bearing Fault Diagnosis Technique Based on Singular Values of EEMD Spatial Condition Matrix and Gath-Geva Clustering. *Applied Acoustics* **2017**, *121*, 33–45, doi:10.1016/j.apacoust.2017.01.023.
 50. Chui, C.K.; Mhaskar, H.N. Signal Decomposition and Analysis via Extraction of Frequencies. *Applied and Computational Harmonic Analysis* **2016**, *40*, 97–136, doi:10.1016/j.acha.2015.01.003.
 51. Cicone, A. Nonstationary Signal Decomposition for Dummies. *Advances in Mathematical Methods and High Performance Computing* **2019**, *41*, 69–82, doi:https://doi.org/10.1007/978-3-030-02487-1_3.
 52. Rao, S.S. *Mechanical Vibrations*; 6th ed.; Pearson: University of Miami, 2017;
 53. de Boor, C. *A Practical Guide to Spline*; Springer New York: New York, 1978; Vol. Volume 27;
 54. Koza, J.R.; Poli, R. Genetic Programming. In *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*; Burke, E.K., Kendall, G., Eds.; Springer US: Boston, MA, 2005; pp. 127–164 ISBN 978-0-387-28356-2.
 55. Roetzel, W.; Luo, X.; Chen, D. Chapter 6 - Optimal Design of Heat Exchanger Networks. In *Design and Operation of Heat Exchangers and their Networks*; Roetzel, W., Luo, X., Chen, D., Eds.; Academic Press, 2020; pp. 231–317 ISBN 978-0-12-817894-2.
 56. Li, H.; Yuan, D.; Ma, X.; Cui, D.; Cao, L. Genetic Algorithm for the Optimization of Features and Neural Networks in ECG Signals Classification. *Sci Rep* **2017**, *7*, 41011, doi:10.1038/srep41011.
 57. Storn, R.; Price, K. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* **1997**, *11*, 341–359, doi:10.1023/A:1008202821328.

58. Qin, A.K.; Huang, V.L.; Suganthan, P.N. Differential Evolution Algorithm With Strategy Adaptation for Global Numerical Optimization. *IEEE Transactions on Evolutionary Computation* **2009**, *13*, 398–417, doi:10.1109/TEVC.2008.927706.
59. Das, S.; Mullick, S.S.; Suganthan, P.N. Recent Advances in Differential Evolution – An Updated Survey. *Swarm and Evolutionary Computation* **2016**, *27*, 1–30, doi:10.1016/j.swevo.2016.01.004.
60. Muratoglu, A.; Tekin, R.; Ertuğrul, Ö.F. Hydrodynamic Optimization of High-Performance Blade Sections for Stall Regulated Hydrokinetic Turbines Using Differential Evolution Algorithm. *Ocean Engineering* **2021**, *220*, 108389, doi:10.1016/j.oceaneng.2020.108389.
61. Muratoglu, A.; Yuce, M.I. Design of a River Hydrokinetic Turbine Using Optimization and CFD Simulations. *Journal of Energy Engineering* **2017**, *143*, 04017009, doi:10.1061/(ASCE)EY.1943-7897.0000438.
62. Barszcz, T.; Czop, P.; Zabaryłło, M. Development and Tuning of a Simplified 1D Model for Generation of Transient States in Large Turbomachinery. In Proceedings of the Proceedings of the 13th International Conference on Damage Assessment of Structures; Wahab, M.A., Ed.; Springer: Singapore, 2020; pp. 541–554.
63. Ahmad, M.F.; Isa, N.A.M.; Lim, W.H.; Ang, K.M. Differential Evolution: A Recent Review Based on State-of-the-Art Works. *Alexandria Engineering Journal* **2022**, *61*, 3831–3872, doi:10.1016/j.aej.2021.09.013.

List of Figures

Figure 1. Power plant machine hall: top – old arrangement; bottom – new arrangement. 12

Figure 2. Turbine hall of 13K215 units installed in a power plant (top) and cross-section of the turbine type (bottom). 13

Figure 3. Intermediate Pressure (IP) turbine rotor: top - disassembly of the old one; bottom - new one assembly..... 14

Figure 4. Steam turbine parts: (a) – HP cylinder assembly; (b) – 200MW class LP rotor assembly; (c) – two LP rotors of +450MW class unit in situ. 16

Figure 5. Typical accelerometer sensor: top – real sensors examples; bottom – sensor components schematic..... 18

Figure 6. Typical velocity sensor: top left – real sensors examples, top right – on-site velocity sensor assembly; bottom – sensor components schematic..... 19

Figure 7. Arrangement of eddy-current probes in the bearing housing; top– schematic, middle left – physical sensors example, middle right – typical sensor components; bottom – field assembly on a journal bearing. 21

Figure 8. Relative expansion sensor assembly..... 22

Figure 9. Eddy-current probe installation example; top – schematic arrangement by Eisenmann [3], bottom – physical assembly of the eddy-current sensors inside thrust bearing pedestal for different type of measurements. 23

Figure 10. Once-per-revolution event. Timing difference and rotational speed measurement [4]. 24

Figure 11. Phase lag angle, angular velocity, and amplitude extraction procedure [4]..... 25

Figure 12. Reference mark and vibration signal during field measurement..... 26

Figure 13. The Jeffcott rotor model (a) [6]; End view of the Jeffcott rotor and its coordinates (b) [6].27

Figure 14. Bode plots of the rotor response phase (top described by equation (5)) and amplitude (described by equation (3), (4), and (6)) [8]. 28

Figure 15. Bodé plot of the real system. 29

Figure 16. (a) – centrifugal force simulation; (b) – real data example of the unbalance rotor system response. 30

Figure 17. The layout of a 200MW type turbo-set. The bearings are numbered from the HP side (the figure prepared by the authors)..... 37

Figure 18. The layout of a +500MW type turboset. 37

Figure 19. Shaftline inertia curve (coast down example) of the 200MW class unit. 39

Figure 20. Start-up curve of the 200MW class unit..... 39

Figure 21. Examples of DAIUs from leading manufactures. The top: left – Siemens [18], right – Meggitt [19]; in the middle: left - Bently Nevada [20], right – OROS [21]; in the bottom: left – Emerson [22], right - Bruel and Kael [23]..... 42

Figure 22. An example of the vibration measurements setup – acquisition units (two DAIUs) connected to the stationary monitoring system in the control room. 43

Figure 23. ADRE Sxp dynamic channels setup..... 45

Figure 24. One-per-revolution reference mark configuration tabs. 46

Figure 25. Portable equipment diagnostic features capabilities..... 49

Figure 26. Coast down chart examples.	50
Figure 27. Transient data structure	52
Figure 28. Transient data points for one sensor.	53
Figure 29. Transient data points for all sensors at single rotational speed instance.....	53
Figure 30. Non-equidistant spaced transient data points.....	54
Figure 31. Raw transient data example: top – whole transient data point set; bottom – zoom of the transient representative interval.	56
Figure 32. Examples of different settings of the spacing parameter interval between the consecutive rotational speed instances: (a) spacing: 25RPM; (b) spacing: 50RPM; (c) spacing: 100RPM; (d) spacing: 150RPM.	59
Figure 33. Automatic anomaly detection flowchart method.....	63
Figure 34. OpEn and its upper and lower values.....	64
Figure 35. Example of reference transient data set with OpEn values given.	65
Figure 36. OpEn 2D baseline measurements flowchart.....	67
Figure 37. Severity parameter visualization versus real data values: top – RMSE norm; bottom – “Min Out of OpEn” and “Max Out of OpEn” values.	71
Figure 38. OpEn 3D baseline measurements flowchart.....	73
Figure 39. Baseline vector examples.	74
Figure 40. Example of OpEn 3D ellipsis set: top - mean baseline and ellipse at each speed point during transient; bottom - 3D view (including turbogenerator rotational speed) of the ellipse set for the whole transient state for one sensor only.	75
Figure 41. OpEn 3D acceptance region and severity parameter definition.....	76
Figure 42. OpEn 3D ellipse severity parameter visualization: top – actual data inside of an ellipse; bottom – actual data outside of an ellipse.....	77
Figure 43. The MD3 method flowchart.....	84
Figure 44. Example of the decomposed signals simulating runout, resonance and unbalance and their superposition.....	86
Figure 45. Differential Evolution (DE) algorithm scheme.....	88
Figure 46. Basic monitoring and diagnostic system arrangement.....	89
Figure 47. Flowchart of the proposed system.....	90
Figure 48. Monitoring-protection and diagnostic systems in power plants supported by the automated FDI system (OpEn & MD3 modules).....	91
Figure 49. Example of reference one transient data for whole turboset.	116
Figure 50. Transient data during measurements course. Transients 01-04 were non-satisfactory. Transients 05-10 were satisfactory. Transient 09 was selected as the reference.	120
Figure 51. Example of a transients with misalignment (the initial state – transient no.1 and no.2 and after first improvement – transient no.3 and no.4).....	121
Figure 52. Example of acceptable dynamic behavior.....	122
Figure 53. The evolution of the Root Mean Square Error (RMSE) vs HP-IP alignment incremental improvement.....	123

Figure 54. Graphical representation of the OpEn 3D distance matrix: 7Y sensor before balancing attempts.	128
Figure 55. Graphical representation of the OpEn 3D distance matrix: 7Y sensor before balancing attempts.	132
Figure 56. 7Y vibration vector data distance from the acceptance region evolution during corective actions.	133
Figure 57. Polynomial data fit to the vibration data vector during consecutive transient.	134
Figure 58. Scheme of the test rig used for validation purpose.	135
Figure 59. The Rotor Kit test rig used for model validation.	136
Figure 60. The transient vibration response of the test rig during with and without unbalance, respectively: (a) data from the bearing number 1; (b) data from the bearing number 2.....	136
Figure 61. Outcome of the MD3 method based on three scenarios proposed in Chapter 5 on a test rig with imbalance malfunction simulation: (a) Data obtained from the 1Y sensor, driven side of the rotor; (b) Data obtained from the 2Y sensor, non-driven side of the rotor.	137
Figure 62. MD3 method outcome based on three scenarios proposed in Section 2 on a test rig without mal-function: (a) Data obtained from the 1Y sensor, driver side of the rotor; (b) Data obtained from the 2Y sensor, non-driven side of the rotor.	138
Figure 63. Imbalance coefficients for sensor 1Y and 2Y.....	139
Figure 64. Decomposed function identified by the DE algorithm to fit data from a real object before and after balancing and the corresponding MSE indexes: (a) data from the bearing 8 and the sensor 8Y before balancing; (b) data from the bearing 9 and the sensor 9Y before balancing; (c) data from the bearing 8 and the sensor 8Y after balancing; (d) data from the bearing 9 and the sensor 9Y after balancing.	140

List of Tables

- Table 1. Rotational speed simulation parameters. 30
- Table 2. Dynamic card configuration table. 47
- Table 3. Configuration of one-per-revolution reference mark. 48
- Table 4. Differences between OpEn and standard Spectrum Envelope. 62
- Table 5. Python libraries used in research. 97
- Table 6. Cloud resources used in the research (uploaded by the author from his portable data acquisition unit)..... 98
- Table 7. Function retrieving and preprocessing the transient data set from given location..... 98
- Table 8. Lowest and highest RPM value in transient. 99
- Table 9. CS interpolation procedure. 100
- Table 10. Fncion returning list of all transient data set file names within the folder..... 100
- Table 11. Parabola and Gaussian functions - the decomposed functions. 101
- Table 12. Three models used in MD3 method “tournament”. 101
- Table 13. The MD3 method the fitness function quality parameter. 102
- Table 14. Upper and lower search limits boundries for the DE algorithm..... 102
- Table 15. Baseline measurements functions. 102
- Table 16. Baseline measurements procedure: calculating and saving in the cloud. 105
- Table 17. Baseline data load (for the OpEn 3D case) and creating baseline ellipsis coordinates..... 106
- Table 18. OpEn procedure with severity parameter calculation and charts plotting (for the OpEn 2D case)..... 107
- Table 19. OpEn 3D method functions. 110
- Table 20. OpEn procedure with severity parameter calculation and charts plotting (for the OpEn 3D case)..... 111
- Table 21. MD3 method procedure example. 112
- Table 22. Baseline measurement matrix with an synchronous amplitude in polar coordinate system. 117
- Table 23. Baseline measurement matrix with a phase angle of the first component in polar coordinate system. 117
- Table 24. Baseline measurement matrix with vibration wector in Cartesian coordinate system: x-coordinate. 118
- Table 25. Baseline measurement matrix with vibration vector in Cartesian coordinate system: y-coordinate. 118
- Table 26. Comparison of the OpEn 2D method selection criteria..... 122
- Table 27. OpEn 3D distance matrix: left – transient before balancing attempt; right – after trial mass balancing. 126
- Table 28. OpEn 3D distance matrix: left – trim balance attempt; right – start-up for the long-term operation..... 130
- Table 29. The MSE values are based on the scenario and sensor location for the imbalance simulation on the test rig. 137

Table 30. Parameters of the decomposed functions identified by the DE algorithm and chosen by the MD3 method based on the imbalance data and the MSE as the quality index.....	137
Table 31. The RMSE values, based on the scenario and sensor location, for the reference transient simulation on the test rig.	138
Table 32. Parameters of the decomposed functions identified by the DE algorithm and chosen by the MD3 method based on the reference data and the RMSE as the quality index.....	138
Table 33. The RMSE values based on the scenario and sensor location of the generator rotor with and without imbalance malfunction.	141
Table 34. Decomposed function parameters identified by DE algorithm. Function parameters in imbalance and healthy condition. The MD3 scenario selection number is highlighted in bold font.	141
Table 35. Healthy state operation decomposed function parameters by the DE algorithm.	142