University of Strathclyde Engineering

Automated Interpretation of Boiler Feed Pump Vibration Monitoring Data

Submitted in partial fulfilment with the requirements of the degree:

MSc in Energy Systems and the Environment

Radek Kaczorowski

December 2008

Supervisors:

Dr Graeme West

Professor Stephen McArthur

Copyright Declaration

Abstract

The aim of this project was to analyze data provided by British Energy in order to design and implement a system, which would signal abnormal states of operation. The received data was for a boiler feed pump turbine and consisted of peak-to-peak vibration and governor valve position values.

Boiler feed pumps are significant and expensive power plant components and their failures may lead to serious financial and safety consequences, therefore an alarm system signalling potential abnormalities would reduce those risks.

The project involved applying KDD process to the data in order to attempt to distinguish abnormal operational states without having any additional knowledge about the data in order to remove possible bias. Initial analysis, which involved using Weka machine learning software and applying clustering to the pre-processed data sets, showed that two vibration sets are highly correlated over the whole range. Unfortunately, this analysis did not provide any results which could be used to create an alarm system.

After receiving additional data, the second analysis was performed. This time, it was focusing also on the time aspect, not only the data spread. The new algorithm was designed to investigate data on a point-by-point basis and compare difference between trends of two vibration data sets. A fragment of the data set was used to establish acceptable limits, which were imposed in the remaining data and would trigger an alarm if difference between two trends of vibration data sets was outside that limit.

Further testing of the algorithm on other data sets showed that it can correctly flag any potentially suspicious states as warnings, which should be investigated by the machine operator.

Recommendations for future work include further testing with different data sets and machines. Should the algorithm is verified by the engineers that signalled warnings indeed correspond to abnormal state of operation, then it is recommended to develop a full software application, which could monitor the condition of the machine online.

Table of contents

Copyright Declaration........1
Abstract........2
Table of contents........3
1. Project definition........5
1.1 Brief........5
1.2 Aim........5
1.3 Objectives and deliverables........5
1.4 Conclusions summary........6
2. Diagnostics........7
2.1 Importance of diagnostics........7
2.2 Types of maintenance........10
2.2.1 Reactive maintenance........11
2.2.2 Preventive maintenance........12
2.2.3 Predictive maintenance........13
3. Boiler feed pumps........15
3.1 Description........15
3.2 Stages of boiler feed pumps........15
3.3 Construction........16
3.3.1 Classic boiler feed pump........16
3.3.2 Advanced boiler feed pump arrangement........16
3.4 Design and arrangement........19
3.5 New developments........21
4. Condition monitoring examples........22
4.1 Expert systems........22
4.2 Machine learning........23
4.2.1 Artificial neural networks........24
4.2.2 Evolutionary algorithms........25
4.2.3 Bayesian learning........25
4.2.4 C4.5 algorithm........26
4.2.5 Case-based reasoning........26
4.3 Condition monitoring methods summary........26
5. Data analysis techniques........28
5.1 Knowledge discovery from databases........28
5.1.1 KDD and traditional methods........28
5.1.2 Data mining and KDD........29
5.1.3 The steps of KDD........33
5.2 Data mining........35
5.2.1 Data mining description........35
5.2.2 Clustering........36
5.2.2.1 Iterative clustering........37
5.2.2.2 Incremental clustering........38
5.2.2.3 Probability-based clustering........40
5.2.2.4 The EM model........42
5.2.2.5 Bayesian clustering........43
5.2.3 Data mining summary........44
5.3 Weka........45
6. Case study........46

6.1 Initial analysis attempt 46
6.1.1 Data description 46
6.1.2 Data distribution 50
6.1.3 Clustering 53
6.1.3.1 First file 53
6.1.3.2 Second file 56
6.1.3.3 Clustering summary 57
6.1.4 Combining the data 57
6.1.4.1 Distribution of combined data 57
6.1.4.2 Clustering combined data 59
6.1.4.3 Combined data analysis summary 60
6.1.5 LP/HP ratio 60
6.1.5.1 LP/HP ratio data distribution 61
6.1.5.2 LP/HP ratio data clustering – file 1 63
6.1.5.3 LP/HP ratio data clustering – file 2 65
6.1.6 LP/HP ratio – combined data 66
6.1.7 Initial analysis summary 68
6.2 Final analysis 69
6.2.1 Data description 69
6.2.2 Difference data 73
6.2.2.1 Difference between trends 75
6.3 Validation 80
6.4 Summary of Excel implementation 83
7. Recommendations and conclusions 85
7.1 Future recommendations 85
7.2 Conclusions 86
References 88
Appendix 1 – additional figures 90

1. Project definition

1.1 Brief

In the power generation industry, any failures may lead to blackouts which can have very severe impact on many (or actually most) aspects of our life and can cripple many services, which are crucial to our well-being. Therefore, minimizing outages and preventing plant equipment failures is becoming a critical issue for power plant operators. While condition monitoring of steam or gas turbines is quite popular, boiler feed pumps do not seem to be receiving the same level of attention from engineers.

Boiler feed pump failures may have catastrophic impact on the power plant operation, therefore it was decided to investigate a possibility of developing a system, which would trigger a warning in case of suspicious machine behaviour.

1.2 Aim

The aim of the project was to analyze provided boiler feed pump data in order to develop an alarm system which signals a warning in case of abnormal state of operation and, consequently, can be used to improve the maintenance regime.

1.3 Objectives and deliverables

The initial main objectives of the project were divided into following areas:

1) Background research – investigate condition monitoring and data mining techniques

2) Data analysis – find patterns in data

3) System design – create an alarm system based on data analysis

Due to the fact that the project was evolving mainly due to data availability and initial analysis results the objectives were rather flexible. However, the final deliverables are reasonably close to the initial objectives:

1) Background research
    a. State-of-the-art rotating machinery condition monitoring techniques and technologies
    b. Data analysis techniques and processes applicable to obtained data
2) Data analysis
    a. Initial analysis showed that looking at whole data does not provide any patterns
    b. Final analysis resulted in finding a pattern based on differences in changing of two data sets
    c. The identified pattern has been verified to work on other data sets
3) System design
    a. A system which calculates static limits basing on current and historical data
    b. Validation of the system on another data set

1.4 Conclusions summary

The case study showed that looking at data set as a whole and taking into account its spread does not provide satisfactory results.

Final analysis showed that it was necessary to analyze the data taking into account the time factor. The algorithm used to calculate limits basing on comparison of trends between two signals proved to be successful and it was tested (and verified) on the available data.

2. Diagnostics

In this section the reader is introduced to condition monitoring and how it can impact the operation of the plant item.

2.1 Importance of diagnostics

Diagnostics is a term very familiar to everyone. It is used referring to different everyday aspects of normal life – expressions like medical or car diagnostics are self-explanatory and there is no need to present elaborate definitions, because it is common knowledge. This work will be focused on technical diagnostics, therefore it is required to provide a detailed definition and explanation of this term.

Branch of science called diagnostics deals with identifying the examined state of things through classification into known types, through complete and causal explanation and determining the present state and predicting future development. From this very general definition it can be concluded that diagnostics presents the current state of the examined object and gives us some guidelines regarding its future behaviour. According to Koscielny (2001) definition of technical diagnostics is rather similar, just narrower: it is a field of knowledge including the whole of theoretical and practical issues concerning identification and evaluation of present, past and future states of a technical object taking into consideration its environment.

Definitions given above are very general and do not provide exact clarification how to approach diagnostic problems. This is because it is a very broad field of study and involves various engineering methods which help engineers reaching certain goals. In different branches of industry numerous tools are used when dealing with diagnostics and it would not be possible to include all of them in a short and comprehensible definition.

After familiarization with basic explanation of diagnostics, some more detailed information can be provided. As it was mentioned above, diagnostics deals with past, present and future state of a technical object, therefore according to Koscielny (2001) three following forms can be distinguished:

- Diagnostics: process, whose aim is to determine the current technical state of the examined object. Result of this process is establishing a diagnosis. Exact identification of state of a machine and its elements is necessary in order to perform repairs which will remove all the defects causing the inefficiency
- Genesis: process, whose aim is to establish in what way the states of machine elements were changing from a given time to present moment. It allows to determine the reason of a failure or inefficiency and helps with preventing similar problems in the future
- Prediction: process, whose aim is to determine the state of a machine in future from present time. It is possible only if certain factors are known: state of the machine when diagnosis is set, intensity of certain phenomena causing changes of machine's technical state and also probability of occurrence of unpredictable failures. If those factors are well known and prediction time is short, then there is a chance that the forecast will be correct. However, when above factors are less known and prediction time is longer, then the forecast becomes less accurate.

The above forms are common things which are encountered in everyday life and which seem obvious, but an average person would not be able to define them precisely. They are inevitably connected and dependant on each other.

When dealing with complicated problems where many symptoms must be analysed, very often it is hard to determine the primary cause of defect and analysts realize how diagnostic process can be difficult. Technological progress on one hand allows to perform more accurate measurements, but on the other hand introduces even more complex machinery.

As it was mentioned above, diagnostics is a branch of science which is not easily defined and due to complexity of encountered problems there are not any strict rules of applying a solution, but there are some guidelines, which prove to be very useful in practical situations. According to Koscielny (2001) when dealing with faults, there are three phases, which should be distinguished:

- fault detection – noticing the existence of fault and a moment in time, when it occurred
- fault isolation – determining type and position of fault
- fault identification – defining scale and character of fault variability in time

Among terminological suggestions it is worth mentioning those developed by Safeprocess Committee (according to Koscielny, 2001):

- monitoring – task performed in real time aimed at collecting and processing variables and recognizing abnormal behaviours (indicating alarms)
- supervision – monitoring the object and undertaking actions to maintain its regular operation when faults occur
- protection – actions and technical means eliminating potentially dangerous course of process or preventing effects of such a course

As mentioned by Koscielny (2001), in the recent decades the development of technological installations was enormous, which had to be followed by increased numbers of various sensors. Even though modern systems are characterized by very high reliability, the complexity still leads to inevitable failure of machine elements sometimes caused by the so-called human factor. It causes long-term disturbances or even stops the technological process. As a consequence, industry suffers great economic losses due to reduced production. In extreme cases, process failures may cause environmental pollution or danger to human life.

That is why role of diagnostics is so important in modern industry. Another very difficult problem is possible confusion of operator, when sensor system detects failures and starts sending alarm signals. Quick reaction of the operator might be difficult due to information overflow. Consequently, this may result in errors causing more failures.

Possibility of occurrence of such situations forces the development of diagnostic systems and introduction of process protection. Computerized systems make work of operators easier and provide additional safety measures, but cannot absolutely replace human links. Thanks to rapid information flow, failure detection and isolation takes less time. It protects from dangerous process variations and prevents economic losses.

According to Koscielny (2001) another very important role of diagnostics is decreasing repair costs. In most industry plants, reviews of process and sensor devices are periodic. This means that regardless of technical state, process must be stopped and all the devices reviewed. In most cases repairs are not necessary, because state of machines is good. Introducing automatic diagnostics and repairs based on current technical state results in great decrease of costs in comparison with periodic reviews.

2.2 Types of maintenance

Maintenance is a very important element of machine operation. Generally, it is performed when there is a problem with equipment – steps must be taken to fix it. However, maintenance means to keep something in good condition, what implies that it should be performed as a preventive action in order to prevent a piece of machinery from breaking, on equipment which does not need repair yet.

Unfortunately, according to Piotrowski (2001) it is still common for many private and public companies not to maintain the equipment properly and just fix or replace it after it is broken. This can have a negative effect on the operation of whole plant.

Most pieces of machinery have a certain life expectancy and maintenance suggestions (adjustments, lubrication, parts replacement, etc.) – this information is supplied by the manufacturer. Neglecting the suggested maintenance regime may have a negative impact on life expectancy of the equipment and may cause safety and financial risk.

Various maintenance approaches have been developed by the industry in order to make sure that equipment will reach or even exceed its life expectancy without problems.

Further in this section following types of maintenance will be discussed: reactive, preventive and predictive.

According to Table 1, in the United States of America reactive maintenance is still the most popular:

Type of maintenance	Popularity [%]
Reactive	>55%
Preventive	31%
Predictive	12%
Reliability centred/other	2%

Table 1. Types of maintenance and their popularity according to Piotrowski (2001)

2.2.1 Reactive maintenance

This is the simplest approach, which basically focuses on using a piece of machinery until it is broken. No actual maintenance actions take place in order to improve the condition of equipment.

Reactive maintenance has strong pros, which can also be perceived as cons in certain situations. In case of brand new machinery, it is expected to run practically flawlessly without any problems. Applying reactive maintenance means the company is not going to spend any resources on looking after the installed equipment.

This approach can save a significant amount of time and money according to Piotrowski (2001), assuming there are no failures. Ignoring manufacturer's maintenance suggestions causes the design life expectancy to be lower and it should be expected that quicker replacement of this piece of machinery might be required.

The biggest problem with reactive maintenance is that failure of one piece of machinery may impact the operation or even lead to damage of other plant equipment. In this situation, the cost increase due to unexpected downtime and capital investment into new equipment may be very significant. Man-hour related expenses are expected to increase dramatically due to overtime work (the failure may occur during unsocial hours) required to fix the problem as soon as possible.

Additionally, such situations cause significant safety risks for operators, who due to reactive maintenance regime are not aware of current condition of equipment they are working with.

Reactive maintenance can be applied to machinery which is easy and quick to replace, its failure will not impact other equipment or process and replacement parts can be stored on site. Otherwise inventory cost may prove this solution to be unfeasible.

2.2.2 Preventive maintenance

According to Piotrowski (2001) the second type of maintenance is called preventive maintenance. This approach focuses on performing maintenance actions in regular calendar time or machine operation time intervals in order to detect and prevent degradation of equipment condition. In practice, preventive maintenance controls and can even extend design life of a product by maintaining the degradation at an acceptable rate. Additionally, regular maintenance actions usually increase (or maintain at high level) the efficiency of maintained machinery.

Preventive maintenance approach follows the maintenance guidelines supplied by product manufacturer and helps maintain high reliability and design life expectancy. According to Piotrowski (2001), applying this type of maintenance may result in a significant cost decrease in comparison with reactive maintenance – usually by 12% to 18%. The amount of money saved can be much higher and depends on several factors – previous maintenance regime, equipment reliability and plant downtime.

Preventive maintenance is generally much better than reactive maintenance and allows better control over assets. Plant downtime can be adjusted flexibly in order to minimize costs and maximize maintenance work done. Additionally, this saves energy and is a more environmentally friendly approach since the number of catastrophic failures is reduced.

The main disadvantage of preventive maintenance is that unnecessary actions may be performed and, consequently, human effort and money can be wasted.

Preventive maintenance does not absolutely eliminate catastrophic failures, but it dramatically decreases their frequency in comparison with reactive maintenance.

2.2.3 Predictive maintenance

The next type of maintenance is predictive maintenance which takes into account the actual condition of the equipment based on measured degradation of the parts and indicates present and predicted future state of monitored machinery.

As opposed to preventive maintenance, action is only taken when the measurement indicates that it is required rather than on a regular basis (as in case of preventive maintenance). According to Piotrowski (2001) applying preventive maintenance approach, sometimes machine elements in perfect condition (which could be used for much longer without any danger of failing) will be replaced just because of the maintenance schedule. Therefore, the main advantage over preventive maintenance is that wasted human effort and wasted expenses are reduced to minimum.

Predictive maintenance focuses on analyzing the state of machinery before taking any action. If the condition is still satisfactory, then there is no need to replace any elements. Predictive maintenance is used to define which maintenance tasks are required – this approach dramatically reduces unneeded tasks and downtime.

A properly setup predictive maintenance system can practically eliminate catastrophic failures. Additionally, knowing the condition of machinery, maintenance tasks can be scheduled in most convenient time and plant downtime can be minimized. Consequently, there will be no need to store spare parts, because using the maintenance system data, it is possible to order required parts in advance, what will reduce inventory costs. According to Piotrowski (2001), upgrading from preventive to predictive maintenance system can on average save between 8% and 12%. However, depending on many factors, those savings can be as high as even 40%.

Knowing the current condition of plant equipment will increase plant reliability and functioning. The components will be replaced when their condition is below satisfactory. Therefore, quite often this time can be much longer than the design life, what provides significant savings. Additionally, due to the fact that equipment condition is known at all times, safety risks are reduced dramatically and that may have a positive impact on staff morale, who know they are working in a reliable environment.

The main disadvantage of predictive maintenance is initial capital cost, which may be quite significant – depending on the sophistication of the machinery and monitoring equipment. In order to take full advantage of the system, staff members must be trained to utilize the system’s functionality.

However, with long-term company commitment, preventive maintenance will provide significant savings and can have a very positive effect on the whole plant.

This project is focused on predictive maintenance and this concept will be developed further in the report.

3. Boiler feed pumps

This section is focused on boiler feed pumps – their description and design. Boiler feed pumps are extremely important in plant operation. They control how much water and in which manner water is fed into the boiler.

3.1 Description

The purpose of boiler feed pumps is to boost the feed water pressure sufficiently to overcome boiler pressure to allow entry into the boiler and provide sufficient throughput to satisfy the operational requirements of the boiler. The normal procedure is to use pumps, which increase the feed water pressure in stages.

3.2 Stages of boiler feed pumps

According to Woodward *et al*. (1991) there are three stages of boiler feed pumps:

1) Feed Suction Pumps (FSP) – These are at the first stage of pressure raising and boost the feed water pressure to provide the suction for the start and standby feed pumps and main boiler feed pump.

2) Start and Standby Boiler Feed Pumps (SSBFP) – two 50% capacity motor driven Start and Standby Boiler Feed Pumps are installed per unit for boiler start up, shut down and as back up for the main boiler feed pump.

3) Main Boiler Feed Pump (MBFP). One steam turbine driven pump is designed to provide 100% boiler feed water requirements for the unit.

Output from the boiler feed pumps is fed through the HP heaters via the feed regulating valves.

3.3 Construction

3.3.1 Classic boiler feed pump

Figure 3.1 shows a typical example of an early main boiler feed pump. As it can be seen, it contains six stages on a flexible shaft. This design has a few disadvantages – sudden loss of water may result in pump damage. According to Woodward *et al*. (1991) in case of such a failure, the bolted casing will require a long outage to remove the fault within the internal part of the pump.

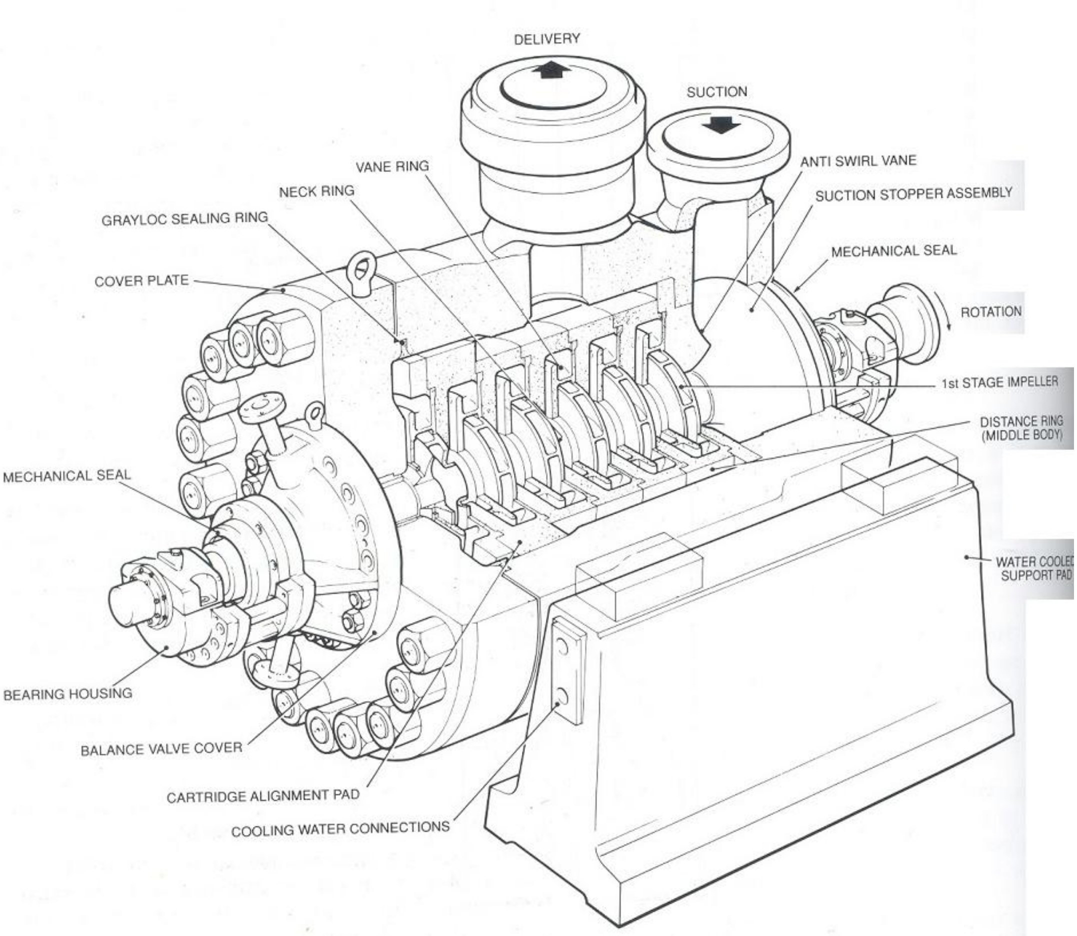


Figure 3.1 Old boiler feed pump arrangement from Woodward et al. (1991)

3.3.2 Advanced boiler feed pump arrangement

Therefore, the industry needed to come up with a new design focusing on minimum downtime and rather long design life. The result was a unit able to deal with extreme situations such as dry-running or any types of shocks without any problems or damage to internal components. Additionally, this new design offered a construction allowing quick replacement of essential components and shorter outages.

According to Woodward *et al*. (1991) in terms of construction, the casing is a forged steel barrel with welded branches (suction and discharge). In order to prevent erosion, stainless steel cladding can be found on most surfaces. Pump supports are arranged in a way which should allow for thermal expansion and give positive restraint to movement upwards and downwards. Sliding keys are responsible for thermal allowances and proper alignment with the drive. A typical example of an advanced boiler feed pump can be seen on figure 3.2.

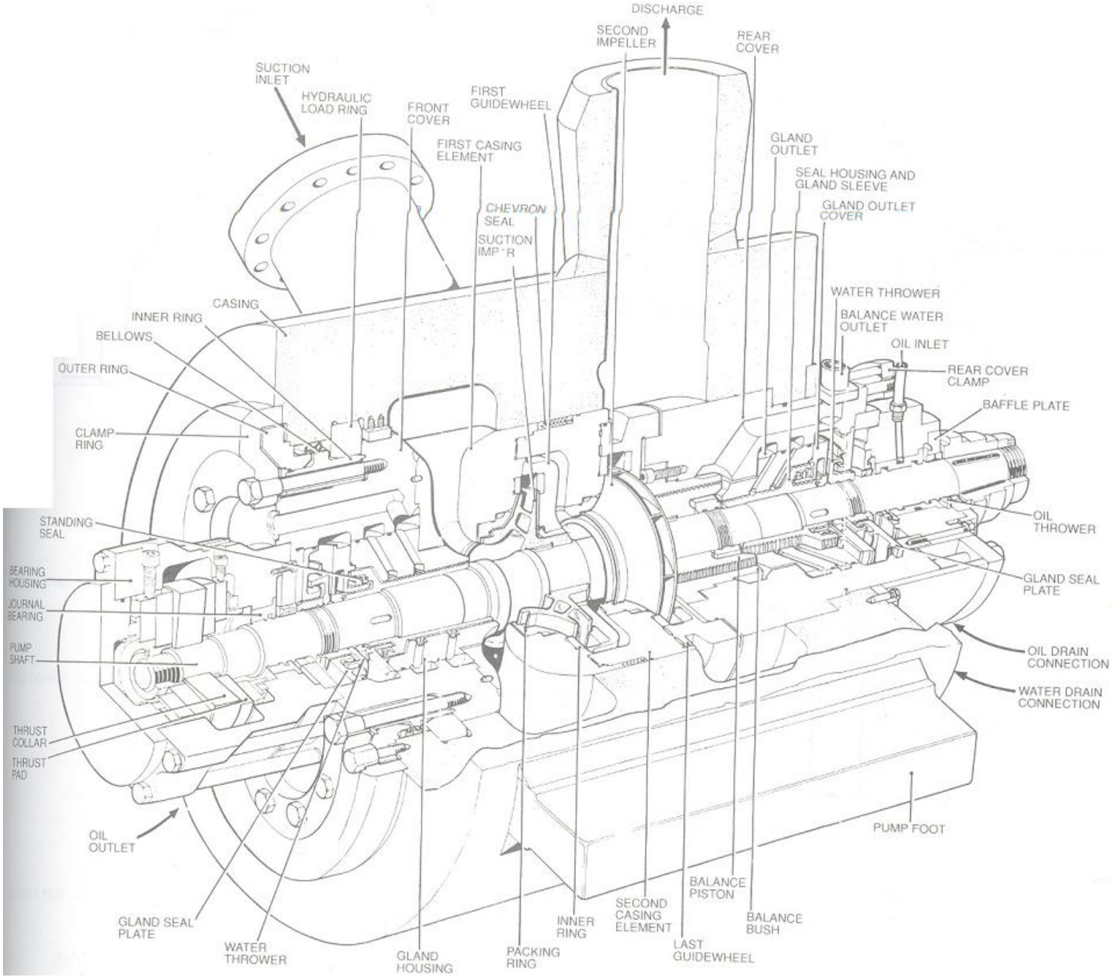


Figure 3.2 Advanced boiler feed pump arrangement from Woodward et al. (1991)

Pump bearings (both thrust and journal) are in housings connected to the casing or the cartridge. According to Woodward *et al.* (1991) such a solution causes the pump to be more resistant to loads exerted by pipework and, additionally, helps to minimise problems with shaft misalignment.

The pump interior assembly, called the cartridge, together with the integrated bearings is build in such a way that it can be easily and safely transported as a single unit for simplified maintenance. The whole cartridge is assembled by its manufacturer, so no additional serious adjustments in order to provide concentricity are usually necessary. As described by Woodward *et al.* (1991) only minor setting up is required after the cartridge is installed inside the pump in order to obtain the expected coupling alignment.

Upon installation of the cartridge into the pump barrel, both suction and discharge ends are moved apart and placed in the respective ends of the casing. Only three joints are required during the installation process – the suction ring and main housing ones. Even though there are some variations of the pump design, the cartridge replacement should be compliant with this process in order to ensure minimum downtime and cost. In order to secure quick replacement process of the cartridge, the bolted cover from previous model has been removed and replaced by a new version with a self-sealing joint system, which dramatically reduces maintenance time. Since this is a rather sophisticated system, in order to minimise replacement time, the operator has to use special equipment specifically designed for this purpose.

High quality stainless steel is a material used for the impellers. The castings require detailed radiographic and dimensional inspection in order to deliver expected performance. According to Woodward *et al.* (1991) the design life for the impellers is around 45,000 hours operating at almost 100% of the design flow.

The arrangement of diffusers can be either axial or radial, because both setups proved to be successful with this boiler feed pump design. In earlier design there was a problem with high pressure pulses between the impeller and diffuser blades and it was a significant safety risk. Advanced boiler feed pumps solve this problem by adding axial diffusers. On the other hand, radial diffusers tend to improve flow stability and the maximum efficiency is slightly higher than in case of axial arrangement. However, this can be neglected due to losses from gland leakage.

In case of some power stations, where electricity generation is stopped during main boiler feed pump outage, forced cooling of boiler feeds pumps has been introduced.

This feature reduces outage time and allows for quick return to service and full capacity – it is a significant cost reducing factor.

3.4 Design and arrangement

In order to achieve a target of creating a pump able to perform dry-running without any problems, according to Woodward *et al.* (1991) the pump has to survive the following conditions without any damage:

- Sudden reduction of suction pressure, while the pump continues to maintain expected head. Additionally, after the pressure is restored, the pump should be able to smoothly accept the new state without a need of manual intervention.
- In case of a dramatic incident causing a total loss of water – this is an extreme situation, but the pump should survive it without damage and be shut down safely until the external failure is taken care of.

The design described above led to a creation of an advanced class pump, which has the following common elements:

- Reduced number of stages and drastically stiffened shaft provide high rotor rigidity and decrease shaft deflection
- Clearances inside the pump are larger in order to provide dry-running capability
- The bolted cover used in older models is replaced by a more advanced self-sealing system operating in high pressure
- A balance drum is installed to reduce the axial hydraulic thrust
- Advanced class pumps are designed to minimize downtime and permit quick replacement of internal pump elements with new parts due to so called cartridge design

The stiffened shaft used in this design and mentioned earlier allows to operate at relatively high rotational speed ranges (over 6 000 RPM) to obtain expected head per unit stage. Same as in the case of older pump arrangements, the design capacity margin is increased in order to counter internal wear in between outages.

According to Woodward *et al.* (1991) there are several design and arrangement issues, which need to be taken into consideration in order to understand the impact of boiler feed pump on plant availability.

According to Woodward *et al.* (1991) the main elements which constrain the optimum arrangement are:

- The economic study, which must take into account following factors:
    - Capital cost
    - Running cost
    - Repair and maintenance costs
    - Impact of loss of availability
- A study to make sure that a problem with one pump element does not dramatically affect the start-up of the main plant or total resultant capacity.
- Making sure that the plant can operate at an acceptable level when a large load is rejected by the generator unit. This means that pump drives must be designed to work in this specified condition.
- The pump system must be able to provide an acceptable pressure margin in case of reducing turbine load and consequent pressure decay.
- It is required that at least two pumps are able to perform plant start-up. In case of turbine drive, there has to be an additional steam supply – for example from an auxiliary boiler.
- In case of multiple pumps operating in parallel, both sets are expected to take over other pump's duties in an emergency situation.

- Since boiler feed pump is a crucial power plant element, it has to be capable of delivering sufficient capacity in extreme situations. Therefore, it is rather common to design boiler feed pumps with a capacity margin in order to handle larger than expected flow. Additionally, there is also a generated head design margin, which takes into account internal pump wear between maintenance actions.

3.5 New developments

One of the most important features of a boiler feed pump is its availability, because power station output depends on its downtime. Therefore, a lot of research has been done lately in this field. Additionally, according to Woodward *et al.* (1991) since boiler feed pump is an expensive plant item, there is a significant research interest in lowering the capital investment and maintenance (including part replacement) cost.

In order to achieve high availability, there was a focus on delivering robust pump design and providing easy replacement parts. This has been partially achieved with advanced class boiler feed pumps. The cost of added redundant capacity has actually be outweigh by benefits in case of failure. Normally, such a failure would result in lost generating power and it can be avoided by added redundant capacity.

4. Condition monitoring examples

This section is focused on examples of condition monitoring techniques used in diagnosing expensive rotating machinery, especially turbines. The availability and reliability of those important plant items has crucial impact on power generating plants and their downtime. Failure of a gas or steam turbine can be very expensive not only in terms of lost profit due to downtime when waiting for replacement, but also additional damage to structures and equipment caused by that failure. Therefore, companies are willing to invest significant sums of money in research devoted to condition monitoring methods, especially for most important and expensive plant equipment. According to Todd *et al.* (2007) early detection and classification of faults can reduce catastrophic failures to minimum.

4.1 Expert systems

Expert systems are one option to implement condition monitoring of turbines. They have been designed to help specialists by using rules created from knowledge gathered from various experts on plant equipment. According to Todd *et al.* (2007) the main problem with this approach is that the process of knowledge extraction is expensive and very time-consuming, because the number of true experts in very specialized fields is very limited and it requires a significant amount of resources to correctly transfer their knowledge.

Expert systems have over a twenty year history with turbine condition monitoring systems. They were designed to examine data from sensors in order to establish the condition of plant item. One of the systems described by Gonzalez *et al.* (1986) analyzed sensor data and basing on corresponding fault probabilities established a possibility for each failure. This system required an engineer to review the results and confirm or alter them.

Another example of expert system described in by Gemmell (1995) has a turbine model implemented and utilizes monitoring data from every element. The problem with knowledge based systems is with events which happen rarely and there is not much expert information on them. However, in case of few catastrophic failure tests, the described systems actually managed to perform well.

The most popular and successful expert system used for turbine condition monitoring is called TIGER described by Milne *et al.* (1997). It actually is build from a combination of two models – rule-based and model-based. The model is designed to compare the rules derived from knowledge elicitation to the current condition using sensorial data. The rules are created based on physical modelling of machine behaviour and expert knowledge coming from experience. TIGER system was very popular within the industry and has successfully worked with many installed turbines.

According to Todd *et al.* (2007) the main disadvantage of expert or rule-based systems is that they are not capable of learning from new situations and actually helping to expand knowledge beyond current expert experience. Additionally, the design process of those systems is very time and resource consuming and requires access to knowledge of specialist engineers.

4.2 Machine learning

Machine learning is a very wide area of knowledge and it is not possible to cover it all in this report. However, it is necessary to mention a few examples of application of machine learning in condition monitoring.

Machine learning is described by Simon (1983) as “any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population”. Therefore, machine learning systems are able to extract new knowledge from gathered data in order to perform their assigned tasks better and improve performance.

4.2.1 Artificial neural networks

An example of machine learning method used in condition monitoring is application of artificial neural networks (ANN). According to Todd *et al.* (2007) they are similar to human nervous system because the system is created from a group of connected neurons with assigned knowledge and weights. This solution does not need expert knowledge in order to create rules – they are derived from available training data. ANN approach saves time by creating very complicated relationships using minimum expert knowledge and requiring little training time. However, in case of incorrect result, due to the fact that ANN works as a black box, it is very often not possible to understand the nature of a misdiagnosis. This major disadvantage immediate disqualified ANN as a viable condition monitoring tool for many industries.

In case of expert systems, the problem with no explanation does not exist, because the rules are derived in a logical way from expert knowledge and are expected to be transparent. It is important in case of erroneous diagnosis causing a catastrophic failure, when consequently an investigation is undertaken. Clarity and good understanding of derived rules are crucial in finding and correcting problems with the implemented condition monitoring system. With unclear rules it is almost impossible and may have a negative effect on plant operator confidence. The knowledge that the system was designed using knowledge from experts, who devoted their careers to become specialists in relevant fields, can have a positive effect on the working environment.

Expert systems, even with all their problems, appear to be very popular within the industry and it is expected that this trend will continue, because intuitively it is a natural evolution of expert knowledge. However, new methods are being developed which use the experience of expert based systems, but only as a training set for machine learning condition monitoring system. Those artificial intelligence methods are called symbolic machine learning according to Luger *et al.* (1998), which can be further divided into analytical and inductive techniques. The former is focused on extracting information from knowledge and creating rules based on an expert system training set. Inductive techniques analyze the supplied data and identify underlying patterns using numerous training data ranges.

4.2.2 Evolutionary algorithms

According to Todd *et al.* (2007) another machine learning method or rather set of methods is called evolutionary algorithms. Those algorithms try to find a matching hypothesis from a random selection of various hypotheses. Numerous attributes are attached to each one in order to establish which one should be able to advance to the next stage. Evolutionary algorithms involve many sophisticated transformations – crossovers, mutations and reproductions in order to find the best solution.

This method is applied to unprocessed condition monitoring data and it had some success within the industry. The main specific application of evolutionary algorithms is finding useful features from analysis of raw data. An evolutionary algorithm school called genetic programming can be used according to Guo *et al.* (2005) to establish a set of main features from raw data and then use it as inputs for neural network model in order to classify faults. This information can be obtained applying statistics and genetic programming operators. As described by Chen *et al.* (2001), genetic algorithms can again extract features from raw and transformed data and use the results as input data to a classifiers responsible for fault diagnosis of bearings.

4.2.3 Bayesian learning

According to Todd *et al.* (2007) Bayesian learning is another example of machine learning algorithm. It is a probabilistic technique which estimates the results knowing the probabilistic distribution of important features. Optimal solutions are selected by analyzing those probability values attached to gathered data. For example, Bayesian learning has been used in induction motor condition monitoring system – described by Haji *et al.* (2001). Additionally, a diagnostics system using Bayesian network has been mentioned by Chien *et al.* (2002). It was used on a network distribution feeder. The dependencies within the network were established using expert knowledge.

4.2.4 C4.5 algorithm

Next algorithm described in this section is an inductive machine learning method called C4.5. According to Todd *et al.* (2007) this technique is able to create decision trees basing on supplied training data. The training data is tested at every non-terminal node of the tree and each consequent branch is a probable value of feature. In case of all training data belonging to the same classification, a terminal node is created. An example of application of C4.5 algorithm is described by Mejia-Lavelle *et al.* (1998), where it's used to determine the power factor interval class. C5 algorithm is an upgraded version of C4.5 algorithm and has been used to extract rules diagnosis faults in using transformer data according to McArthur *et al.* (2004).

4.2.5 Case-based reasoning

Case-based reasoning is a very intuitive technique, sometimes referred to as lazy machine learning method. According to Watson (1997), it has been successful in many fields not at all related to engineering, especially in technical support. It basically stores all the initial cases based on training data and tries to match new data to existing cases. If the result is close enough, then the new instance is added to an existing case. Otherwise, after being accepted by the system operator, a new case is created and then the whole process is repeated for the next instance. Case-based reasoning has been successfully used in condition monitoring as well – an example is described by Stanek *et al.* (2001), where CBR methods were used to find circuit breaker faults on a electricity distribution system. Another engineering example of CBR application is described by Devaney *et al.* (2005), where it has been successfully used to monitor the condition of General Electric gas turbines.

4.3 Condition monitoring methods summary

According to Todd *et al.* (2007) neural networks (ANNs) have been used successfully in some condition monitoring applications. However, this method is very data intensive and requires substantial amount of training data supplied. Unfortunately, in case of some plant equipment (especially expensive rotational plant items), there is not much data available, because failure events have luckily been very rare and extremely little data is available. Another problem with ANNs is that due to their construction, the rules created by the system are impossible to trace back and verify by a specialist engineer.

According to Todd *et al.* (2007) Bayesian learning and its variations has a great potential of success, but there might be difficulties in obtaining the required probability data, because sometimes historical data is practically unavailable or there is too little of it. It is of course possible to roughly estimate those probabilities, but there is a significant chance that it could be erroneous or not accurate enough.

In case of instance-based learning methods according to Todd *et al.* (2007), lack of training data sets can be a major obstacle, because it might be not possible to setup an initial system capable of dealing with at least most common faults. However, those algorithms are capable of expanding and learning about new cases or instances just by being applied online.

5. Data analysis techniques

5.1 Knowledge discovery from databases

In the modern world different types of data are collected all the time in many various fields. Due to increases storage and processing capabilities of computers, the amount of data collected increases very quickly. According to Fayyad *et al.* (1996) it is not possible for human being or even less complicated algorithms to analyze the data and extract useful information. Therefore, is it essential that sophisticated new algorithms are created in order to deal with dramatically increasing volumes of electronic data. Those new methods and tools belong to a growing field of knowledge discovery in databases (KDD).

In practice, KDD is responsible for techniques for understanding and retrieving information from data. According to Fayyad *et al.* (1996) the main problem KDD has to deal with is processing low-level voluminous data into something more useful and easier to understand, for example a model, short report or process description. Basically, KDD is expected to use data mining methods in order to discover underlying patterns in data.

5.1.1 KDD and traditional methods

According to Fayyad *et al.* (1996), traditionally, in order to extract useful information from data, manual analysis had to be performed. Data specialists tend to periodically process the data manually in order to produce reports about, for example trends or changes in trends. The result of this analysis has significant influence on future decisions and plans. Data analysis has been used in many sectors – from retail and marketing to science and finance. The main disadvantage of this system is that it relies on a specialist spending a significant amount of time on analyzing the data and being closely familiar with it. Additionally, manual data analysis is a very slow process and the result can very often depend on analyst’s biased personal opinion.

As the amount of data gathered and required to be analyzed increases, it becomes virtually impossible to manually process all the information and create reports in acceptable timeframe. The magnitude and complexity of data suggest that the analysis process should be somehow performed by computers, because it is not feasible anymore to do it manually anymore.

According to Fayyad *et al.* (1996), many business or engineering decisions rely on large amounts of data being analyzed very quickly and accurately and it is rapidly becoming impossible to do it without any special data mining methods. In principle knowledge discovery from data tries to solve the problem of having too much data to analyze manually.

5.1.2 Data mining and KDD

The interest in research on knowledge discovery from data results in reports about successful data mining application in various fields. According to Fayyad *et al.* (1996), in scientific applications, astronomy is probably the leading area in terms of data overload and KDD has been used to data mine image data collected from various places around the world. Other applications include, but are not limited to:

- Marketing
- Investment
- Fraud detection
- Manufacturing
- Telecommunications
- Data cleaning

The popularity of Internet, which definitely is an information-rich environment, the importance of intelligent data analysis applications is growing very rapidly. Especially various search engines, which provide results based on user interests or specification, for example in the areas of music preferences or searching for information in newspaper archives. The number of successful KDD systems is increasing rapidly – more details can be found in article by Fayyad *et al.* (1996).

The notion of data mining is very often used to describe the process of finding useful information and underlying pattern in data sets. Later on, according to Fayyad *et al.* (1996), a term knowledge discovery in databases (KDD) was introduced. Theoretically speaking, data mining itself is just one (very important) step in KDD process. Other steps of KDD will be described later in this chapter. Using data mining methods without any knowledge and understanding of the whole process can provide meaningless or invalid results.

KDD is a process which involves different disciplines in order to be successful. In order to extract useful information from data, the following research areas are incorporated (according to):

- machine learning
- pattern recognition
- databases
- artificial intelligence
- knowledge acquisition for expert systems
- data visualization
- high-performance computing

Researchers working with KDD need to have significant knowledge at least in some of those areas and have a target to achieve useful knowledge from voluminous data. One may ask a question what the actual difference between data mining and KDD is and why the former is only a single step in the KDD process. Well, according to Fayyad *et al.* (1996) data mining focuses specifically on methods to extract knowledge from data – it is just a blind tool. KDD, on the other hand, investigates problems like data preparation, storage, access and focuses on the actual objective and selection which machine learning techniques will be most appropriate in this case. The KDD process is much more sophisticated and interdisciplinary than data mining itself. Additionally, it is responsible for optimization of algorithms used to make sure they are suitable for size and type of data analyzed.

Very often supplied data is imperfect – there are incorrect or missing values. That is why knowledge discovery from data relies so heavily on statistics, because it can deal with those uncertainties for a selected data range. Data mining is very often looked at with caution and suspicion, because according to Fayyad *et al.* (1996) the early data mining systems after analyzing the data long enough found underlying patterns, which were unfortunately misleading and insignificant. However, over the years the improvements to the data mining methods practically eliminated this problem and nowadays data mining can be safely used as long as it is not just a blind tool applied at random.

According to Fayyad *et al.* (1996), the primary element shaping the KDD process is the database itself. The type and condition of the database itself will have a major impact on what tools will be used to analyze the data. In terms of hardware limitations, it is difficult to manipulate the database when there is not enough computer memory available to process it as a whole. Fortunately, modern computers have enough memory to handle most even very large databases. However, there are special algorithms to make efficient use of memory in case of huge databases or less powerful computers.

There is a whole research field devoted to databases and it is called data warehousing. According to Fayyad *et al.* (1996), its aims are to provide easy (preferably online) access to clean collected data in order to support decision making. Data warehousing helps to set foundations for KDD process via two activities:

- Data cleaning: aim is to store the data in a logical way and prevent from storage of missing or incorrect data.
- Data access: transparent and intuitive methods must be designed in order to provide quick and easy access, also in case of legacy data, which was stored offline

After all the procedures and methods for storing and accessing data are established, the next step is to decide what kind of analysis is required. The main tool used with data warehousing is online analytical processing (OLAP) – according to Fayyad *et al.* (1996). The aim of OLAP is to provide interactive analysis tools for basic analysis and simplification of the data. However, KDD is designed to make most of this process automatic and, in general, is more advanced than OLAP.

According to Fayyad *et al.* (1996), in KDD *data* are a set of cases in the database and *pattern* is a description of a model applicable to the data. Therefore, the aim of KDD is to analyze the data and find a model or a structure which would describe the data. In principle, the aim of a KDD is to provide a high-level description of low-level data. The process involves many stages responsible for preparing the data, finding patterns, checking results against requirements and finally, repeating the process if required. The resultant patterns are supposed have a high degree of certainty that they will be valid with new data. Additionally, the results should be *innovative* (there is no need to learn something already known) and *useful* in further analysis of new data. Finally, the resultant pattern from KDD process should be *understood* and accepted by the analyst to make sure it is not just a computational error.

According to Fayyad *et al.* (1996), it is possible to objectively estimate some of the qualities of the patterns derived using KDD – especially *certainty* (by using the pattern with new data) and *utility* (for example by amount of money saved due to better predictions). However, some qualities are very difficult to quantify, because they depend on the analyst – it is up to him to decide whether he understand the solution and if it is novel enough. According to Fayyad *et al.* (1996), there is another very important quality describing the results – it is called *interestingness* and takes into account multiple factors (some of them mentioned earlier) in order to derive a single figure describing the result of KDD process. After explaining briefly the basic notions in KDD, it can be seen that pattern actually corresponds to knowledge, if the result receives higher interestingness than the minimum required by the analyst, because something new can be learned from it.

The difference between data mining and KDD is that the former is just one step and the latter describes the whole process including data preparation and taking into account additional constraints – for example computational, which can be later on used to adjust the data mining method used. The whole KDD process applies many steps to the selected database in order to extract knowledge from it. According to Fayyad *et al.* (1996), data has to be pre-processed and transformed before data mining techniques can be applied to it. Afterwards, the derived patterns have to be evaluated in order to select the most appropriate solution for this case. As it can be seen, data mining is just a tool without any intelligence attached – it will perform the job being assigned, but won't evaluate its usefulness.

5.1.3 The steps of KDD

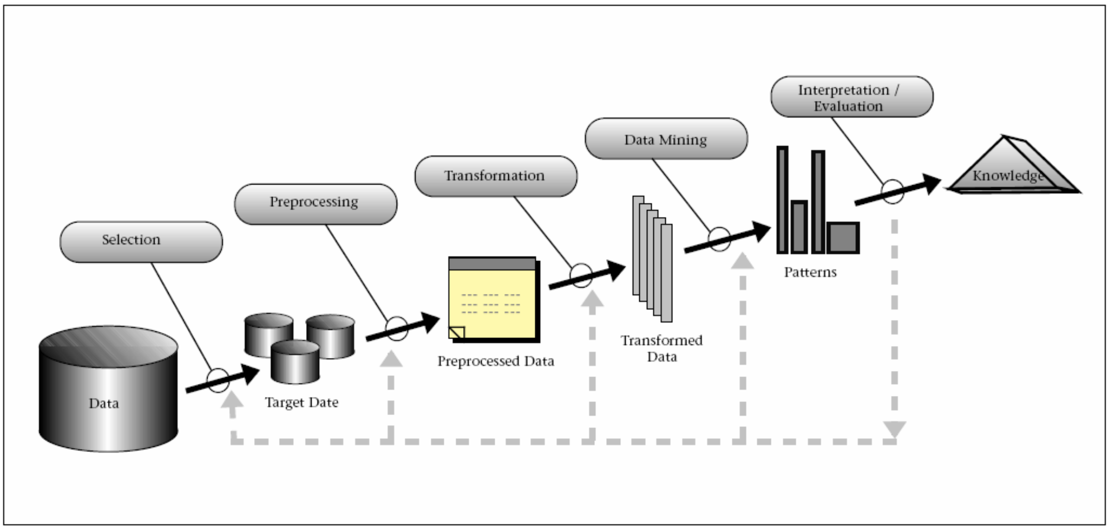


Figure 5.1 Steps of KDD from Fayyad et al. (1996)

The whole process supports user interaction and iteration in order to find the most suitable solution. According to Fayyad *et al.* (1996), the core steps are shown on Figure 5.1 and include:

1) The analyst has to understand the scope, gather all prior knowledge and identify the aims of the KDD process
2) Second step is to select whether the whole dataset will be analyzed or just a part of it. In the latter case, a new target dataset with chosen data should be created.
3) Next step is to apply data pre-processing and cleaning – it involves tasks like:
    - removing noise (or at least reduction) if applicable
    - collecting of information about noise (in order to offset it)
    - deciding how to handle missing data
    - offseting known changes to the data

4) Reduce data dimensions (if possible) by finding additional information about it. Try to find a better way to represent the data, attempt to reduce number of variables.
5) Review data mining methods and select one matching goals and expectations established in step one.
6) This step is based on exploration. There should be at least a rough idea which data mining methods are best for established goals and most suitable for available data. Now there is a need to select a specific algorithm to search data for underlying patterns. Additionally, the analyst has to make sure that the selected model or algorithm is capable of providing required qualities.
7) This step is the actual application of data mining – the analyst is looking for patterns in the selected data range. The aim is to represent those patterns in an understandable way – in a form of rules, structure trees, clusters, regression, etc. In order to fully use the capabilities of data mining techniques, the previous KDD steps have to be thoroughly done.
8) After the data mining process is finished, this step is focusing on understanding and interpretation of resulting patterns. If it is necessary, the analyst should return to any of the previous steps and iterate. Additionally, the resulting models should be visualized in order to improve the interpretation process.
9) The last step in KDD process is a very important one, because now there is an opportunity to apply the resultant model to new data, create a report describing the process and summarizing the results. Additionally, it is an opportunity to check the results, eliminate any errors. If comparison with previously established knowledge or rules shows inconsistencies, then they should be resolved.

The learning from data process can be rather time consuming and may contain multiple iterations between two or more steps. Data mining step seems to be the most important the in the process, but without other steps it would not be as useful and possibly could provide incorrect results.

5.2 Data mining

As described earlier in this section, according to Fayyad *et al.* (1996) data mining is only one step in learning from data process. There are many data mining methods which can used to gain knowledge from data and they all depend on what kind of information is needed and the outcomes of previous steps during learning from data process. From many data mining techniques only a few are actually used in this project and they will be described later in this chapter, but first the next section will focus a bit more on the data mining process as described by Witten *et al.* (2000).

5.2.1 Data mining description

First of all data mining is a practical thing, which focuses on actual tasks, not only theoretical issues. Therefore it can be (and usually is) very important when dealing with real problems – not only engineering ones.

According to Witten *et al.* (2000) data mining focuses on finding and explaining patterns in analyzed data sets – it allows to understand what kind of problem the analyst is dealing with and prepare future predictions of possible behaviours – such information can have a positive impact on whatever task is being focused on and allows long-term planning. For example, maintenance schedule of plant equipment can be based on applying data mining techniques on historical condition monitoring data. This may dramatically reduce costs or downtime. Additionally, equipment life can be extended – this can have a significant impact on smooth plant operation and also safety of working environment.

In data mining process supplied data can have very uneven quality. Usually an analyst should receive a data set with examples of different situations, for example what kind of drugs should be prescribed given a certain set of symptoms. In case of engineering applications and especially condition monitoring problems, an analyst should receive data sets with corresponding conditions – for example types of failures of their severity. Using this provided data and applying data mining techniques, it is possible to create a monitoring system which will analyze the data practically immediately and determine current equipment condition. Using this approach the operator is aware, whether plant equipment is in good or bad condition without the need to shut it down and perform costly analysis.

5.2.2 Clustering

The main data mining tool used in this project is clustering. According to Witten *et al.* (2000) this technique is used when the supplied data is supposed to be divided into natural groups indicated by data distribution. It is expected that those clusters represent a pattern in data and help in understanding the mechanism causing certain instances fall into one cluster rather than the other. It can be shown in multiple ways – graphically or as a mathematical description of all clusters.

There are different ways the groups manifest themselves. Sometimes the identified clusters might be overlapping, so the instance might fall into more than one group. The other extreme is that the clusters are exclusive and an instance can only fall into one group. In a situation, when a probability model is being used, each instance has a certain probability of belonging to a cluster. Another option is creating a hierarchical division. There is a large selection of possibilities how the data can be displayed. It really all depends on what the analyst is dealing with and what is to be achieved.

In this section a few clustering methods will be described. The first one is k-means clustering which forms clusters from numeric data sets creating disjointed clusters. According to Witten *et al.* (2000) it is a popular simple and straightforward method, which has been used for some time now.

Another clustering technique described in this report is an incremental clustering method called Cobweb. It takes into account so called *category utility* which corresponds to the quality factor of a cluster.

The last one is a statistical clustering technique taking into account various probability distributions. This method does not determine to which cluster an instance belongs, but rather what is the probability that an instance belongs to a certain group.

5.2.2.1 Iterative clustering

This method (called *k-means* clustering) is a simple iterative distance-based technique. In order to use it efficiently, one actually has to look at data distribution first and specify the parameter *k*, which is a number of clusters expected the data to be divided into. In the next step, the algorithm chooses *k* random cluster centres (within the data range) and for each instance calculates the distances to those cluster centres using the Euclidean distance formula. Afterwards, each instance is simply assigned to the nearest cluster. When all instances have an assigned cluster, for each group the centroid (mean value) of all instances is calculated, which become the new cluster centres. Next, the whole algorithm is repeated using newly calculated centres. Repetition does not stop until for two consecutive iterations, the same instances are assigned to the same group and, therefore, cluster centres are and will remain unchanged.

According to Witten *et al.* (2000) this technique is rather simple yet effective in most situations. It should be noted that as with all clustering methods, the resultant centroids are only local extremes and may be totally different during another execution of the algorithm if initial random cluster centres are different. It is possible for this method to fail to find suitable centroids, when the choice of initial random cluster centres is unfortunate.

In order to maximize the possibility of finding a global extreme, it is wise to repeat the algorithm quite a few times with many variations of initial centroids and compare the results. Assuming it does not take an unreasonable amount of time to run the algorithm, using this approach the best set of clusters for a given data set can be selected.

There is quite a significant number of variations of the classical k-means method according to Witten *et al.* (2000). Some focus on implementing hierarchical clustering by executing the algorithm with k = 2 for the given data set and then repeating the process within each new group. Some improvements focus on making the algorithm faster, because the basic version can be time consuming for certain data sets due to multiple iterations each involving performing numerous distance calculations. Therefore, there are algorithms using reasonable and simple approximations which speed up the whole process. They, of course, can have decreased accuracy, so the quality of results might decline. It is important that the analyst is capable of deciding whether the benefits outweigh the cost of applying those approximations.

5.2.2.2 Incremental clustering

As mentioned earlier, simple k-means algorithm repeats the process for the whole data set until the resultant clusters become stable. This clustering method works in a different way – it just adds instance by instance incrementally. Imagine a tree with leaves and roots representing instances, an example of incremental clustering is shown on figure 5.2. In the beginning, there is only one root and then more roots and leaves are added one by one. Adding a new instance may be as simple as just adding a new root/leave or as complex as redoing the whole tree or a part of it. The most important issue is deciding how a new instance should affect the existing tree. As described by Witten *et al.* (2000) in order to do this, a new quality called *category utility* is introduced – it measures the quality of a division of a group of instance into smaller clusters – the exact process will be described later.

The first couple of instance added to the new tree structure form new one-instance clusters attached to the tree. Next, each added instance is assessed individually and basing on category utility it is determined whether it is going to be a suitable first instance for a new cluster. In order to prevent the algorithm from being dependant on the order in which instances are introduced to the structure, there is a possibility to rebuild the tree during the process of adding new instances. Using category utility, sometimes adding a new node may result in a merger of existing nodes with the newly added one to form a new substructure. A way of approaching this problem would be investigating all possible pairs of nodes and checking if they qualify for merger. Unfortunately, if this process was to be repeated every time when a new instance is added, it would require a lot of computational power and would not be efficient, because some work would need to be repeated every time a new instance was being considered.

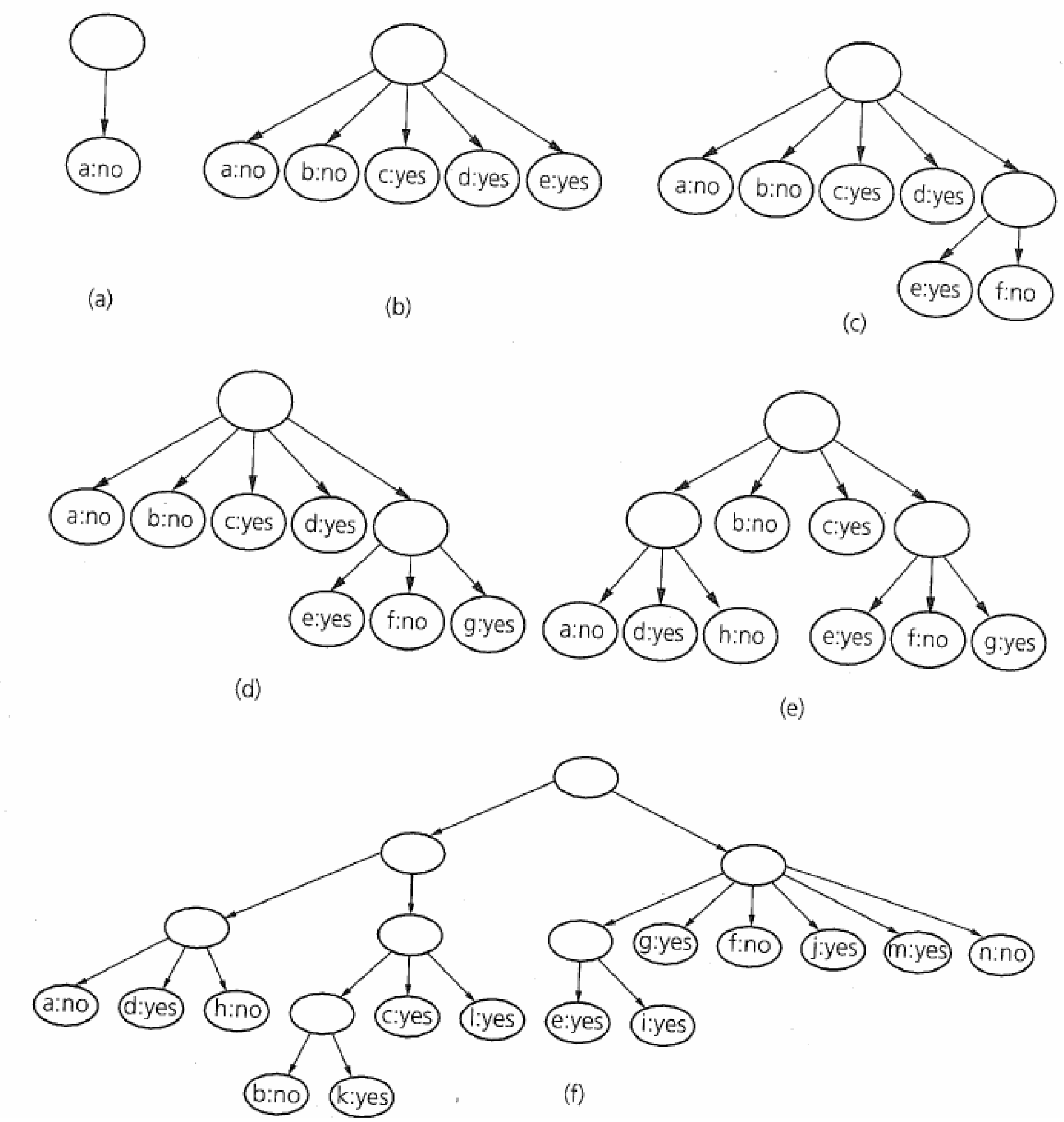


Figure 5.2 Example of incremental clustering from Witten et al. (2000).

However, according to Witten *et al.* (2000) in order to save time, whenever the pairs of nodes are evaluated for merging the instance with an existing one, the two best options for a certain level are memorized. Usually, then the best match will form a new cluster with a new instance unless according to category utility it is better for that new instance to be in a separate cluster. However, before that happens, the two best options mentioned earlier are considered for merger. If it appears to be feasible, this new cluster containing those two best options is created and then is it assessed whether the new instance should be on its own or joined to this newly created subsection.

In order to provide maximum flexibility, a splitting operation can be performed. As in previous case, when the best possible cluster for a new instance is found, it is assessed whether it's worth splitting this cluster. If it is feasible, the cluster is split into two or more parts. Both merging and splitting operations provide a tool to rebuild the tree structure during the algorithm and correct any errors made due to order in which the instances are introduced to the structure.

According to Witten *et al.* (2000) this algorithm works in the same way for both numerical and nominal attributes. The category utility value is based on mean and standard deviation of the value. Unfortunately, with numerical values in case of single-instance clusters the standard deviation is zero and this results in infinite value of category utility. Therefore, a solution has to be implemented in order to force minimum variance in each case. From theoretical point of view it is claimed that this forces minimum value represents a measurement error, which by definition has to be greater than zero.

In some cases clustering produces a tree with an unreasonable amount of clusters and reducing the quality of data mining. In order to prevent the tree from growing to having too many separate instances, a new parameter is introduced – so called cut-off, which is used to reduce overgrowth as mentioned by Witten *et al.* (2000). If clusters are assessed as similar enough, then there is no point to have them as separate groups. Cut-off parameter defines what this similarity threshold is and it is defined using category utility. If creating a new cluster does not increase category utility by a defined value, the new instance is joined to an existing group.

5.2.2.3 Probability-based clustering

The clustering methods described earlier in this chapter have certain disadvantages, for example in case of k-means clustering, the k parameter has to be specified and its incorrect choice may lead to undesirable results. In case of Cobweb technique, the cut-off parameter and problems with setting the minimum value for standard deviation in case single-instance clusters may cause problems and setting those values incorrectly may have negative impact on the analysis results. Additionally, even with merging and splitting functions described earlier, there is some uncertainly whether the order in which instances are introduced to the structure affects the final results and are the taken measures enough to deal with unlucky order of instances. Another disadvantage is that the need to repeat those algorithms in order to choose the best result increases the factor of human error in the formula.

According to Witten *et al.* (2000) some of the disadvantages mentioned above can be overcome using a statistical approach. Since due to the way how clustering works it is not possible to be completely sure about how the data is divided into clusters, even training examples are not supposed to be assigned to only a single cluster. This is where probability-based clustering comes into play – it calculates probability that an instance belongs to a given cluster (it is assessed for each cluster).

The base for this technique is statistical method called *finite mixtures*. It is a group of k distributions for k clusters, which is responsible for attribute values in each cluster. What it means is that a distribution should have a selection of attribute values specific for a cluster; therefore, it is different for every cluster. In practice, an instance can be only associated with a single cluster, but it cannot be confirmed which one. An instance is assigned a set of probabilities of belonging to each cluster.

In the most basic case, there is only one attribute value and a normal distribution for every cluster with different means and standard deviations. The algorithm is responsible for looking at all the instances and calculating mean and variance for each cluster and how the instances are distributed between them. According to Witten *et al.* (2000) the finite mixture model combines together the distributions and creates a probability density function. In summary, when there is a set of instances, the finite mixture model uses them to calculate the mean and variance values for the clusters and probabilities of instance belonging to each cluster. In this case either there is a training model provided or the parameters defining the finite mixture model are known.

5.2.2.4 The EM model

Sometimes there is no training set provided or the mixture model attributes are not available. In this case, according to Witten *et al.* (2000) an analyst has to use an approach similar to the already described k-means method – it contains iteration element. The first step is guessing the parameters of the finite mixture model and basing on this calculating the probabilities for each instance. Next step is recalculating the model parameters using the new probability values and repeating the whole process. Such an approach is called the EM (expectation-maximization) algorithm. Expectation – calculating the probabilities for each instance; maximization – calculating the model parameters. The probabilities basically act like weights according to Witten *et al.* (2000).

As mentioned above, this algorithm uses iteration in a similar way to k-means. In case of the latter one, the iteration stops, when clusters stabilize and centroids are not changing between iterations. In case of EM algorithm it is a bit more complicated unfortunately – the algorithm aims to reach a stability point, but in practice is not able to reach it. In order to control this process, another parameter has to be introduced – the possibility of the instances being part of the dataset using the values of mixture model parameters.

In practice, the formula described by Witten *et al.* (2000) calculates the quality of clustering and it is supposed to be increased with every iteration (for the EM algorithm). Since no normalization procedure is applied, this likelihood is not actually a probability; it can go beyond the zero to one interval. However, greater values are considered as better quality clustering. In real application, the value is usually presented in a form of logarithm – by summing logarithm values of components in order to avoid multiplication, but it is still required to apply iteration in the EM algorithm until the increase in this value is satisfactorily low. Usually, during the first initial iteration, this quality is expected to increase dramatically until a point when it is almost stable.

The EM algorithm is still looking at the local maximum instead of the global one. Therefore, according to Witten *et al.* (2000) in order to choose the best solution, it is a good idea to run the algorithm a few times in order to have a better insight and select the most useful solution. The best solution should have the highest likelihood or logarithm of likelihood value – this should correspond to best quality of clustering from available sets of results. In case of multiple clusters EM algorithm can get very complicated and require significant amount of time and computational power.

5.2.2.5 Bayesian clustering

According to Witten *et al.* (2000) the problem with EM algorithm is that when it is assumed that all attributes are dependent on each other, the solution might be overly conservative and suffer from overfitting. This can be a problem especially when there are many clusters and parameters present. In the worst case, there is only one instance belonging to every cluster – in such a situation data mining does not improve the understanding of the data set at all. It can happen to a finite fixture model when the distributions are too narrow to fit reasonable amount of instances. Therefore, newer algorithms force the clusters to contain at least two instances.

When dealing with a large number of parameters, it is unsure which ones are dependent on each other – it increases computational time and it is more difficult to choose the best solution. Unfortunately, the likelihood of data belonging to the data set increases also when the overfitting occurs; it might compromise selecting the most suitable result.

In order to suppress creation of new parameters and prevent overfitting, a so called Bayesian approach should be implemented. In this implementation, every parameter has a distribution assigned. Next, before a new parameter is introduced it is required to multiply its probability by the overall quality figure. Since the likelihood quality will be multiplied by a value less than one, it will be decreased. This way the algorithm is penalized for increasing the number of parameters in the system. Therefore, new parameter will only be introduced if it will really benefits the solution.

In a same way it is possible to reduce the final number of clusters by penalizing the system by dramatically decreasing solution quality whenever a new cluster is added.

According to Witten *et al.* (2000), Weka software, which will be described later in this section, has a function called AUTOCLASS which is actually a Bayesian clustering scheme. It uses finite mixtures model with prior probabilities on all parameters. In order to estimate the parameters, EM algorithm is used and repeated several times to hopefully obtain a global minimum. Additionally, AUTOCLASS takes into account different numbers of clusters, different covariances and different types of probability distribution. The main drawback of AUTOCLASS is that it is very complicated and requires a significant amount of time and computational power for larger data sets.

5.2.3 Data mining summary

In this chapter, three different clustering methods have been discussed. As you can see, they all have certain advantages and disadvantages and produce different types of results. However, according to Witten *et al.* (2000) they are all capable of taking data as input and performing clustering on a given dataset. It is to be decided by the analyst which technique is the most desirable and which would provide best results. If it is feasible, it might be a good idea to try all of them using, for example, data mining software Weka and compare the results.

Clustering helps to visualize the data and find underlying patterns and rules behind the association with certain cluster. Additionally, if clustering is used on training data, the results can be later used to create classification rules for decision support systems. In case of probabilistic clustering, the rule or decision system has to support using multiple weighs, which represent probabilities of instances being assigned to a certain cluster.

Next important role of clustering is to fill in missing data values, because some attributes might be unavailable or corrupted. Therefore, basing on training data, it is possible to estimate missing values. In case of all methods described in this chapter, there is an underlying assumption that attributes are independent. Some actual implementations of the algorithms allow for tying multiple attributes together if there is a known dependence between them and in this case they are modelled together.

There are some limitations, which will not be described in this report. Further information can be found in a book by Witten *et al.* (2000).

Before applying any clustering method, it is important to prepare the data first according to for example KDD process described earlier in this chapter.

5.3 Weka

Weka is a software package developed by Witten *et al.* (2000). It stands for Waikato Environment for Knowledge Analysis. The programming language chosen to write Weka was Java, because it provides a selection of methods for data processing and evaluation various data mining techniques.

Weka has an implementation of various modern data analysis and modelling algorithm – detailed algorithms have been described by Witten *et al.* (2000). Weka provides a graphical user interface for easy access to its functionality. Additionally, it is a free tool on a GNU General Public License. Input data has to be provided in a single file prepared according to Weka specification.

Weka is a powerful application supporting following data analysis tasks:

- data pre-processing
- clustering
- classification
- regression
- visualization
- feature selection

In this project, Weka was used for some of those tasks – detailed description can be found in the next chapter.

6. Case study

In this chapter a case study is described, where KDD process was used to extract knowledge from boiler feed pump data. In the first part of this chapter, the initial analysis attempt was described, where the data mining methods applied did not derive any useful knowledge from data.

The second part of this chapter is focused on analysis of a second data set supplied by British Energy and using a radically different approach than in the initial analysis. As a result a system checking for alarms has been created in Excel. Provided that this solution is validated and approved by a British Energy engineer, in future this system can be implemented on-line.

According to British Energy the goal of the data analysis was to create a system which would trigger an alarm in case of abnormal state of operation.

6.1 Initial analysis attempt

This section of the report contains the description of application of knowledge discovery from data process to boiler feed pump data supplied by British Energy.

6.1.1 Data description

Data supplied by the British Energy comes in two files representing two different periods:

1) File 1: from 14-06-2007 to 27-11-2007
2) File 2: from 26-11-2007 to 08-04-2008

As it can be seen above, the supplied data spans over several months, therefore it is quite a large data set and may correspond to multiple changing states of the machine. There are about fourteen thousand instances for each quantity in each file.

Files supplied by British Energy include measurement of three different quantities:

1) High pressure (HP) side peak to peak vibrations – measured in micrometers
2) Low pressure (LP) side peak to peak vibrations – measured in micrometers
3) Governor valve position – measured in percents

All the measurements were taken every 10 minutes for each quantity for practically the whole period mentioned above. Due to extremely low sampling frequency it was not possible to apply special techniques designed to analyze vibration data with high sampling frequency – for example FFT.

What makes this dataset so difficult to analyze is the fact that only raw data was received without any expert or any other knowledge added to it. Therefore, it was not known which parts of data correspond to normal or abnormal state of operation. Perhaps for the whole measurement range, the boiler feed pump was in a normal state and any spikes or suspicious measurement results are just regular behaviour or noise.

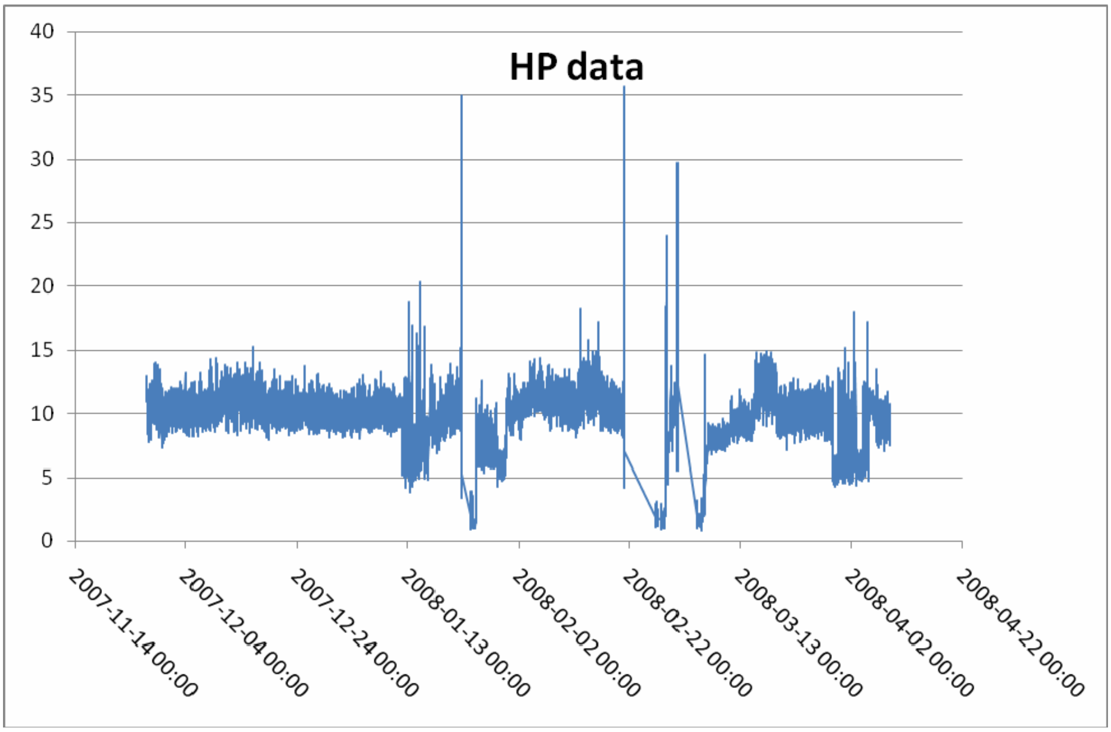


Figure 6.1. HP data from file 1

It seems initially, that the analyst is not in a very advantageous situation here. This might be true, however, lack of any prior knowledge can actually be an advantage. Without any expert knowledge attached to the supplied data, there is no risk of any kind of bias towards known rules and solutions. This way, applying knowledge discovery from data process, maybe it will be possible to discover new rules and underlying patterns in the data, which might improve the fault identification and classification process and help experts with creation of new rules.

Since the goal of this data analysis process is to establish a system able to trigger an alarm in case of abnormal state of operation, the next step in the KDD process is to decide how to divide the data and which parts of it should be used in the analysis process. In this case, it was decided to use the whole data set to check if it can be further divided into smaller parts corresponding to various states.

In order to show the data on graphs Microsoft Excel software was used. On figure 6.1 HP data from file 1 is shown. Figure 6.2 presents LP data from file 1 and on figure 6.3 in appendix 1 governor valve data from file 1 can be seen.

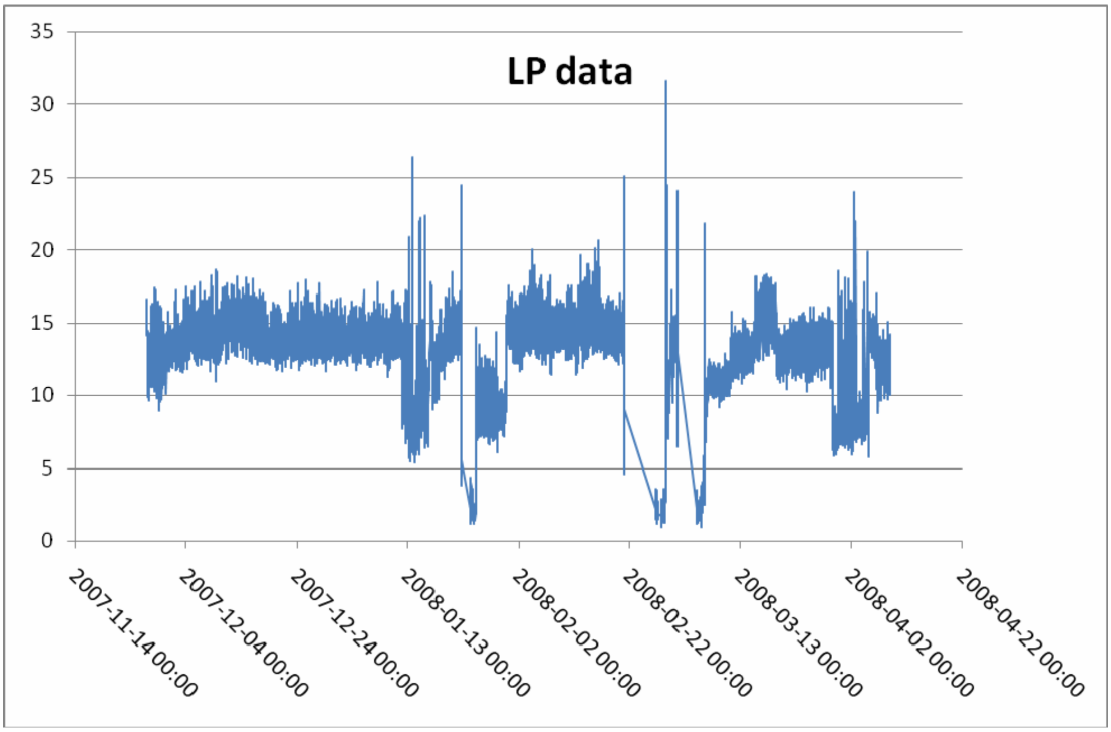


Figure 6.2. LP data from file 1

Figures 6.4 shows HP data from data file 2, on figure 6.5 presents LP data from file 2 and figure 6.6 in appendix 1 presents governor valve data from file 2.

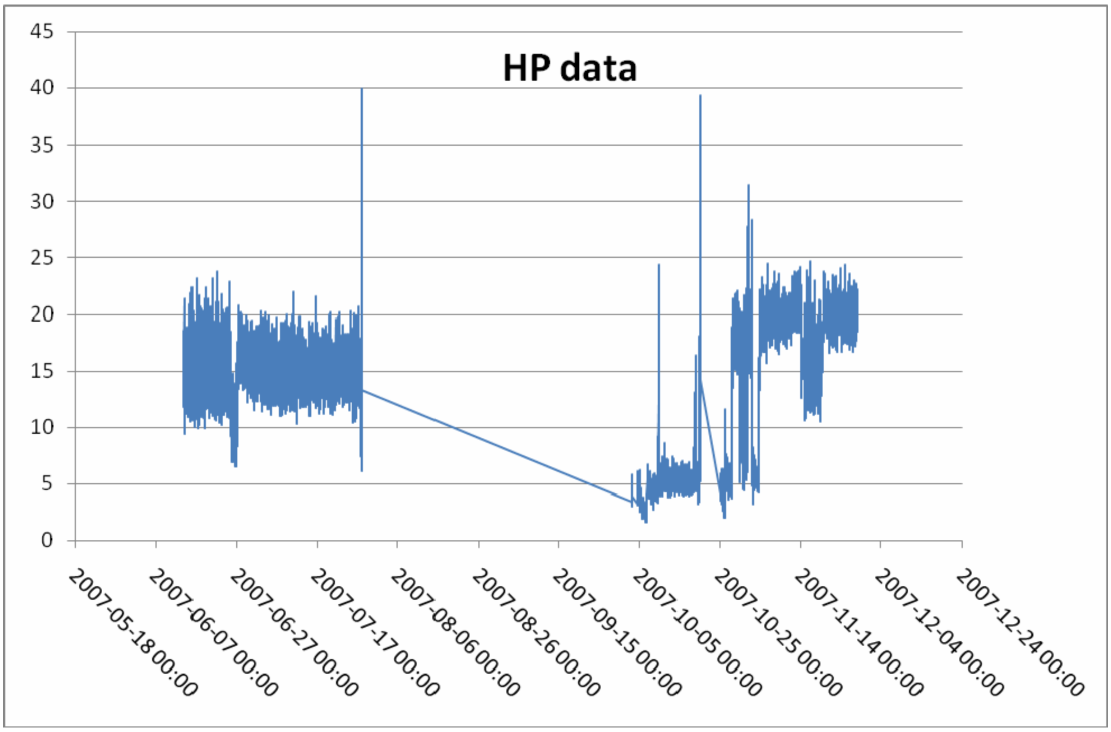


Figure 6.4. HP data from file 2

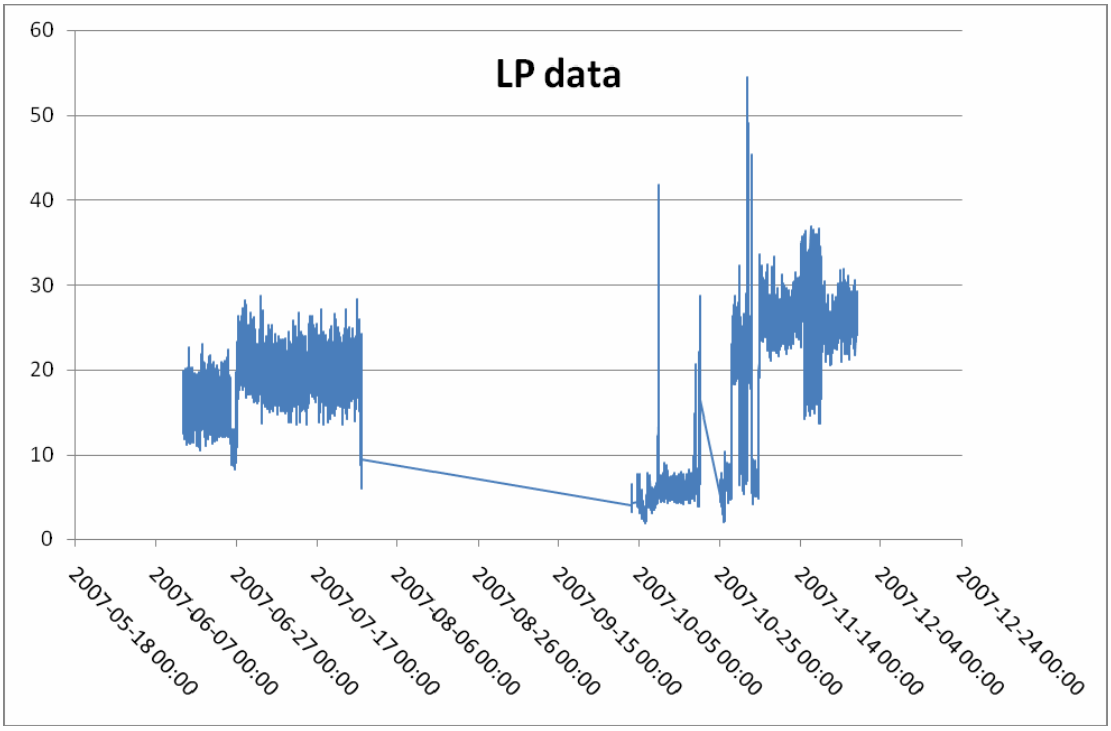


Figure 6.5. LP data from file 2

Since there was no additional information about the data or knowledge which parts could be treated as noise, handling of noise was ignored. In step three of KDD process, the analyst should decide how missing data should be treated. Fortunately, in this case, there is no missing data and this step can be ignored.

6.1.2 Data distribution

In the next step, before analyzing which data mining method could be used to extract knowledge from data, it was attempted to present the data in a different way, so it would be easier to apply data mining to the supplied data set. Using Weka software described in section 5.3, data distribution of each measured quantity was investigated. This was applied to raw data without any averaging or data manipulation.

The aim is not to show exact values and precise distribution, but rather have an appreciation of visual representation of the data in order to select the most appropriate data mining method a few steps later. It can be seen on figure 6.7 the data distribution for HP side vibration taken from file 1. On the graph one can immediately notice two or even maybe tree distinctive areas, around which the data is focused. Looking at figure 6.8 showing distribution of LP side vibration data, again two or possibly three areas, where data is concentrated, can be noticed.

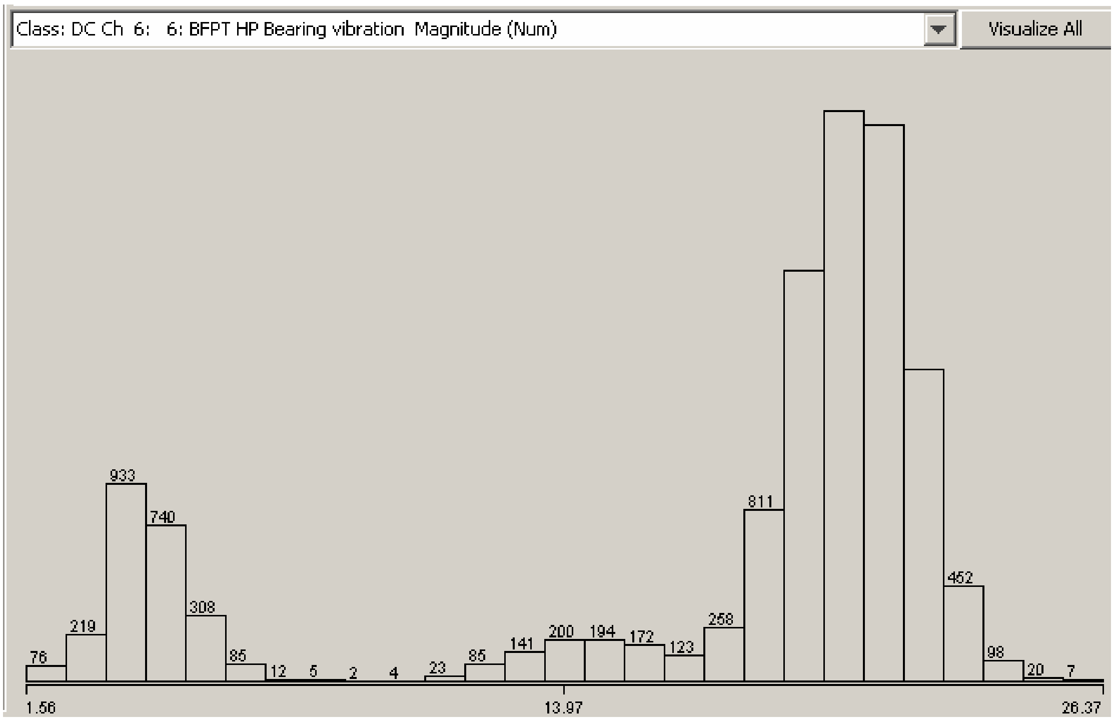

Class: DC Ch 6: 6: BFPT HP Bearing vibration Magnitude (Num)

Figure 6.7. HP data distribution – from file 1

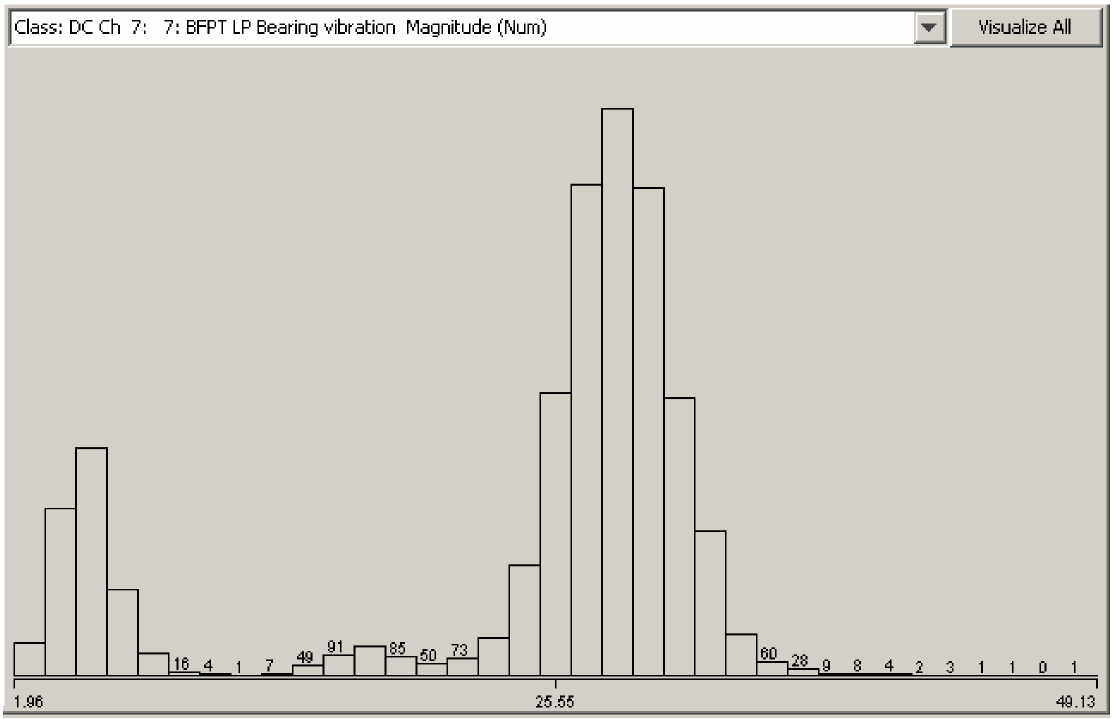


Figure 6.8. LP data distribution – from file 1

In case of figure 6.9 in appendix 1, which presents governor valve distribution data possibly three or four areas of data concentration can be seen, which could be investigated. Before applying data mining techniques, there is a need to look at data distribution for the second file first.

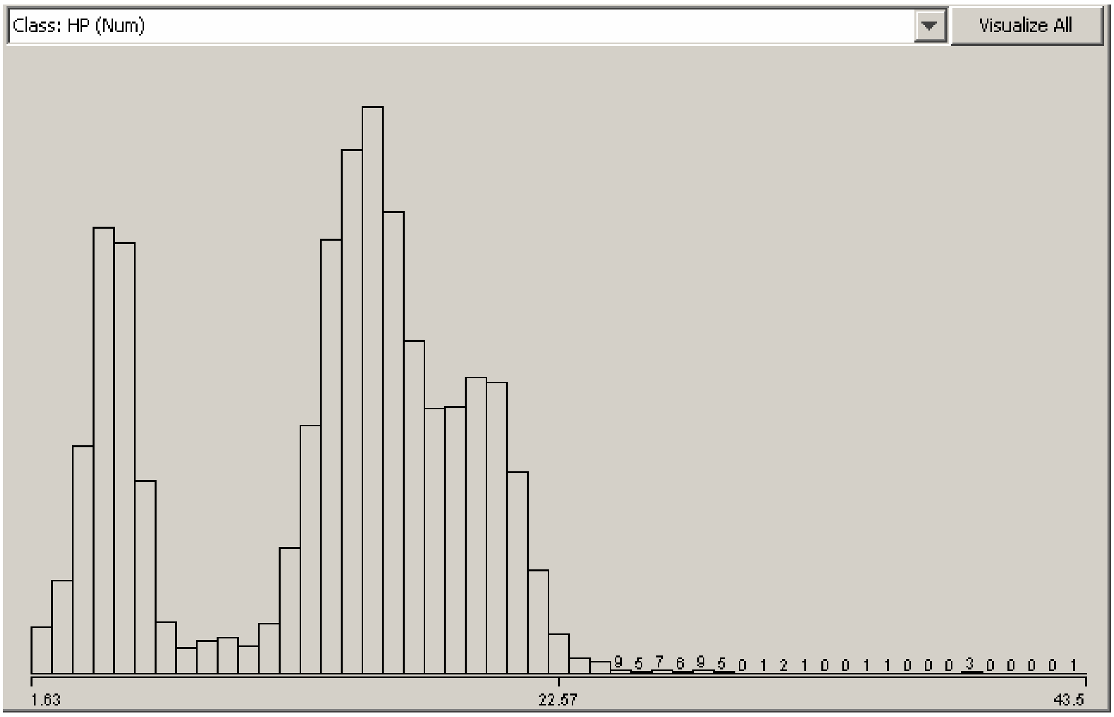


Figure 6.10. HP data distribution – from file 2

In case of second file, data distribution was also prepared in Weka software. The results shown on figure 6.10 for HP side data show three areas of data concentration. Similar results can be observed on figure 6.11, where LP side data are presented. It can be noticed that there is a certain similarity between HP and LP data distribution.

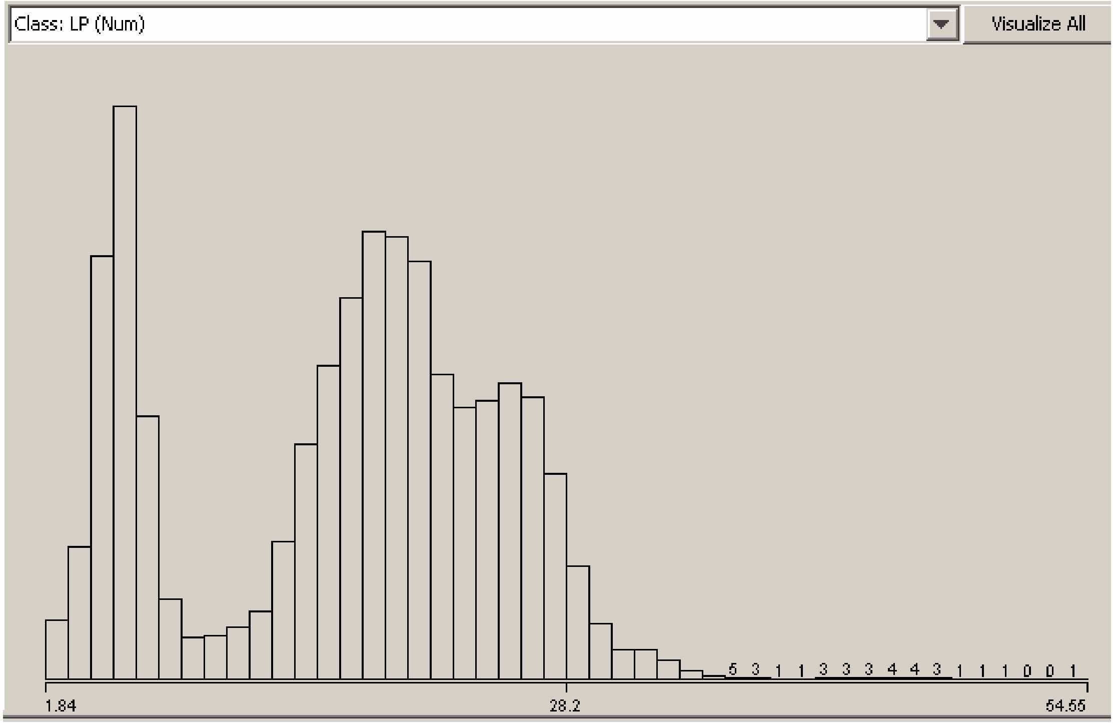


Figure 6.11. LP data distribution – from file 2

Figure 6.12 in appendix 1 presents governor valve position data for the second file. Similarly as in the first file, possibly four areas of data concentration can be identified on the graph.

Next step in KDD process is to look at possible data mining techniques and assess which one will be most suitable in the analyzed case. After looking at data distribution diagrams for both files, it can be noticed that each quantity has two, three or even four areas, around which data instances are concentrated.

Due to the fact that according to data distribution diagrams, data instances are gathered into natural groups, it was decided that clustering would be the most suitable data mining method in this case. Data instances form natural clusters, therefore it might be a good idea to investigate this approach.

6.1.3 Clustering

After choosing a direction in terms of data mining methods, according to KDD the exact algorithm which should be used with the data should be specified. Since from data distribution plots it is known roughly how many clusters can be expected, it was decided to apply the fastest and simplest clustering method – k-means clustering, which was described in chapter 5 of this report. If a simple method is capable of performing the analysis, there is no point in trying to overcomplicate data mining.

6.1.3.1 First file

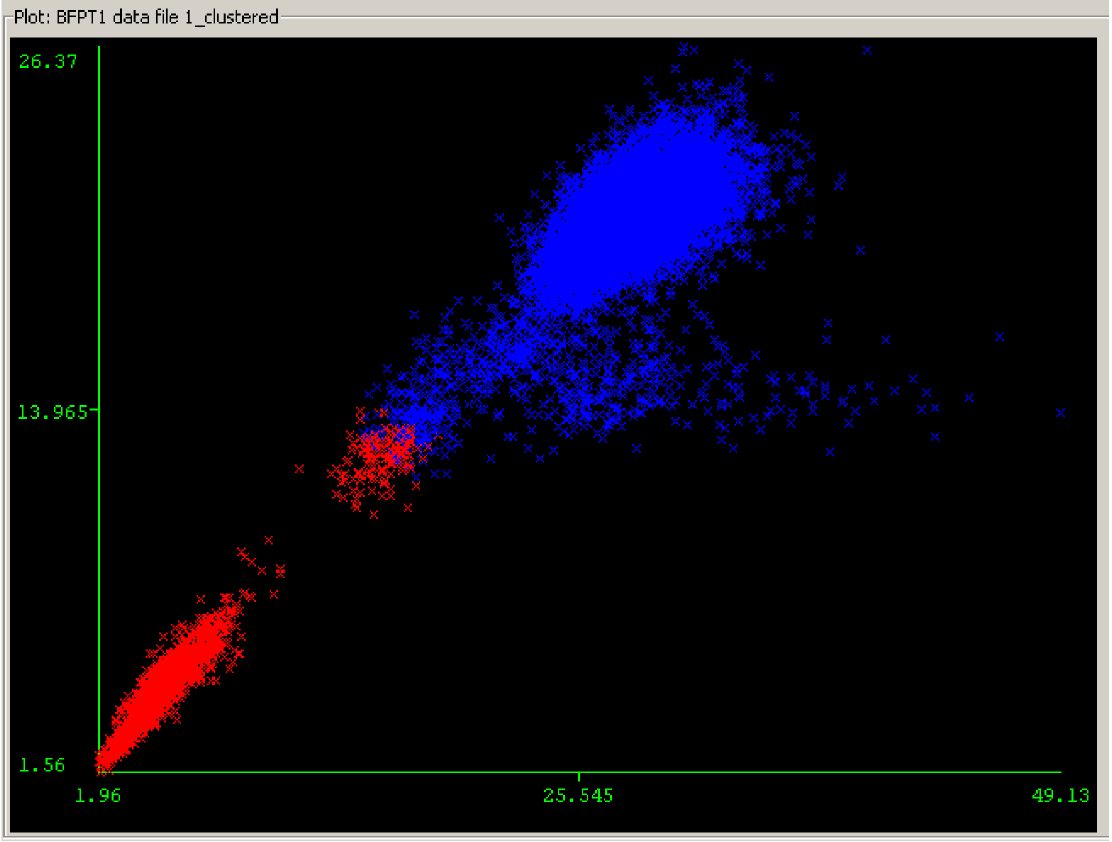


Figure 6.13. LP vs. HP data plot divided into two clusters – from file 1

Step seven in KDD process is applying data mining methods – the results will be described in this section. K-means clustering was performed in Weka software, because it is capable of performing such an analysis and presenting the results in a very convenient way. At this point, it was decided to focus mainly on HP and LP vibration data and ignore governor valve position for the time being.

Clustering was performed for each HP and LP dataset and presented as an LP (on the horizontal axis) versus HP (on the vertical axis) data plot with different colours representing clusters.

First attempt was to try to divide the data into two clusters. Resultant LP versus HP data plot for file 1 can be seen on figure 6.13. It can be noticed that one cluster is formed for lower range of vibrations and a second one for higher vibrations on both sides. This is a case, when data mining extracts knowledge from data, but this knowledge is not very useful.

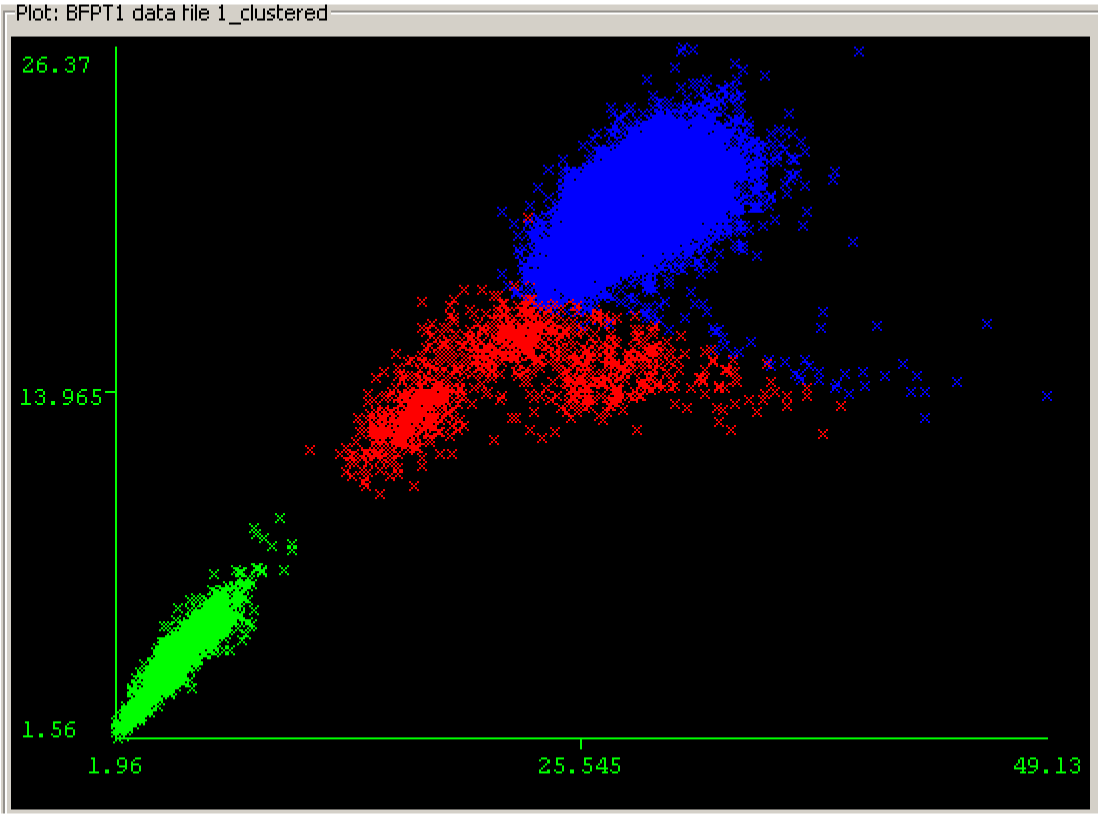


Figure 6.14. LP vs. HP data plot divided into three clusters – from file 1

According to data distribution shown on figures 10 and 11, it could possibly be expected to have three clusters in the data, therefore the next step is to perform k-means clustering with k = 3 and see if any knowledge can be extracted from the results.

Results can be seen on figure 6.14, where LP versus HP data graph is presented. It can be noted that clusters are formed in a similar way as on figure 6.13 in case of only two clusters. Clusters are created in lower, middle and higher range of vibrations on both sides. Again, such a result from clustering does not provide any useful knowledge and cannot be taken any further to create a suitable alarm system.

From shown results a trend can be seen here – adding more clusters will not help with extracting any knowledge from the data. Therefore, in order to confirm this theory it was decided to perform k-means clustering with k = 5. Results (shown on figure 6.15) are pretty much as expected. Clusters correspond to increasing vibrations on both sides and do not provide any additional useful knowledge.

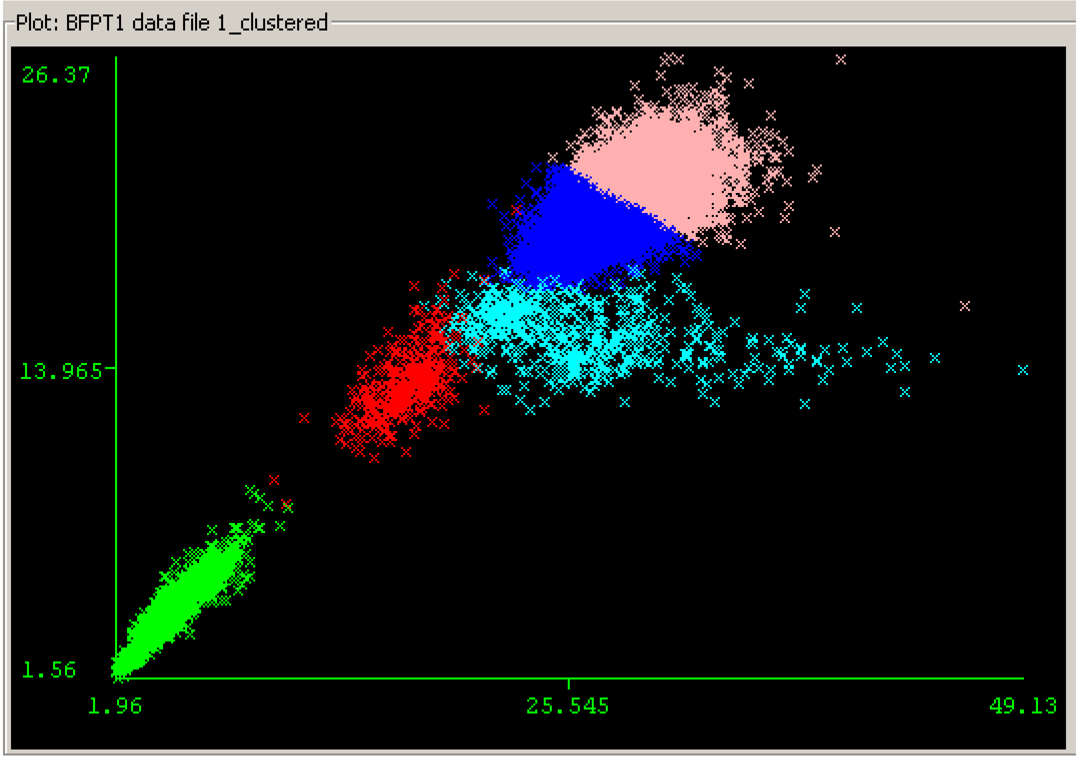


Figure 6.15. LP vs. HP data plot divided into five clusters – from file 1

Analysis performed so far did not provide any useful knowledge and adding further clusters will not improve the results. Therefore, according to KDD process, it is necessary to return to one of the previous steps and iterate. In this case it was decided to use the second supplied data file, which contains similar amount of data instances.

6.1.3.2 Second file

The data mining method used remains the same and result presentation is also the same as in previous case. At the first attempt, k-means with k = 2 clusters is performed. Results can be seen on figure 6.16. Again, the results do not highlight any possible abnormal states, just state the obvious – that with increasing vibration on one side it can be expected for vibrations on other side to increase as well.

Figures 6.10 and 6.11 show that according to HP and LP data distribution it can be expected to possibly have three clusters in data. Consequently, k-means clustering with k = 3 was performed and the results were checked – they are shown on figure 6.17 in appendix 1. It can be seen that the trend is rather similar to the previous data set and perform k-means with k = 5 clusters just to confirm the theory. Results shown on figure 6.18 in appendix 1 prove that useful knowledge cannot be extracted from this data set.

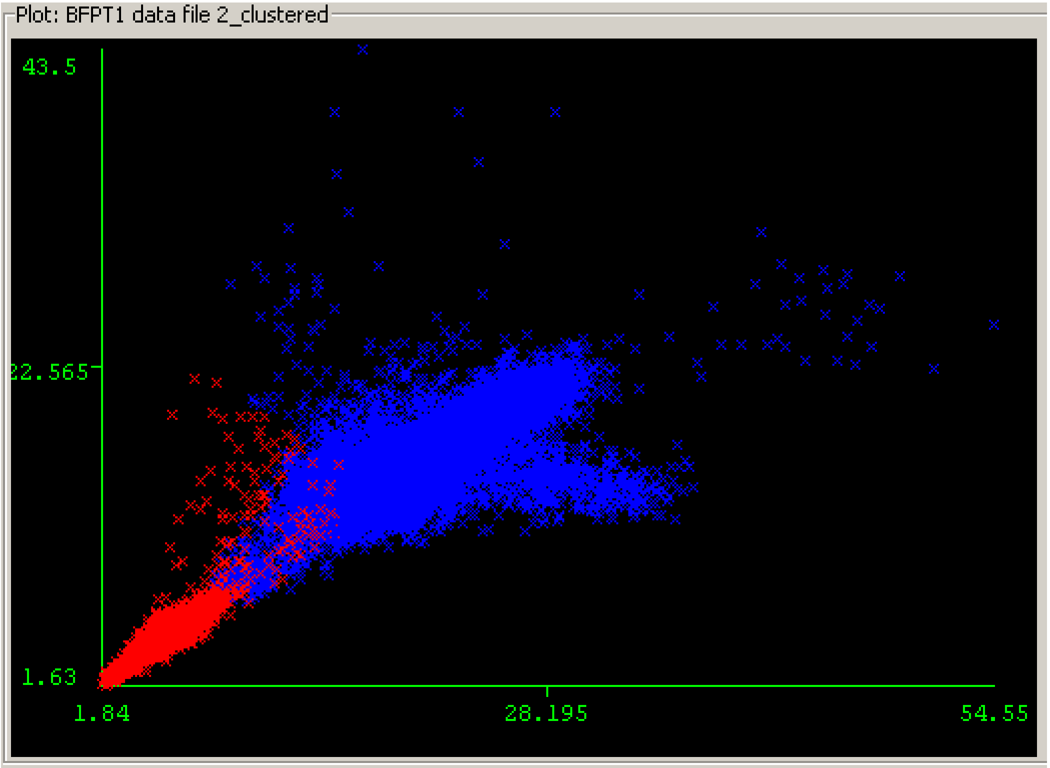


Figure 6.16. LP vs. HP data plot divided into two clusters – from file 2

6.1.3.3 Clustering summary

The analyses performed so far on two files separately show that applying clustering does not provide any useful knowledge. Data distribution graphs suggest that it is expected to have two or three clusters in HP and LP data.

However, those clusters do not correspond to any useful knowledge and cannot be used to design an alarm system. The resultant clusters were only stating the obvious – increasing vibrations on one side should lead to increasing vibrations on the other side. This can be considered useful knowledge, but it was already known, therefore it is not novel.

6.1.4 Combining the data

6.1.4.1 Distribution of combined data

Consequently, another iteration is required. This analysis will focus on combining the data and exploring the whole data range. First step after combining the data is to look at its distribution. Figure 6.19 and 6.20 show the data distribution for HP and LP data in case of combined set. It can be noticed on both that the data tends to concentrate around three areas and consequently, probably three natural clusters could be formed.

Governor valve position distribution data is shown on figure 6.21 in appendix 1. It can be noticed that the data tends to focused around four areas. However, it was decided that governor valve position data will not be used in the analysis process.

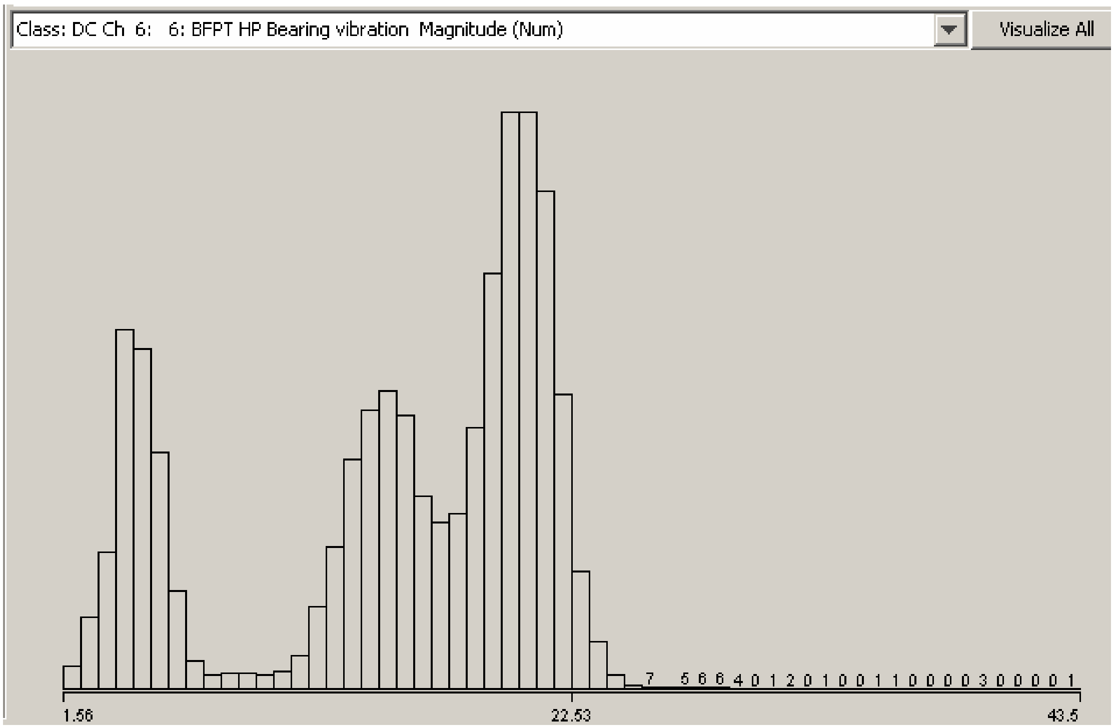

Class: DC Ch 6: 6: BFPT HP Bearing vibration Magnitude (Num)

Figure 6.19. HP data distribution – combined

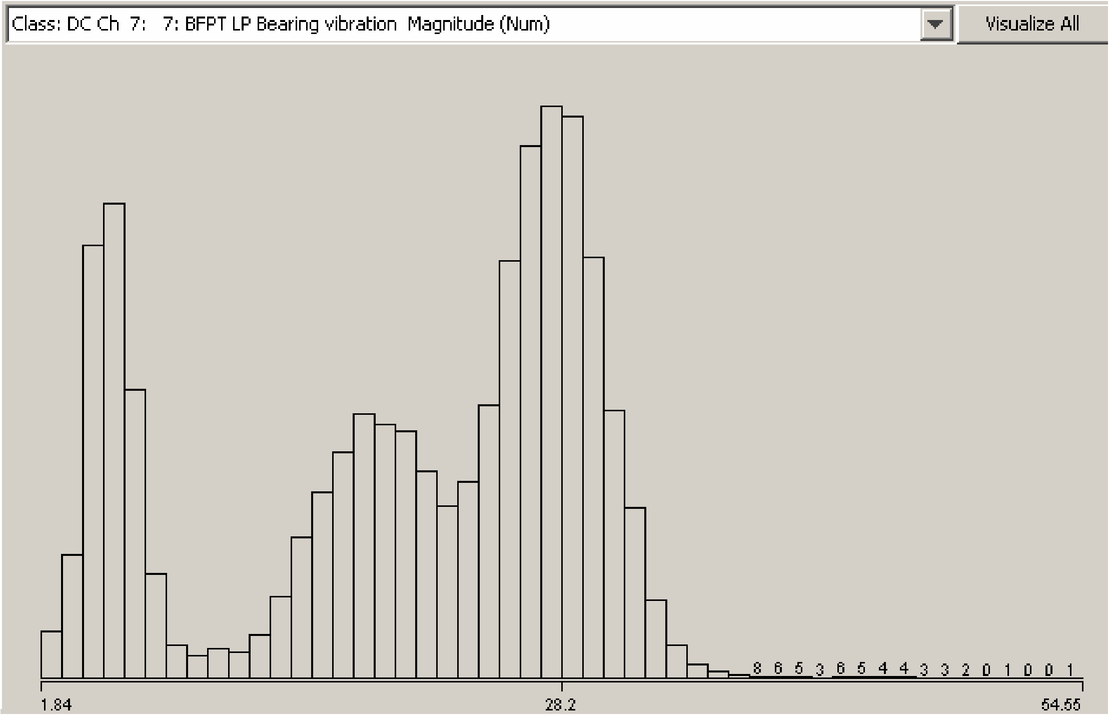

Class: DC Ch 7: 7: BFPT LP Bearing vibration Magnitude (Num)

Figure 6.20. LP data distribution – combined

6.1.4.2 Clustering combined data

After combining the data, for each measured quantity, there were over twenty eight thousand data instances. It was hoped that after combining the data, it would be possible to derive useful knowledge from the new data set.

Again, it was decided to focus on HP and LP data and apply the same clustering method as in previous case – k-means clustering. The resultant data is presented in the same way as before – as an LP versus HP data plot.

Figure 6.22 shows the results with two clusters. It can be noticed that the plot is rather similar to LP versus HP plot for data in the second file. In this case the division between clusters seems to be practically along a straight line and does not give us any additional knowledge.

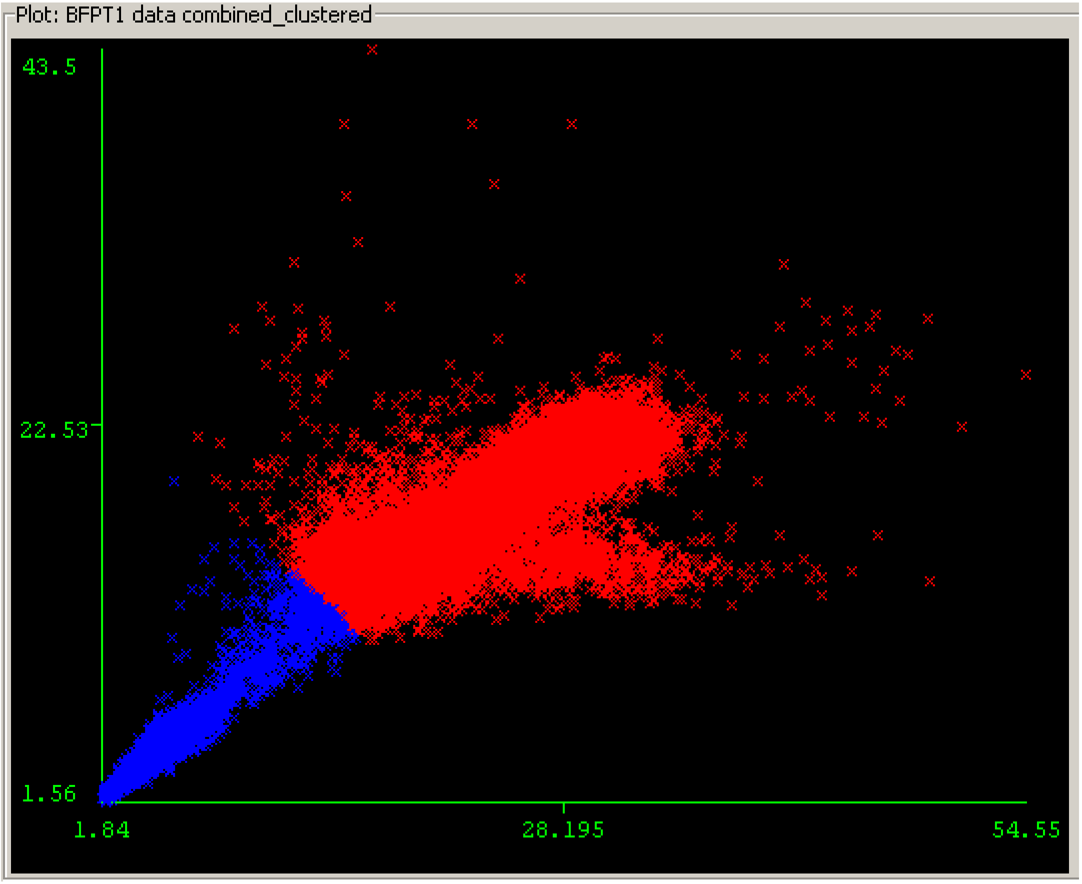


Figure 6.22. LP vs. HP data plot divided into two clusters – combined data

Since the data distribution showed that it can be expected to have three clusters in data, iteration was performed and k-means algorithm with k = 3 was applied. The resultant visualization of clustering can be seen on figure 6.23 in appendix 1. Again, the result does not provide novel or useful knowledge.

In order to confirm that division into further clusters will give similar results as in previous cases, it was decided to perform k-means clustering with k = 5. The visualized results can be observed on figure 6.24 in appendix 1. It can be clearly seen that clusters are divided along inclined lines. This might look nice on a graph, but unfortunately does not provide any additional knowledge.

6.1.4.3 Combined data analysis summary

So far applying k-means clustering for various values of k and different data sets showed that when vibrations on one side increase, it can be expected that the vibrations on the other side to increase as well. Unfortunately, this is not any useful or novel knowledge, because such behaviour was expected, so this solution cannot be used to design an alarm system signalling abnormal machine condition.

Therefore, it is required to have another look at KDD process – from there it can be read that an analyst could go back to any step and iterate in order to obtain expected results. In this case it was decided to go back and attempt to transform the data, present it in a different way and use the new dataset as input into clustering algorithm. In this case study, it was decided to focus on vibration data only, because it was expected that data mining the dependencies between LP and HP side vibrations will provide knowledge, which can be used later on to create an alarm system flagging abnormal boiler feed pump behaviour.

6.1.5 LP/HP ratio

It was decided to keep the focus on HP and LP data and just try to have a look at it from a different point of view. The relationship between LP and HP side vibrations is very important; therefore, it was decided to look at the LP/HP ratio for both data files.

Figure 6.25 shows the plot of LP/HP ratio for file 1 and figure 6.26 in appendix 1 present LP/HP ratio for data in file 2. It can be observed that the plots are roughly at one level with some data instances away from the general trend.

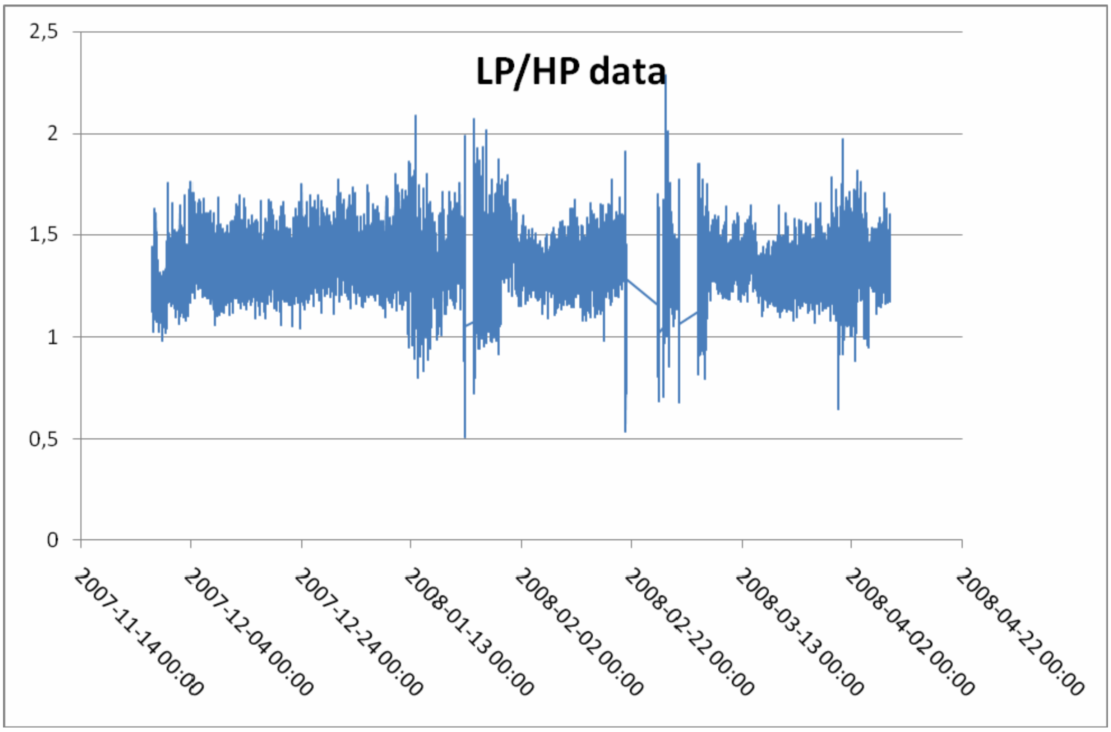


Figure 6.25. LP/HP data from file 1

It is necessary to repeat the KDD process for this case as well, therefore as for previous data, distribution graphs were derived using Weka data mining software.

6.1.5.1 LP/HP ratio data distribution

Figure 6.27 shows the data distribution for LP/HP ratio using instances from file 1. The data is centred around one area, but with spikes indicating possibly a greater number of clusters than one.

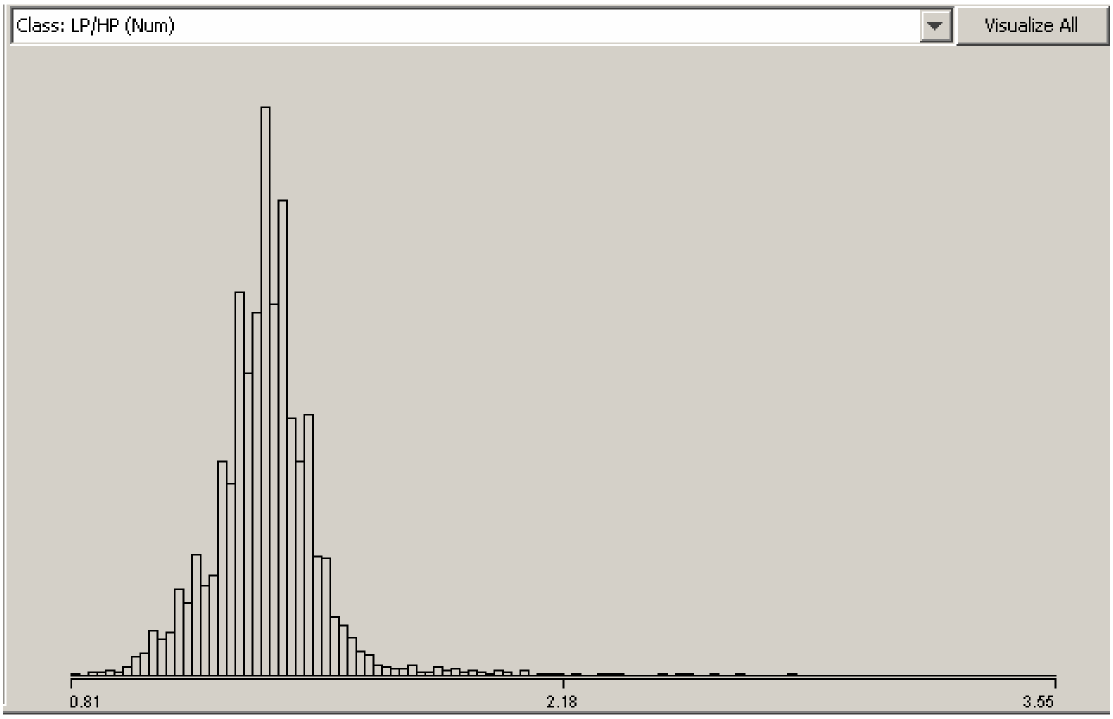


Figure 6.27. LP/HP data distribution – from file 1

Data distribution graph for LP/HP ratio from file 2 is presented on figure 6.28. Again, the instances are clustered around a single centre, but this time it is more uniform with less spikes indicating existence of additional groups.

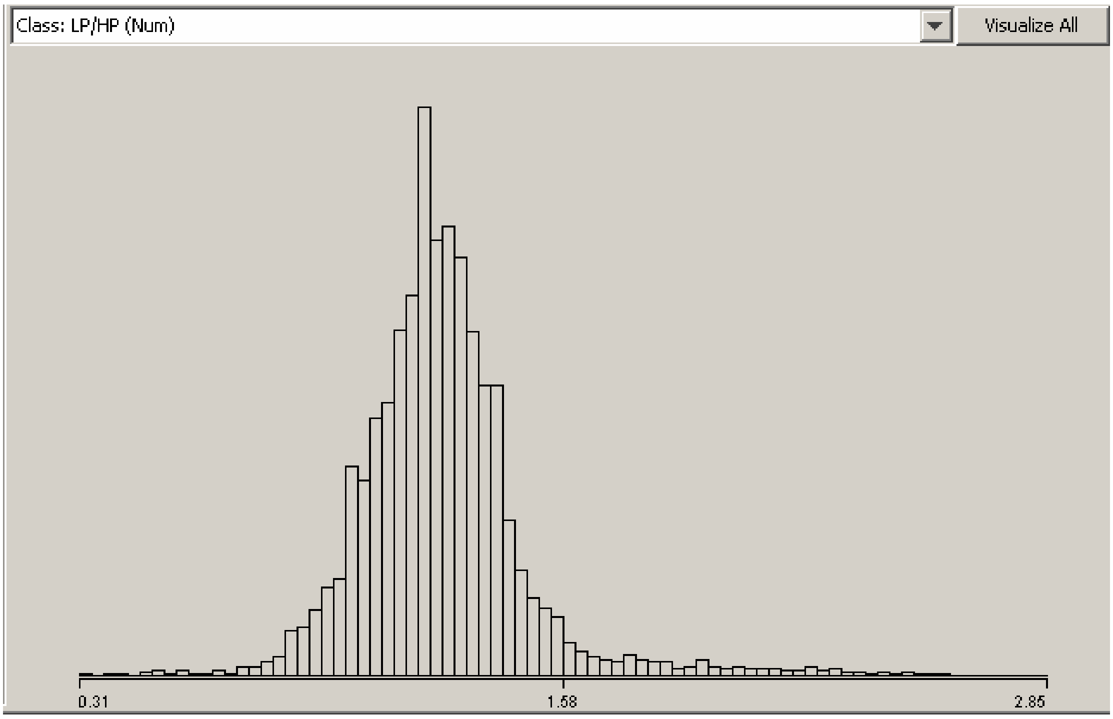


Figure 6.28. LP/HP data distribution – from file 2

6.1.5.2 LP/HP ratio data clustering – file 1

Since the data seems to create natural groups and in order to maintain consistency, again clustering was used, specifically k-means clustering. The analysis was started with parameter k = 2 in order to check if it is possible to distinguish two separate clusters in given data.

The results are presented in a slightly different way than in previous cases. On the horizontal axis is an instance number (corresponding to measurement time) ranging from one to around fourteen thousand. Vertical axis shows LP/HP peak-to-peak vibrations ratio, which does not have any unit.

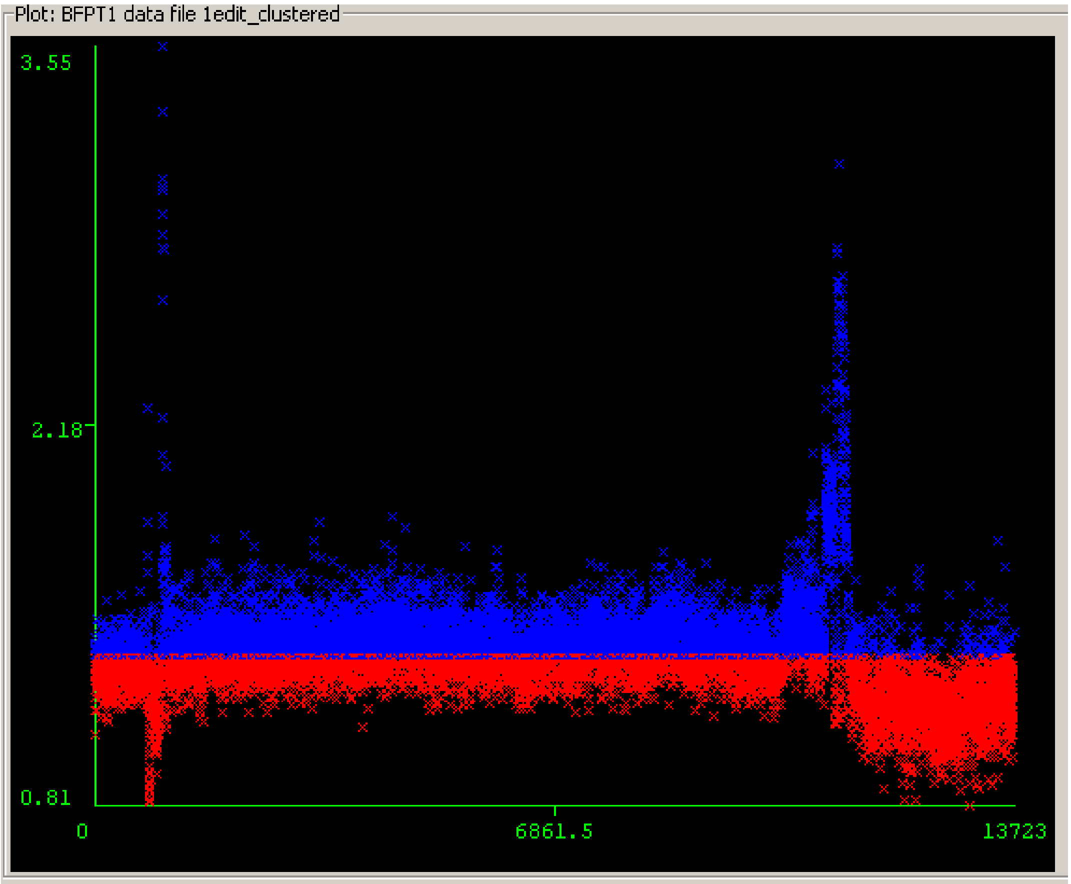


Figure 6.30. Instance number vs. LP/HP ratio divided into two clusters – from file 1

In case of LP/HP ratio for data from file 1, results are shown on figure 6.30. It can be noticed that the data was divided into clusters along a horizontal line practically in the middle of the data range. Unfortunately, this result does not provide knowledge, which it was hoped to extract.

Next step was to increase the number of clusters to k = 3 and apply k-means clustering. Resultant visualization is shown on figure 6.31. This looks slightly more promising, because one cluster (blue coloured instances) seems to contain data a bit further from the main stream.

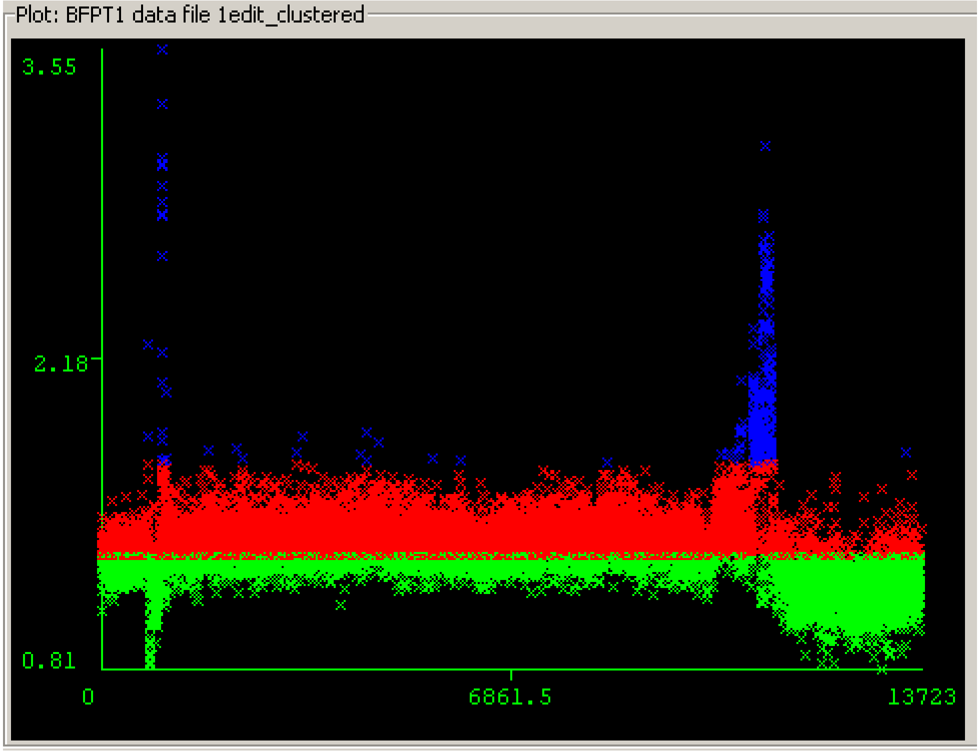


Figure 6.31. Instance number vs. LP/HP ratio divided into three clusters – from file 1

Before deciding on usefulness of results from clustering LP/HP ratio data, it was required to check what happens when parameter k in increased to five. The results are shown on figure 6.32. Similarly as in previous case, data instances with large LP/HP ratio are gathered in one cluster (light blue colour). This might be of use; however, the rest of the data is divided along horizontal lines, what indicates that those are not really natural clusters and are actually forced upon the data. Therefore, it is not of much use.

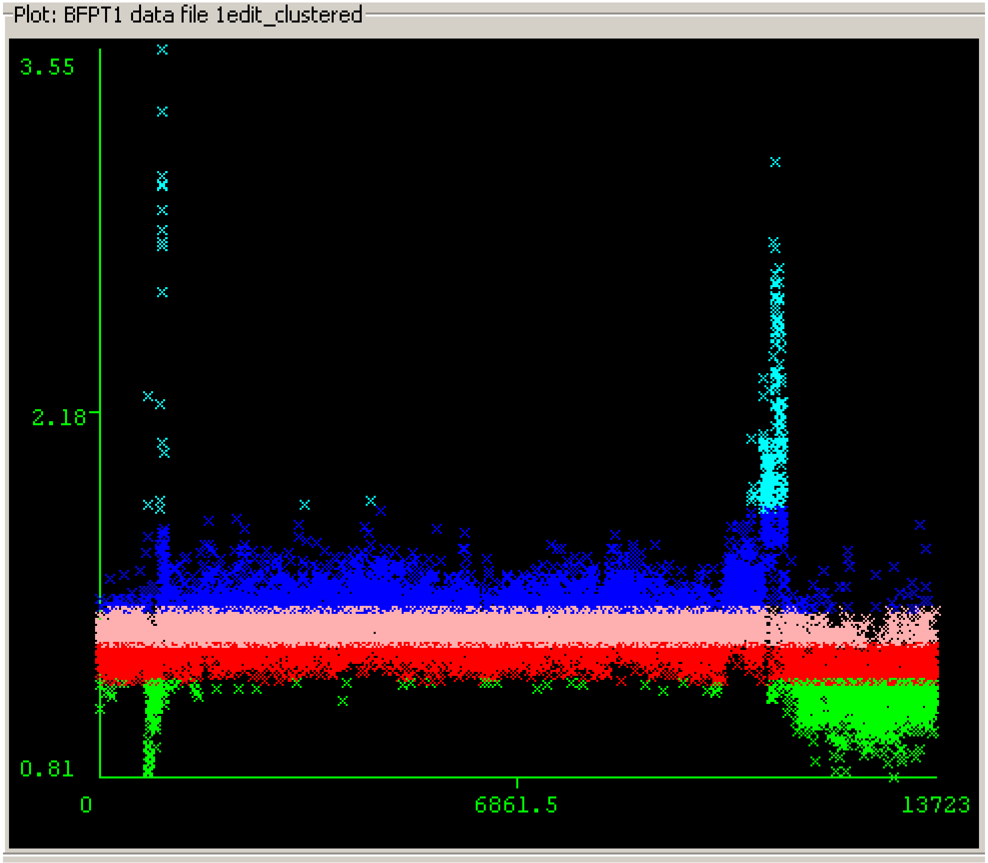


Figure 6.32. Instance number vs. LP/HP ratio divided into five clusters – from file 1

6.1.5.3 LP/HP ratio data clustering – file 2

Since LP/HP ratio from file 1 was not as useful as expected, it was decided to have a look at file 2. Figure 6.28 shows that data is shaped as normal distribution without many spikes indicating additional clusters.

Figure 6.33 shows the visualized result for k = 2 clusters. As previously, the data is divided in the middle and will not be of much use. On figure 6.34 in appendix 1 it can see how data can be divided into three clusters – the divisions occur along horizontal lines. Increasing k to five is shown on figure 6.35 in appendix 1 and it can be confirmed that this trend will continue with more clusters and no useful knowledge can be extracted.

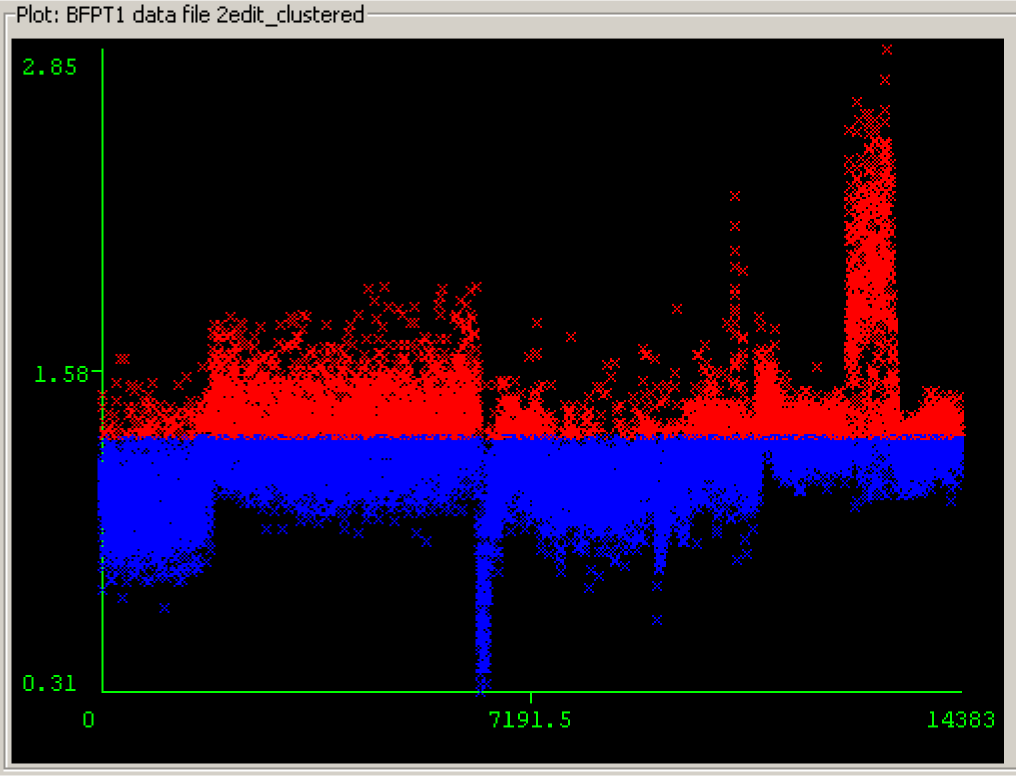


Figure 6.33. Instance number vs. LP/HP ratio divided into two clusters – from file 2

6.1.6 LP/HP ratio – combined data

Looking at LP/HP vibration ratios for separate data files did not prove to be successful. Therefore, in a similar way as with previous data, it was decided to combine it into a single file in order to hopefully extract some useful knowledge from data.

Data distribution (shown on figure 6.36) is similar to the one for file 1 (shown on figure 6.27). It is focused around one area, but there are quite a few spikes indicating that there might be more clusters instead of a single centre.

As with previous analyses, this one also starts with setting k = 2 and trying to divide the data into two clusters. The results are shown on figure 6.37 – it is noticed that due to combining the data, there are over twenty eight thousand instances analyzed in this case. Since k-means algorithm is rather fast and simple, it did not consume significant amount of time on the used workstation. The resultant visualization shows that again there is a horizontal line dividing the instances into two clusters.

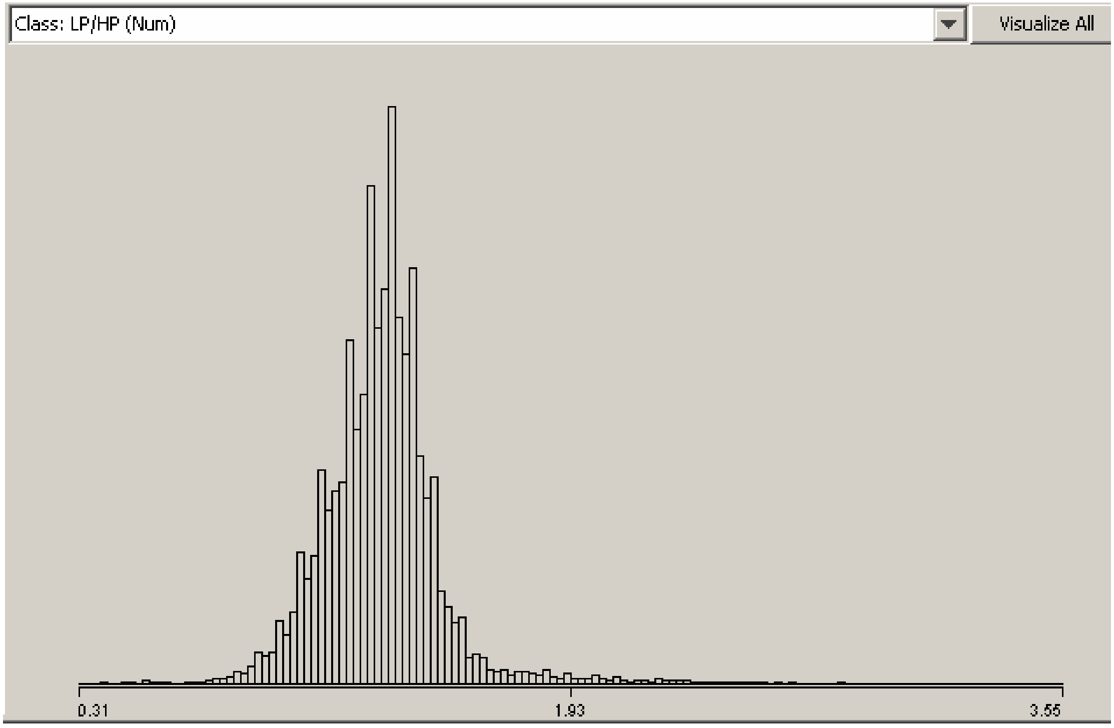


Figure 6.36. LP/HP data distribution – combined

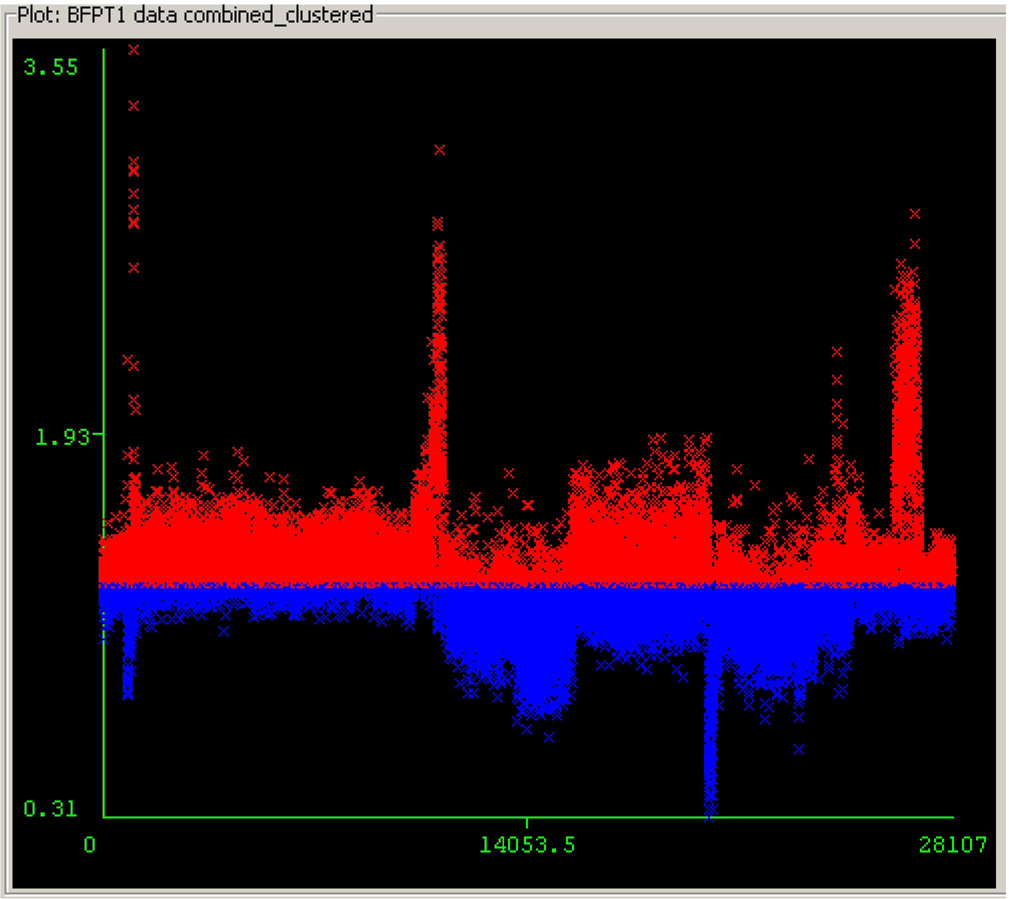


Figure 6.37. Instance number vs. LP/HP ratio divided into two clusters – combined

Next step is to see what happens after increasing the number of clusters to three. Figure 6.38 in appendix 1 shows that as in previous cases, another horizontal line is added and data is divided into three parts now. Changing k to five shows (on figure 6.39 in appendix 1) that this trend will continue and data will be divided further, but without providing any useful knowledge.

6.1.7 Initial analysis summary

In this section a case study was performed to show the application of KDD process to real data. Unfortunately, the process has never been finished, because it was not possible to extract useful or novel knowledge from the data provided by the British Energy.

Applying k-means clustering to raw HP or LP side vibration data only provided knowledge which was obsolete – that with increasing vibrations on one side it can be expected that vibrations on other side increase as well. From LP versus HP plot shown for example on figure 6.13, it can be learned that over the whole data range those two values are highly correlated and it might be difficult to gather unusual data in separate clusters, because there are not many data instances available with data outside main trends.

Converting the raw HP and LP vibration data to a ratio was a good idea, because it allowed visualization of machine behaviour – it was subconsciously expected that the vibration ratio will remain within certain limits and if an instance is outside those limits, it might suggest abnormal behaviour. Unfortunately, applying k-means clustering to LP/HP ratios did not provide useful or novel knowledge. The analysis led to division of data into groups, but unfortunately failed to clearly gather suspicious data in firm clusters. Therefore, it was concluded that there are not enough data instances available to put them in clear separate clusters.

6.2 Final analysis

In this section, the final analysis performed on the data supplied by British Energy is described. This analysis was successful and led to creation of an alarm system working in Microsoft Excel.

6.2.1 Data description

This section deals with analysis of new data supplied by British Energy. The available data is (according to label) – measurement data from Boiler Feed Pump Turbine 2 and is divided into four files:

1) File 1: from 05.05.2008 to 25.08.2008
2) File 2: from 05.02.2008 to 04.05.2008
3) File 3: from 05.11.2007 to 04.02.2008
4) File 4: from 05.08.2007 to 04.11.2007

The files themselves had the same layout as data previously supplied by British Energy and contained the same three variables:

4) High pressure (HP) side peak to peak vibrations – measured in micrometers
5) Low pressure (LP) side peak to peak vibrations – measured in micrometers
6) Governor valve position – measured in percents

The measurements were recorded every ten minutes through the whole data range – same as in previous files. The first step in data analysis process is to look at the raw data. Therefore, on figure 6.40 HP data for file 1 can be seen. Most data instances are hovering around one level, but there are some areas, which might be worth investigating.

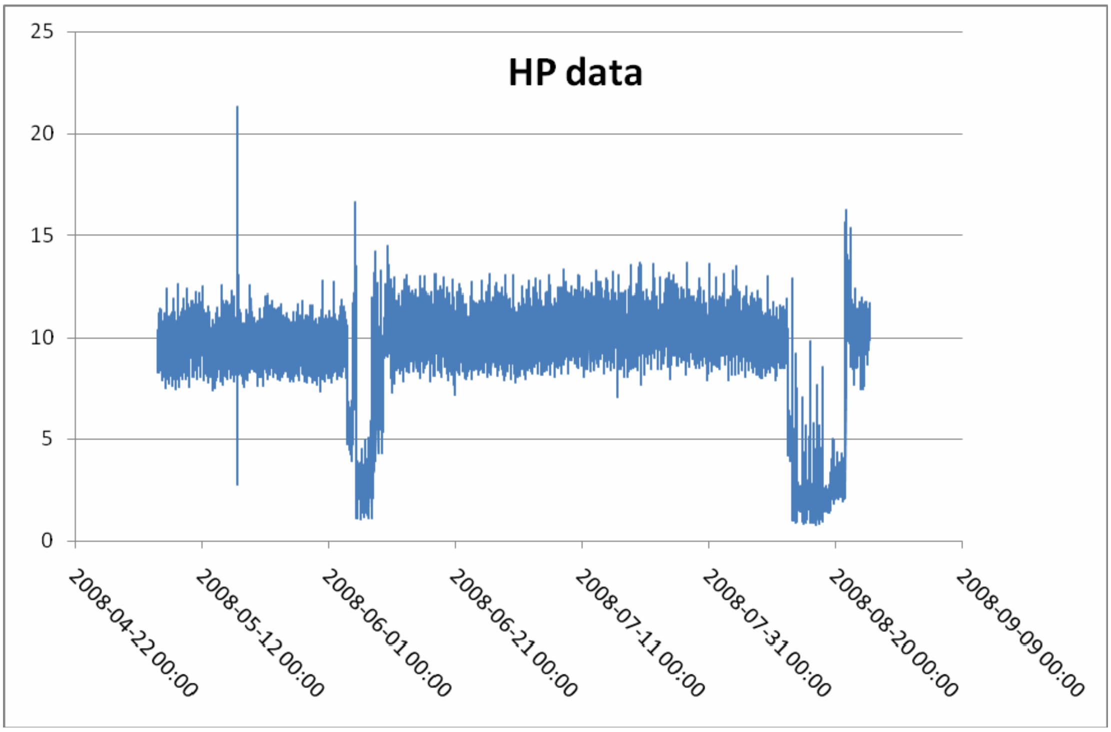


Figure 6.40. HP data from file 1

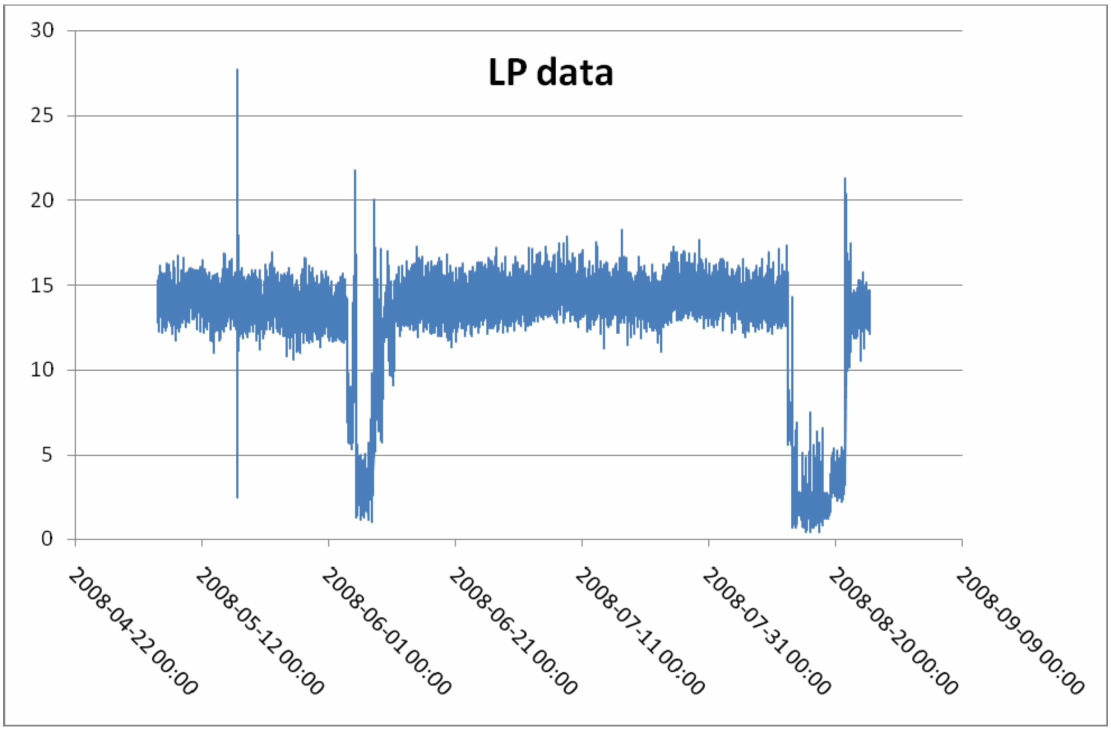


Figure 6.41. LP data from file 1

Figure 6.41 presents to LP data from file 1 – it can be noticed that the profile is quite similar the one of HP data – it might be worth investigating the relationship between those two. On figure 6.42 in appendix 1 the governor valve data can be observed, whose behaviour to some extent can be considered as similar to HP and LP profiles. However, governor valve position data will not be used in the analysis process.

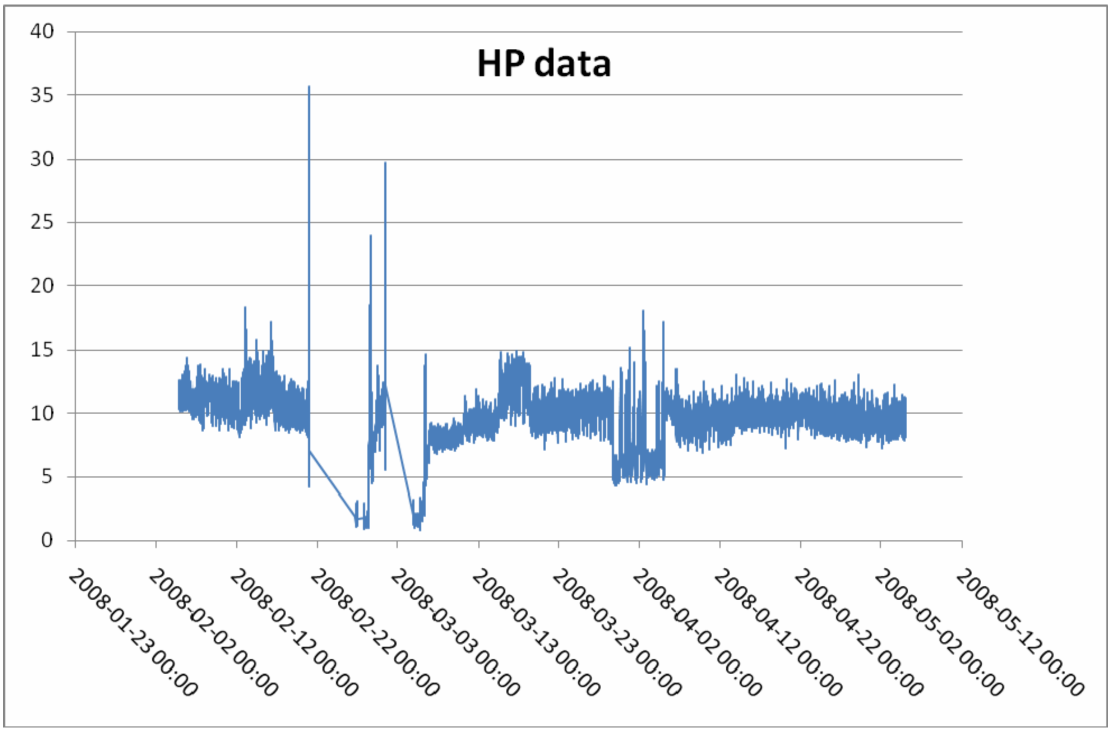


Figure 6.43. HP data from file 2

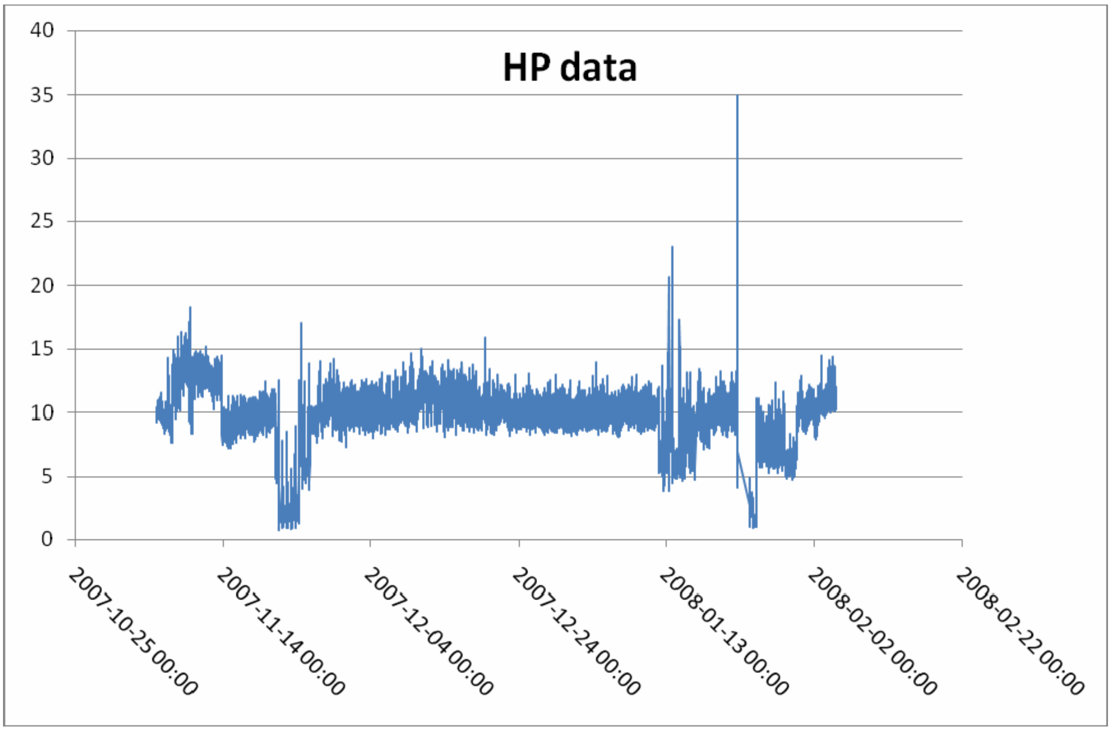


Figure 6.46. HP data from file 3

HP data from the second supplied file can be viewed on figure 6.43. Again, the majority of data instances are around a certain level. The profile is similar for LP data for file 2 shown on figure 6.44 in appendix 1.

On figure 6.45 in appendix 1 the governor valve data values taken from the third file supplied by British Energy can be seen.

Figure 6.49 presents the HP data from file 3 for boiler feed pump turbine 2, while LP data from this file can be viewed on figure 6.50 in appendix 1. Governor valve position data for file 3 is shown on figure 6.51 in appendix 1.

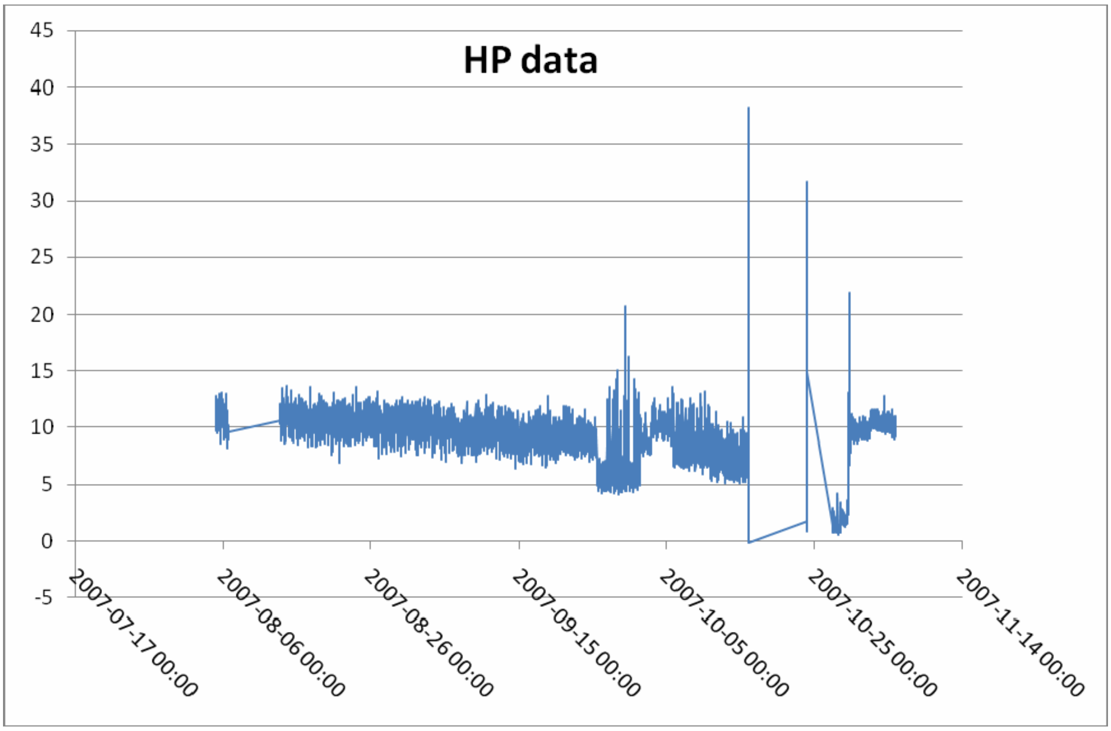


Figure 6.52. HP data from file 4

In case of last file (number four) supplied by the British Energy, the HP data is shown on figure 6.52 and LP data with a similar profile can be viewed on figure 6.53 in appendix 1.

The last graph showing raw data from boiler feed pump turbine 2 is shown on figure 6.54 in appendix 1 – it presents governor valve position data for file number four.

6.2.2 Difference data

In the last subchapter it was described how k-means clustering technique was applied to either raw (HP vs. LP) or transformed data (LP/HP) supplied by the British Energy. Unfortunately, those attempts were unsuccessful and did not provide any useful or novel results which could be later used to design an alarm system flagging up abnormal machine behaviour.

Therefore, after receiving this data set, it was to decided to transform the raw data first and investigate the relationship between LP and HP in order to extract knowledge from the supplied data.

In order to reduce noise and have a better visual representation of data, normalization was applied to the data – each instance was divided by the average of first 1500 measured values. This way there is no need to have the whole database in order to process the data. The next step was applying a moving average of one hundred (100) points to HP and LP data. The resultant dataset is smaller than the original one by 100 points, because due how moving average works, 50 points at the beginning and the end had to be removed.

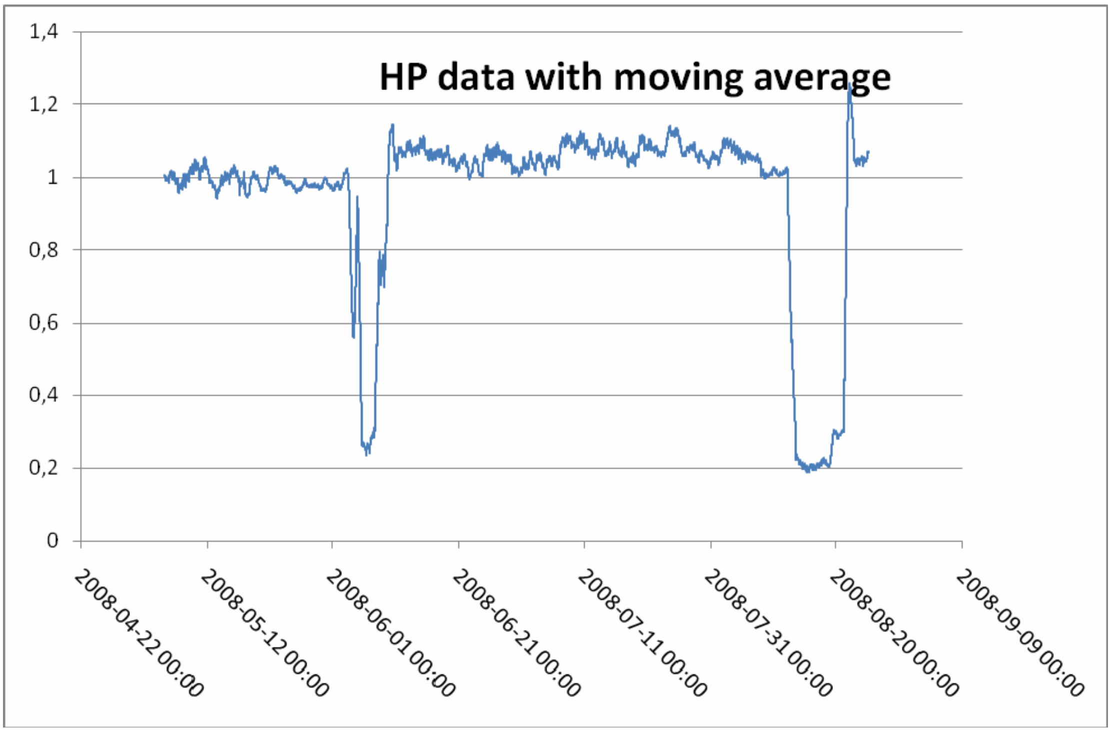


Figure 6.55. Normalized HP data with moving average of 100 points – from file 1

Figure 6.55 shows HP data plot after normalizing and applying moving average of 100 points. Since the data was divided by its average, it is normalized around 1. It can be noticed that the graph is much easier to read and looks much smoother – the sharp edges have been removed and now actually it is possible to see the trends in the data. Additionally, the data is slightly shifted to the right due to being averaged over a certain range.

Figure 6.56 shows LP data after normalizing and applying a moving average of 100 points to the dataset in file 1. The profile has a similar shape to the HP curve and similar consequences of applying moving average can be noticed – the data is less noisy and the profile is much smoother. As with HP data, trends in analyzed data can now be actually noticed.

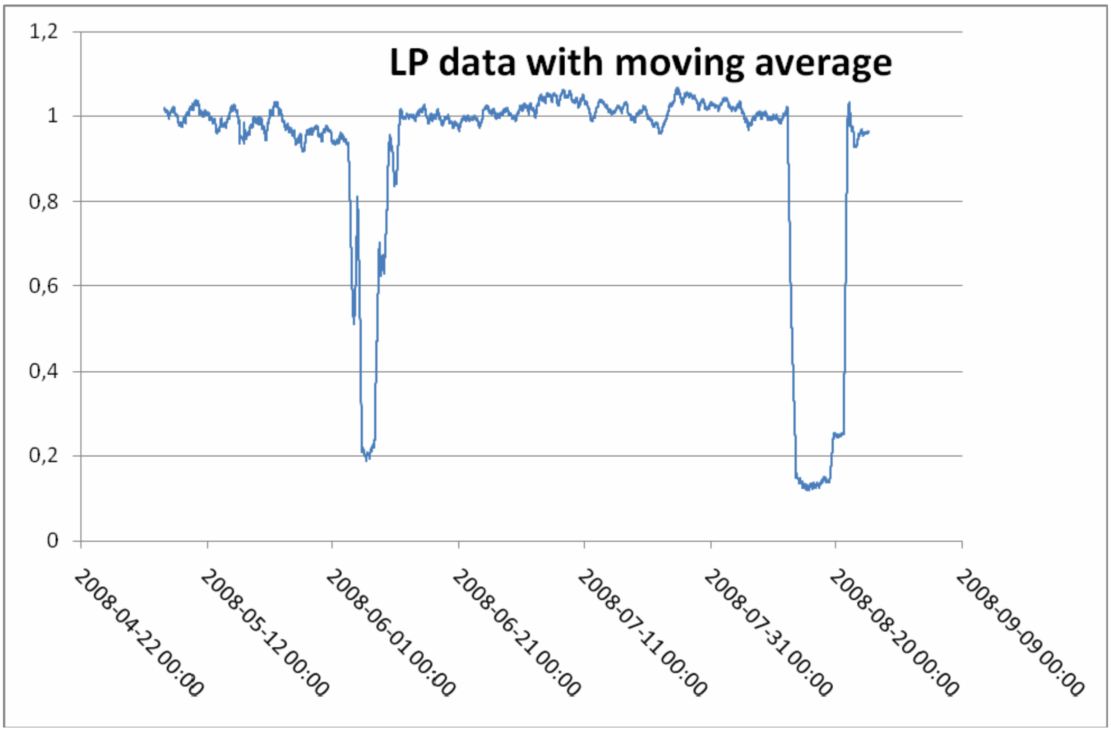


Figure 6.56. Normalized LP data with moving average of 100 points – from file 1

After applying moving average, it was decided to further investigate the relationship between HP and LP side vibrations. Since both datasets are normalized around 1, it is possible to perform operations on both of them. The comparison of trends change between HP and LP side vibrations was of interest.

6.2.2.1 Difference between trends

Since the focus was on a trend change of HP and LP side vibrations, it was decided not to take adjacent points into account, but calculate the difference between instances further apart. In order to choose the best solution, it was decided to investigate differences between points distant by:

a) 100 points
b) 200 points
c) 300 points

Therefore, the difference (for both HP and LP) between a given measurement and a point placed 100, 200 or 300 instances earlier in the data was calculated. This way it was possible to observe how the trends in LP and HP change. Examples for the case of 200-point difference can be observed on figure 6.57 for HP vibration data and on figure 6.58 for LP vibration data.

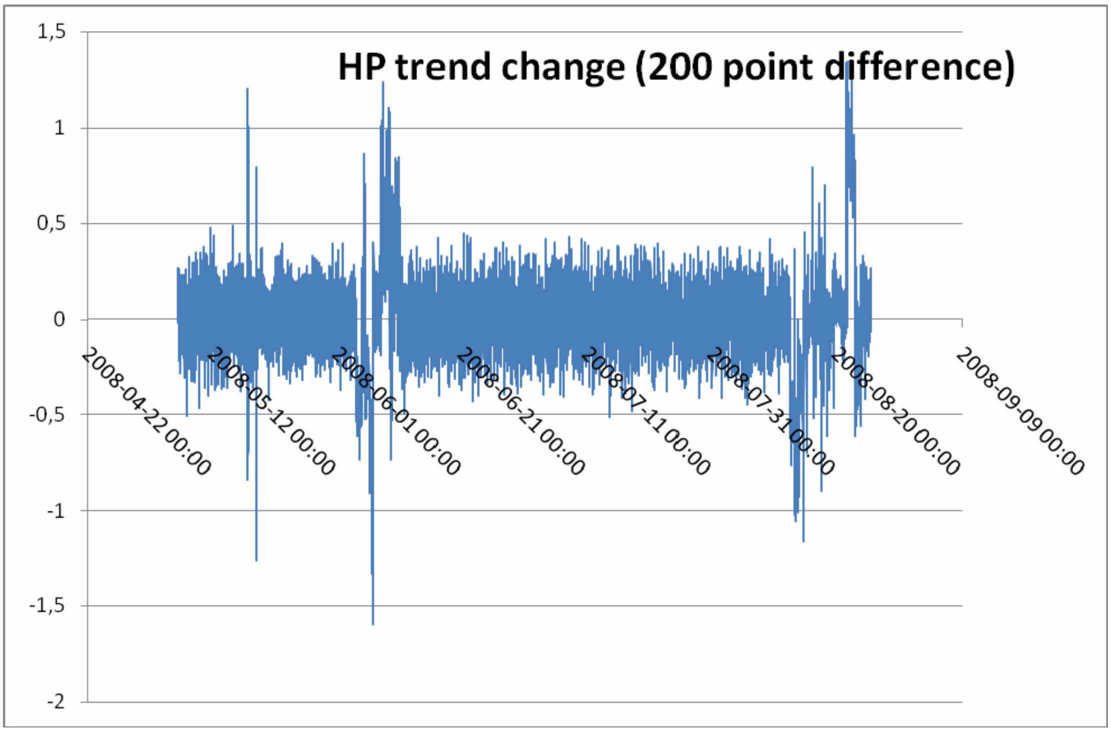


Figure 6.57. HP data trend change with difference of 200 points

It was expected that the HP and LP data will have rather similar trends and if for example one is increasing and the other is decreasing, it might potentially be a problem. Therefore, in order to visualize this difference between two trends, one point difference calculated above was subtracted from the other one (actually it was HP – LP) and moving average of 100 points was applied to the result.

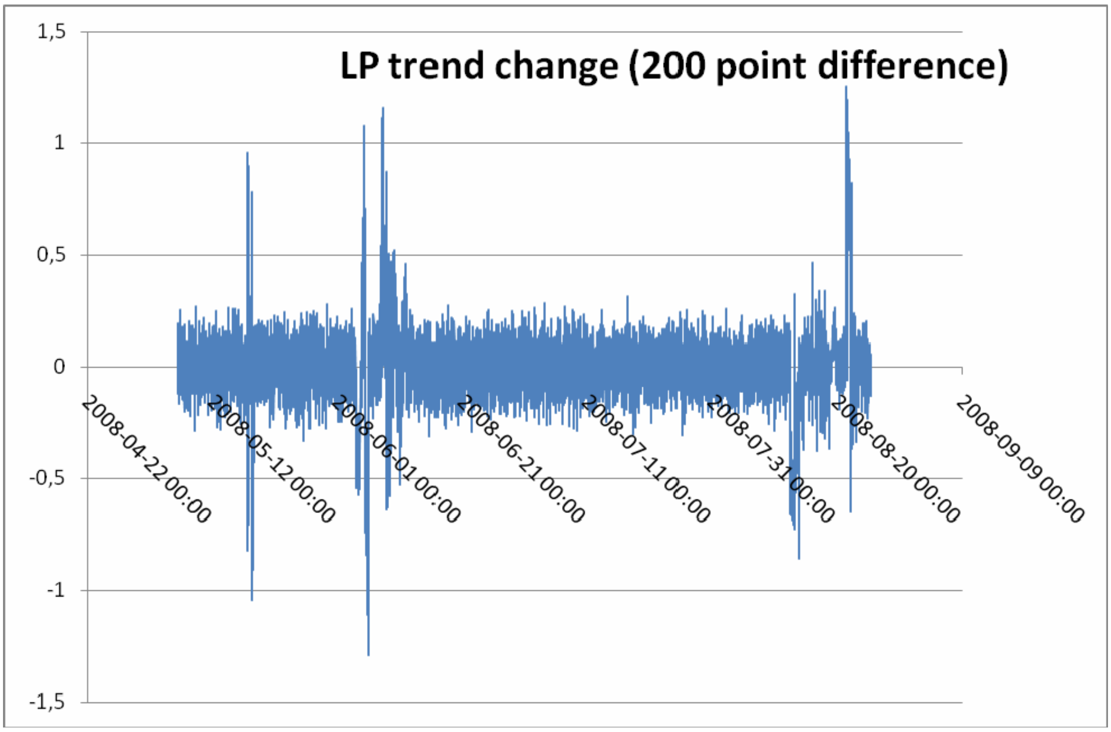


Figure 6.58. LP data trend change with difference of 200 points

The resultant data should allow to visualize potentially problematic behaviour when for example HP vibration are increasing over a period of time and LP vibrations tend to decrease. Such a system can possibly trigger an alarm when such a suspicious behaviour is noticed and should be investigated by an engineer.

In order to select the best solution, as mentioned above it was decided to have a look at comparison between trends calculated from differences between points distant by 100, 200 or 300 instances.

Figure 6.59 shows averaged HP – LP results using 100 point difference and normalized around 1. It is noticed that especially in the beginning, the plot is rather close to 1 and there are a few spikes indicating differences in data. Those areas are possibly problematic and an alarm system should be able to flag a potential problem there. Due to the fact that 100 point difference and 100 point moving average is used, the plot starts after the first 150 data instances.

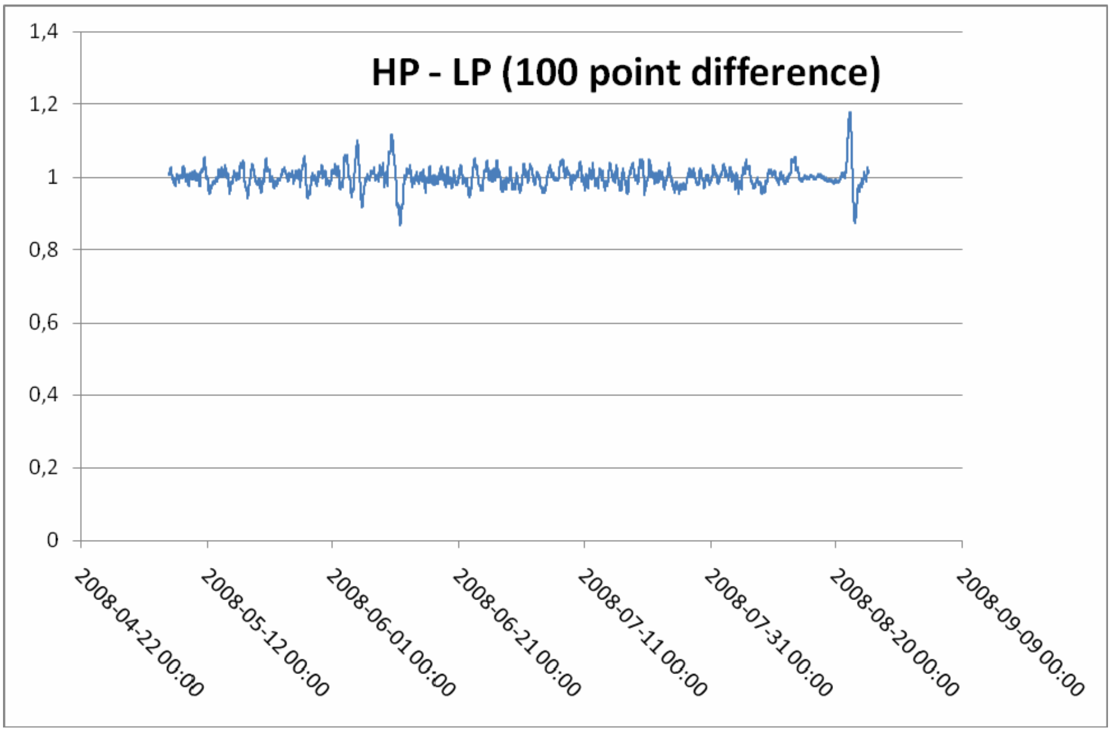


Figure 6.59. HP – LP with difference of 100 points

Next trend comparison was done looking at differences between instances 200 points apart. Figure 6.60 shows the resultant differences after normalizing and applying moving average of 100 points. The plot starts around data instance number 250 due to 200 point difference and 100 point moving average. As in previous case, it can be noticed that the values are close to 1 most of the time, just in areas where HP and LP go in opposite directions, there are noticeable spikes.

In comparison with figure 6.59, where 100 point difference was used, the problematic areas are emphasised with larger spikes in the data and therefore can be easier noticed by a potential alarm system.

The last trend comparison for the case of 300 point difference in order to calculate the vibration behaviour is shown on figure 6.61. Since 300 point difference and 100 point moving average was used, the plot is delayed by 350 points to the raw data. It is noticed that the profile is actually more similar to the one with 100 point difference than the one with 200 point difference. Additionally, with 300 points the problematic areas do not seem to be as strongly emphasised as in case of figure 6.60 – spikes are smaller and less regular.

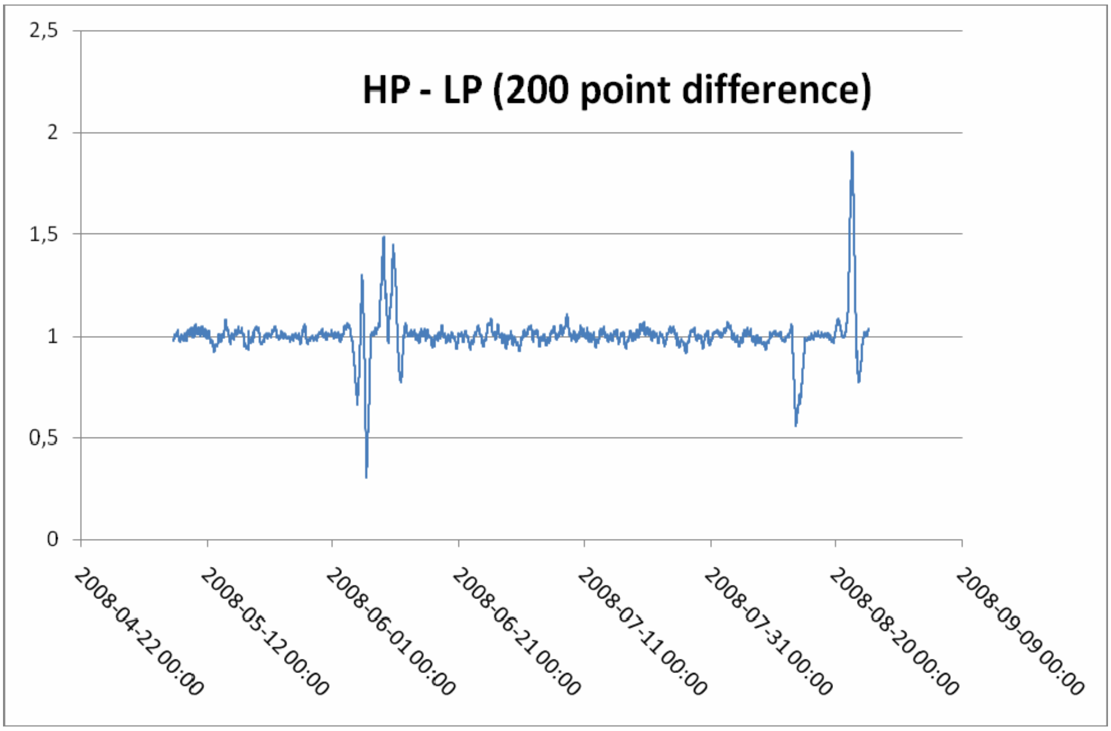


Figure 6.60. HP – LP with difference of 200 points

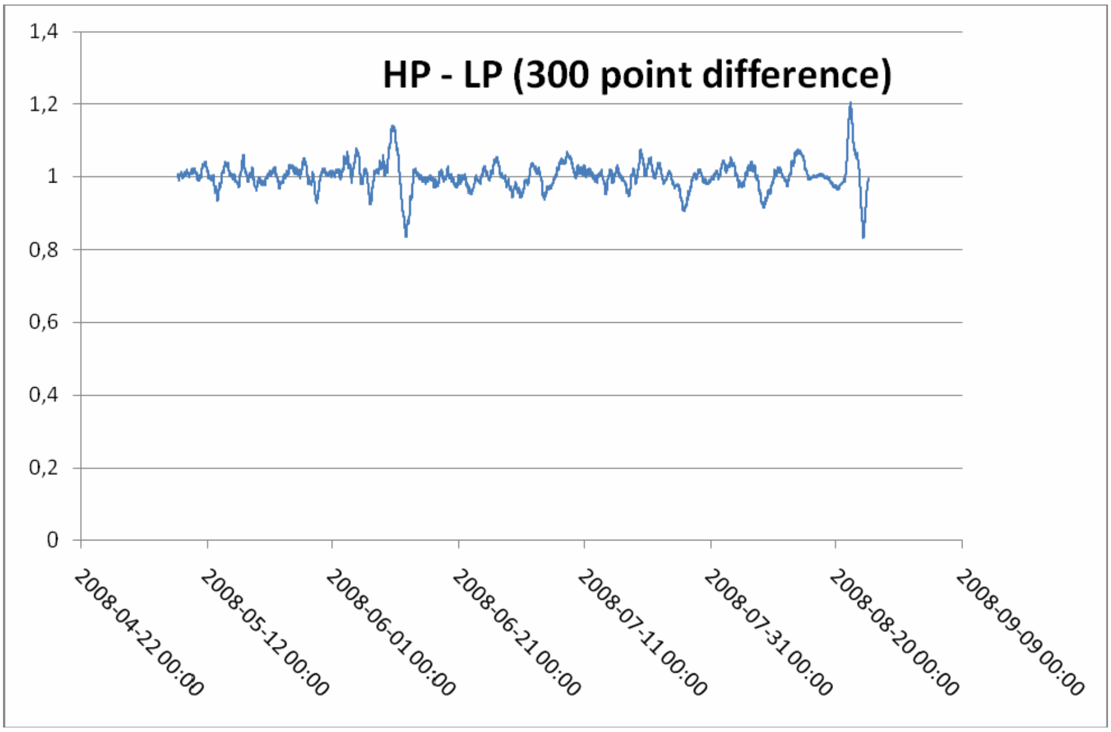


Figure 6.61. HP – LP with difference of 300 points

After looking at three trend comparison algorithms – for 100, 200 and 300 difference between points, the best one to be used for creation of an alarm system is the one with 200 point difference. This is due to the fact that it highlights (with large spikes on the plot) potentially problematic areas while the rest of the profile is fairly stable.

Therefore, it was decided to use this algorithm in order to create an alarm system flagging potential issues. Since it is expected that in the beginning of the operation or measurement there should not be any problems with a boiler feed pump, in order to establish limits for trend comparison it was decided to use the first 1500 points as reference data. From those 1500 points a maximum and minimum values will be extracted, which will be used as upper and lower limits for the alarm system.

Those initial 1500 points equal to about 250 hours or just over 10 days of operation. It will be required to closely monitor the operation of a new machine or a boiler feed pump switched on after maintenance. After that time, the extracted limits can be used to trigger alarm systems.

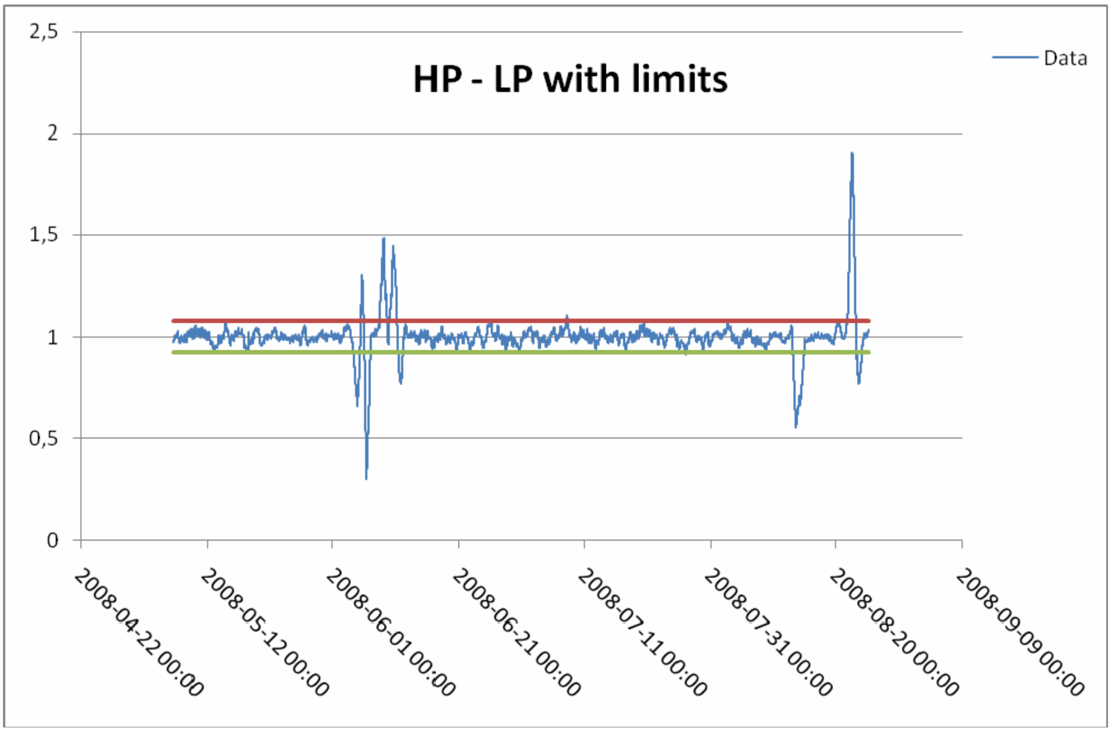


Figure 6.62. HP – LP with limits (file 1)

Figure 6.62 presents an example of how the system will be implemented. As it can be seen, the first 1500 points are used to establish those limits – red line represents the upper limit and green line is the lower limit.

The majority of the data is within the limits, but there are a few significant spikes breaching the limits and triggering the alarm system. Those spikes are consistent with sudden changes in HP and LP vibrations shown on figures 6.55 and 6.56.

Each of those spikes triggering the alarm system should be investigated by an expert and then further classified as a normal or abnormal behaviour.

6.3 Validation

It can be seen on figure 6.62 that the system is able to create limits basing on provided data and check each new data instance if it is within those limits. Otherwise it is flagged as a potential problem.

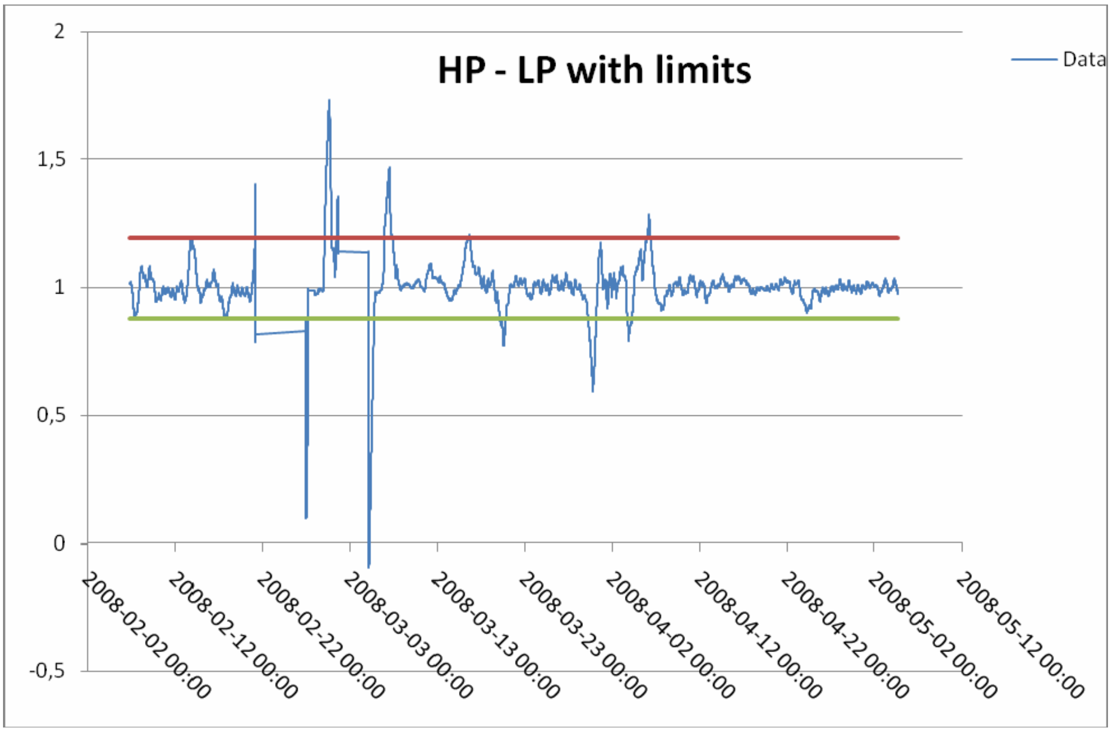


Figure 6.63. HP – LP with limits (file 2)

The system was based on exploration of data in file 1. However, it has to work correctly with other data sets in order to analyze historical data and especially with new data in order to work as a proper condition monitoring system. Therefore, the same algorithm was applied as described earlier to data in other files supplied by British Energy in order to verify that the system will work with them as well.

On figure 6.63 presented results for file 2 can be seen. In this case the initial data is a bit rough and the limits are wider than they were for file 1. However, the limits are still narrow enough to flag large spikes on the plot as potential problematic areas, which need to be investigated by an expert. The limits might be revised after an expert decides that certain machine behaviour is just normal operation or noise.

Applying the derived algorithm to data file number 3 supplied by the British Energy provided the widest limits so far (as seen on figure 6.64). However, even in this case the system is able to notice some large spikes indicating areas, where vibrations on one side are increasing while the trend on the other side shows a decrease. Those results show that the limits are very dependent on the initial 1500 measurements and the machine behaviour has to be carefully observed during those first 10 days after start-up to ensure that the data for normal operation is collected.

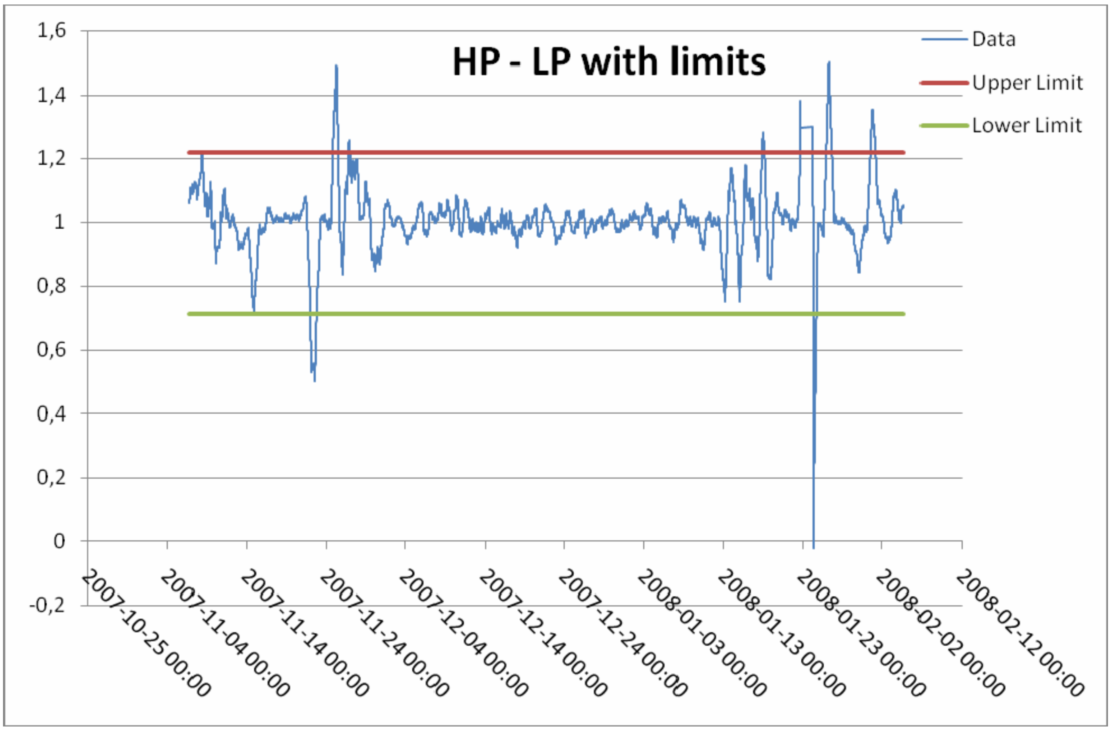


Figure 6.64. HP – LP with limits (file 3)

The last file supplied by the British Energy is actually the one containing the oldest data. Applying the algorithm to this file results in receiving very narrow limits – this is shown on figure 6.65. As in previous cases, there are quite a few large spikes on the plot, which are outside established limits and are triggered as possible abnormal behaviour.

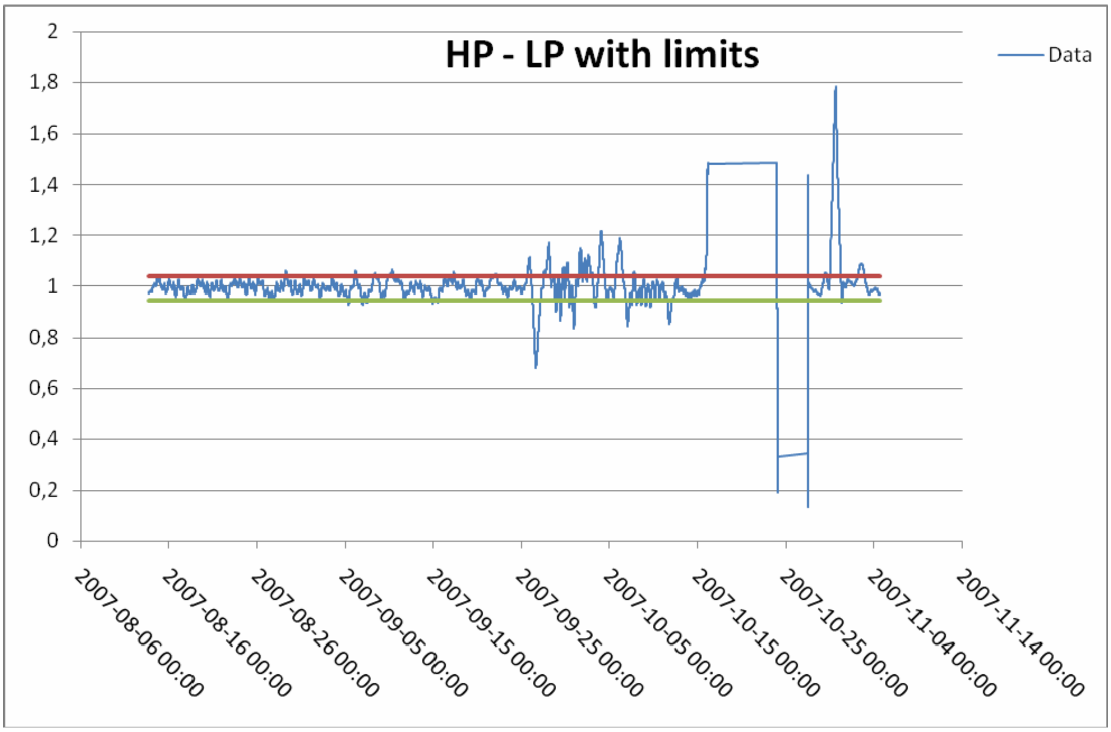


Figure 6.65. HP – LP with limits (file 4)

It was assumed that the data is divided into four files for a reason – maybe a new file was started after a minor maintenance or machine restart. Therefore, the algorithm has been applied in order to calculate the limits for each file separately. However, it is possible that the data was divided into four files for reasons not related with the boiler feed pump equipment (for example for data storage purposes). In such a case, it would be a reasonable assumption that the limits calculated in file 4 correspond to normal machine operation and it is possible to apply them to other files.

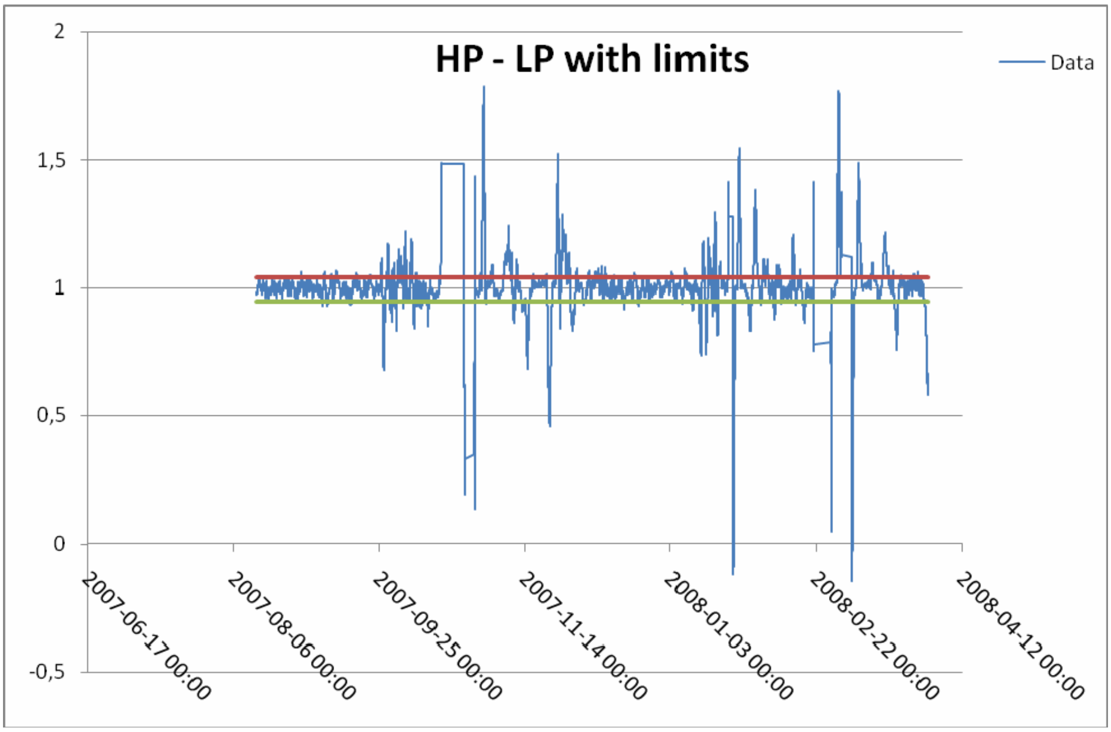


Figure 6.66. HP – LP with limits (combined files)

The results can be seen on figure 6.66. It can be immediately noticed that those limits are very narrow and actually trigger more alarms than limits established for separate files. Many spikes indicating different trends in HP and LP data are outside the calculated limits, but all those breaches still need to be evaluated by an expert who can decide whether it is normal behaviour or a fault.

In this section it was showed that the derived method is able to properly calculate limits from all the files supplied by the British Energy and use as a part of an alarm system responsible for flagging up potentially abnormal behaviour.

6.4 Summary of Excel implementation

The derived algorithm has been implemented in Microsoft Excel, because it is a very popular tool used in practically every office environment and is capable of performing required analysis and presenting the results.

Here is the summary of the algorithm used to create the alarm system:

1) Normalize HP and LP data (divide every data instance by the average of first 1500 points)
2) Apply moving average of 100 points to HP and LP data
3) Calculate the trend difference for both pre-processed HP and LP data using points distant by 200 instances as described earlier in this chapter
4) Calculate the difference HP – LP
5) Apply moving average of 100 points to calculated difference data and add 1
6) Calculate minimum (lower limit) and maximum (upper limit) values for the first 1500 data points
7) Check every new data instance if it is within the limits

This is an outline how my system works in Excel. The main drawback is that it requires 1500 data instances in order to calculate the limits properly and start checking if new instances are within those limits. This equals to over 10 days of operation, when the machine has to be carefully monitored by expert to ensure that it does not develop any faults, because the alarm system will not work yet.

However, during those initial 10 days of operation the upper and lower limits will be dynamic and will be changing based on how many data instances have been entered into the spreadsheet. The limits might actually stabilize after a few hundred points and remain unchanged, but it is essential that during this setup period the plant equipment is closely monitored. Otherwise, the results might be inaccurate and lead to incorrect conclusions.

7. Recommendations and conclusions

7.1 Future recommendations

As mentioned in the previous chapter, the data supplied by British Energy did not arrive with any expert knowledge attached to it. Since it was not known which parts of data corresponded to normal condition, some assumptions about the results had to be made. The most important one was about the difference data. It was assumed that averaged and normalized HP – LP results shown on figures in previous chapter were expected to be close to one for normal operating state. Therefore, it was considered that data outside the limits was corresponding to abnormal state of operation.

The first recommendation or rather actually a necessity is to have the system checked by an expert and see if the alarms triggered by the designed system actually correspond to real abnormal behaviour. This can be used by analyzing historical data with corresponding knowledge or as a carefully monitored part of an on-line system.

After an expert verifies the results of the created algorithm it needs to be tested over a few months with on-line data and possibly adjusted for best performance. If it proves to be successful during the testing period and keeps delivering comprehensive results, then it might be a good idea to develop a full software application written in (for example) Java or C++. This way the data would be connected with the system more directly and the analysis would possibly be faster. Additionally, a proper software application could overcome all the systems limitations due to the fact that it has so far been implemented in Excel – for example data storing and visualization of selected historical data.

Another possible future development of the software could implement case-based reasoning in order to improve the condition monitoring system. For example, a minor breach of imposed limits could be categorized by an expert as normal behaviour and a case could be created, so next time the system does not trigger an alarm. Additionally, if certain breach of limits can be linked to a certain type of fault, a case could be created in order to automatically classify this fault in future.

7.2 Conclusions

The main aim of the project was to create a system, which would analyze supplied boiler feed pump data and trigger an alarm if the input data suggested an abnormal state of operation. This was divided into three parts:

1) Background research – to investigate condition monitoring methods and data analysis techniques
2) Data analysis – to analyze the supplied data and find an underlying pattern which could be used to design an alarm
3) System design – to use the extracted pattern in order to design an alarm system and then validate it using other available data sets

All three objectives have been reached and described in detail in this report. Various examples of condition monitoring techniques available for rotating machinery have been investigated, researched and presented in this thesis.

In terms of data analysis techniques, KDD process with all its advantages has been described in chapter 5 together with detailed explanation what exactly data mining is and how specific techniques actually work.

The most important part of this work was data analysis. In the first part of chapter 6, an unsuccessful data analysis attempt has been described. In this case k-means clustering did not provide any useful or novel knowledge. The pattern derived was already known or at least expected by the analyst. Additionally, the instances which could possibly be interesting, were not numerous enough to be gathered in a firm and stable cluster.

The final analysis of a second data set supplied by the British Energy proved to be successful. In this case various data transformation techniques were used in order to compare how HP and LP side vibration trends were interconnected and how they changed. Basing on reference data the limits were established and a system has been implemented in Excel – breaching the calculated limits would trigger an alarm system.

The designed alarm system has been applied to other data sets supplied by the British Energy and it managed to calculate the limits and highlight possibly problematic data instances.

Now the system has to be validated by British Energy specialists and thoroughly tested on either live or historical data for a few months in order to ensure that it does not trigger false alarms and can correctly notify the operator in case of a fault.

Working with raw data without any knowledge attached to it made it impossible to validate the results. At the moment, it is not possible for the analyst to establish whether the designed system is able to correctly abnormal behaviour, because there is not information what normal operation is and which parts (if any) of data correspond to faulty working conditions. Therefore, before the development of the system can be taken any further, it has to be verified by a boiler feed pump expert, who will be able to add useful knowledge to the data.

References

Chen, P., Toyota, T. and He, Z. (2001). Automated function generation of symptom parameters and application to fault diagnosis of machinery under variable operating conditions. *IEEE Transactions on Systems, Man, and Cybernetics*, **31**, 775 – 781

Chien, C. F., Chen, S. L. and Lin, Y. S. (2002). Using Bayesian network for fault location on distribution feeder. *IEEE Transactions on Power Delivery,* **17**, 785 – 793

Devaney, M. and Cheetham, B. (2005). Case-Based Reasoning for Gas Turbine Diagnostics. *In 18th International FLAIRS Conference*.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine, AAAI.*

Gemmell, B. D. (1995). *A consultative expert system for intelligent diagnosis on steam turbines*. Unpublished PhD thesis. Glasgow: University of Strathclyde

Gonzalez, A. J., Osborn, R. L. and Kemper, C. T. (1986). Online diagnosis of turbine-generators using artificial intelligence. *IEEE Transactions on Energy Conversion,* **EC-1**, 68 – 74

Guo, H., Jack, L. B. and Nandi, A. K. (2005). Feature generation using genetic programming with application to fault classification. *IEEE Transactions on Systems, Man, and Cybernetics,* **35**, 89 – 99

Haji, M. And Tiliyat, H. A. (2001). Pattern recognition – A technique for induction machines rotor broken bar detection. *IEEE Transactions on Energy Conversion,* **16**, 312 – 317

Koscielny, J. M. (2001). *Diagnostics of automated industrial processes*. Warsaw: EXIT.

Luger, G. F. and Stubblefield, W. A. (1998). *Artificial Intelligence: Structures and Strategies for complex Problem Solving*. 3rd edition. Reading: Addison-Wesley.

McArthur, S. D. J, Strachan, S. M. and Jahn, G. (2004). The design of a multi-agent transformer condition monitoring system. *IEEE Transactions of Power Systems,* **19**, 1845 – 1852

Mejia-Lavelle, M. and Rodriguez-Oritz, G. (1998). Obtaining expert system rules using data mining tools from a power generation database. *Expert Systems with Applications,* **14**, 37 – 42

Milne, R. and Nicol, C. (1997). *TIGER: Intelligent Continuous Monitoring of Gas Turbines*. London: The Institution of Electrical Engineers.

Myerscough and W. Wright (eds.) *Modern Power Station Practice*. 3rd edition. Oxford: Pergamon Press plc.

Piotrowski, J. (2001). *Pro-Active Maintenance for Pumps* [Website]. Available from: <http://www.pump-zone.com>

Simon, H. A. (1983). Why should machines learn? *In* R. S. Michalski, J. G. Carbonnell and T. M. Mitchell (eds.) Machine Learning: An Artificial Intelligence Approach. Morgan Kaufmann: Los Altos.

Stanek, M., Morari, M. and Frohlich, K. (2001). Model-aided diagnosis: An inexpensive combination of model-based and case-based condition assessment. *IEEE Transactions on Systems, Man, and Cybernetics*, **31**, 137 – 145

Todd, M., McArthur, S. D. J., McDonald, J. R. and Shaw, S. J. (2007). A Semiautomatic Approach to Deriving Turbine Generator Diagnostic Knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, **37**, 979 – 992

Watson, I. (1997). *Applying Case-Based Reasoning: Techniques for Enterprise Systems.* San Francisco: Morgan Kaufmann.

Witten, I. H. And Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations*. San Francisco: Morgan Kaufmann.

Woodward, A. R., Howard, D. L. and Andrews, E. F. C. (1991). Condensers, pumps and cooling water plant. *In* D. J. Littler, E. J. Davies, H. E. Johnson, F. Kirkby, P. B.

Appendix 1 – additional figures

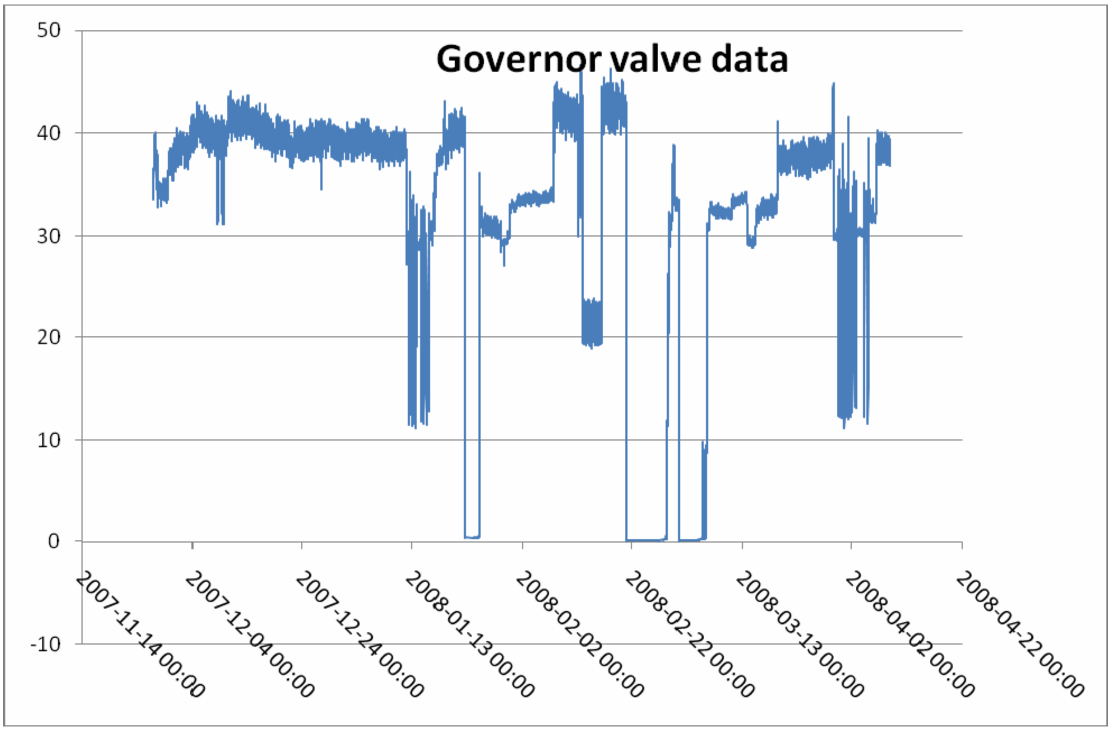


Figure 6.3. Governor valve data from file 1

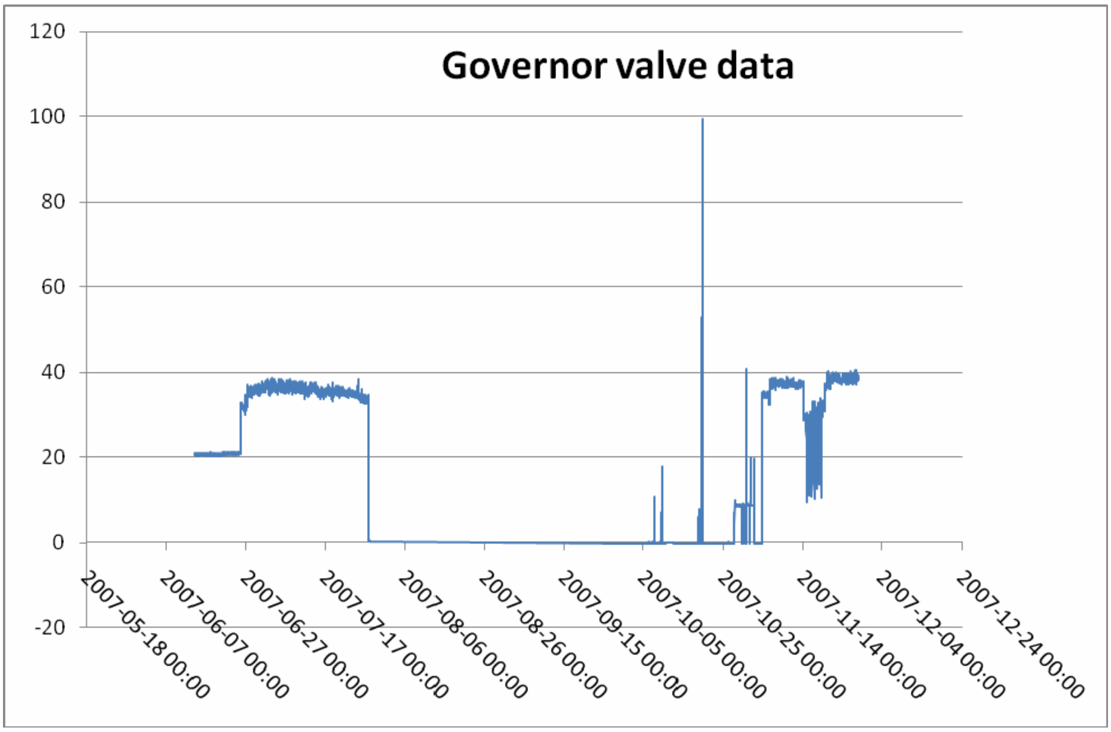
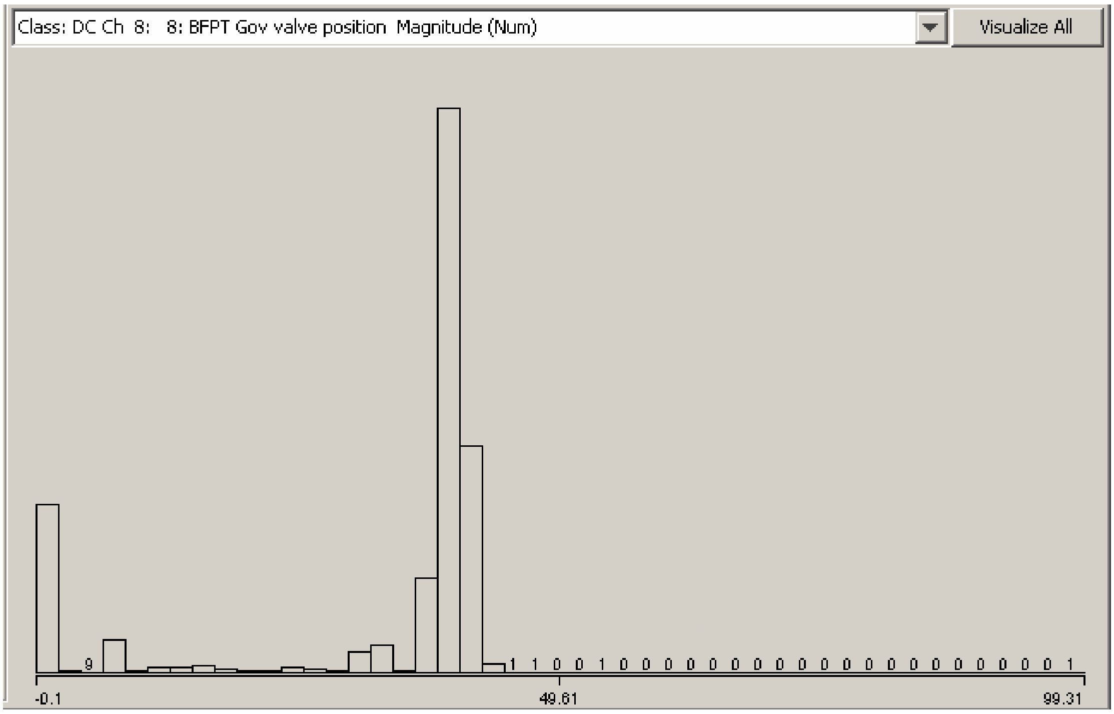
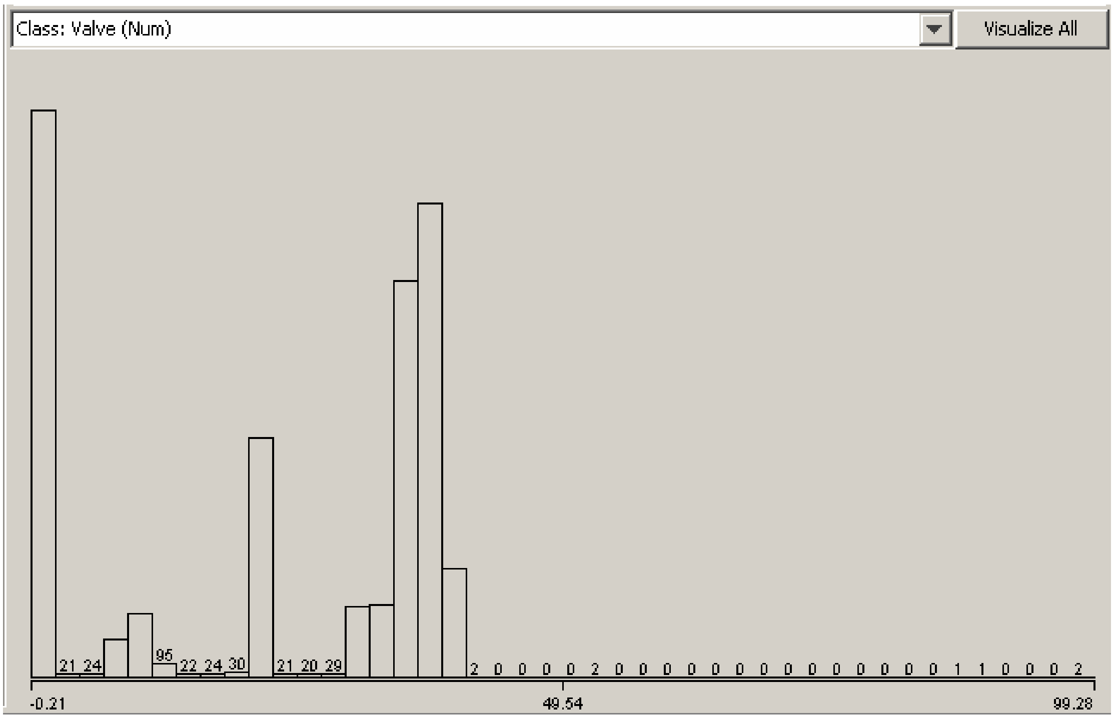


Figure 6.6. Governor valve data from file 2



Class: DC Ch 8: 8: BFPT Gov valve position Magnitude (Num)

Figure 6.9. Governor valve data distribution – from file 1



Class: Valve (Num)

Figure 6.12. Governor valve data distribution – from file 2

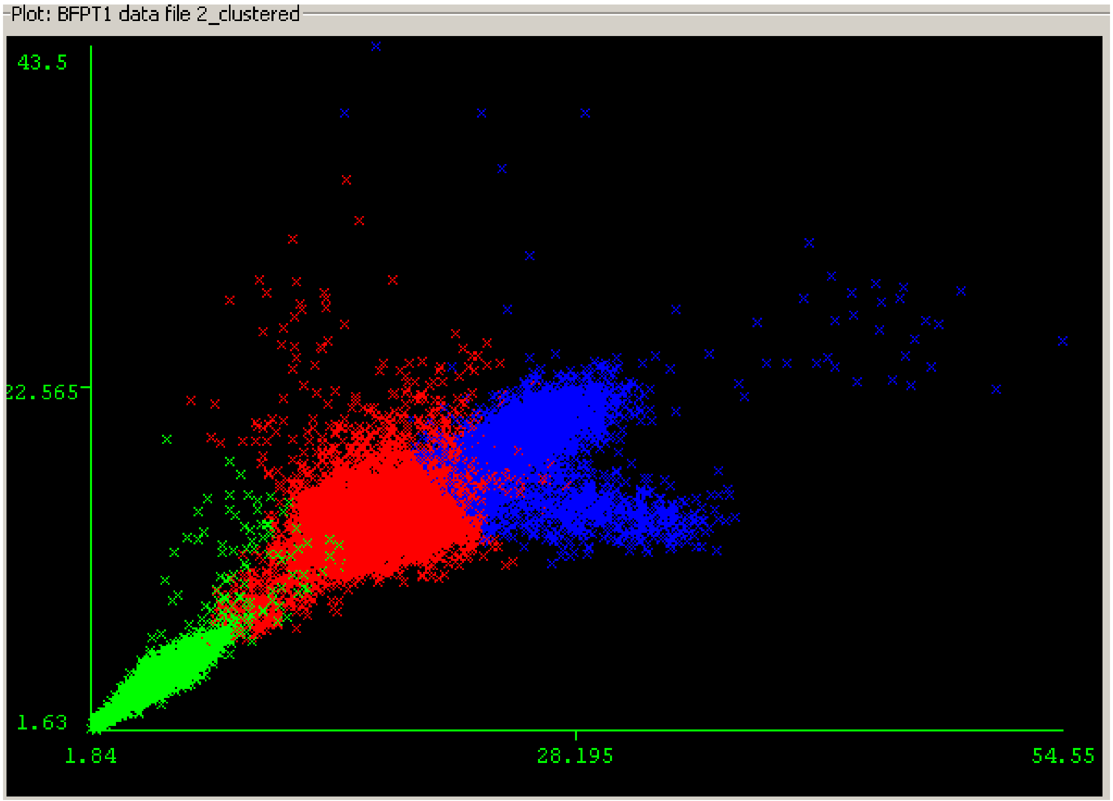


Figure 6.17. LP vs. HP data plot divided into three clusters – from file 2

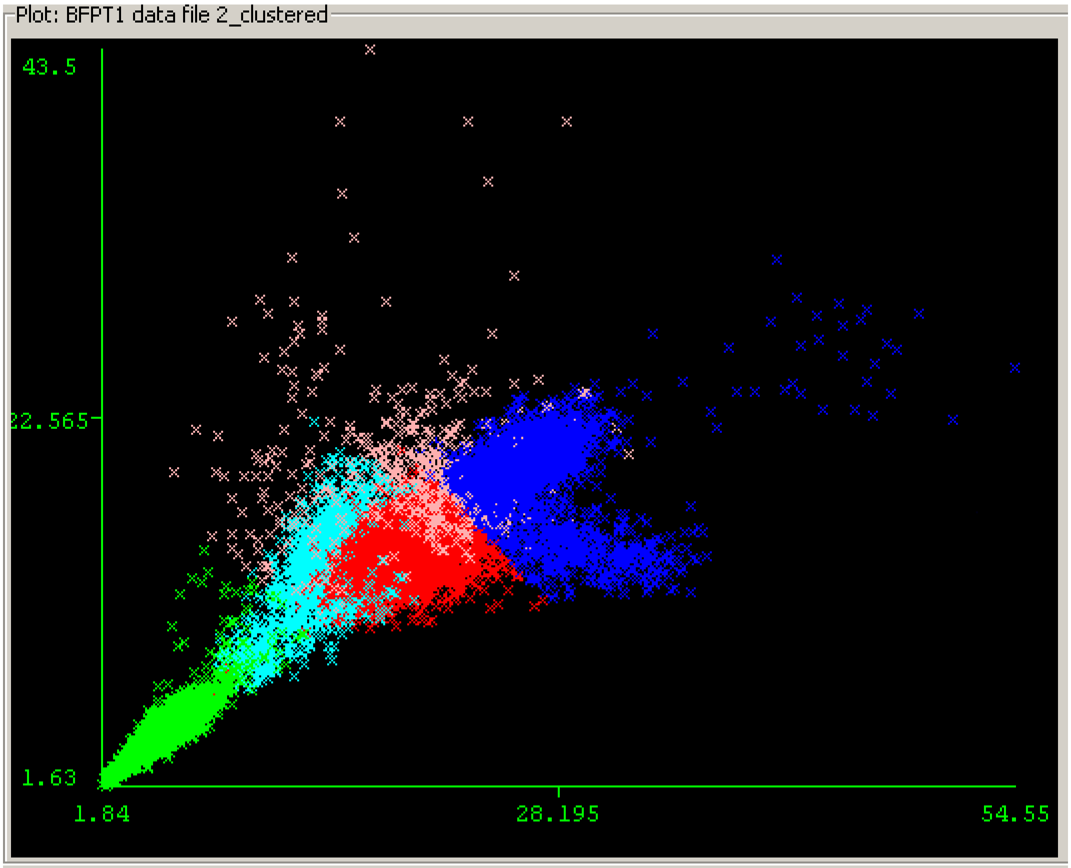
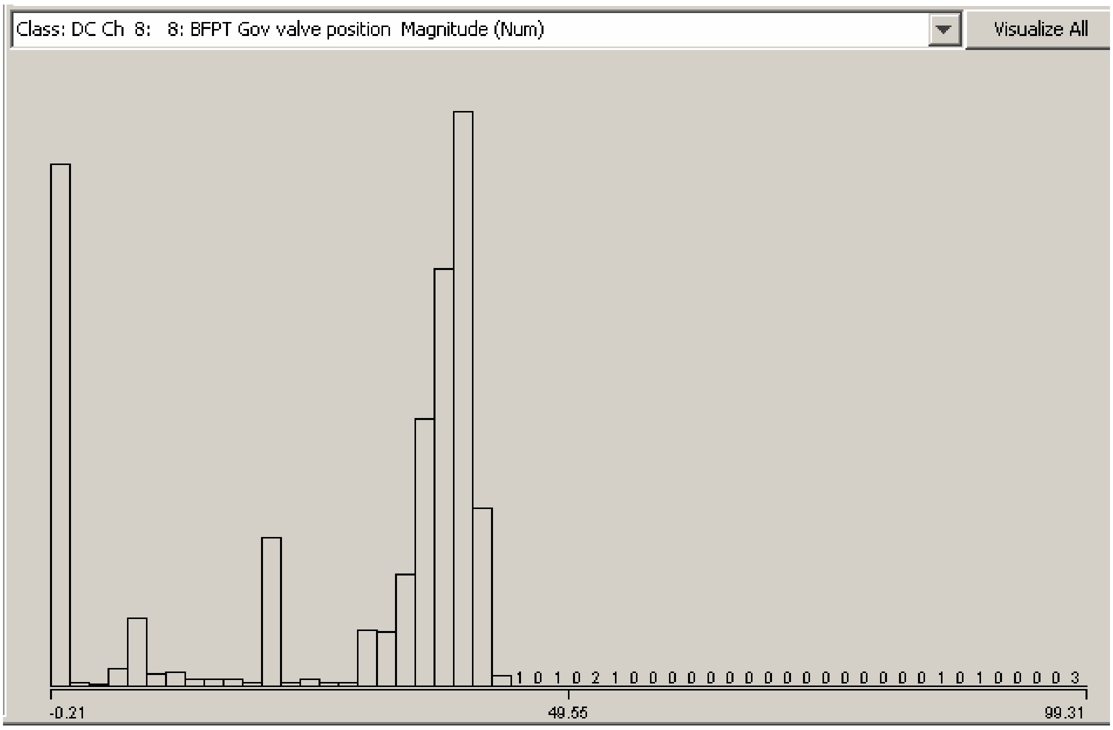


Figure 6.18. LP vs. HP data plot divided into five clusters – from file 2



Class: DC Ch 8:   8: BFPT Gov valve position Magnitude (Num)

Figure 6.21. Governor valve File 1 combined distribution

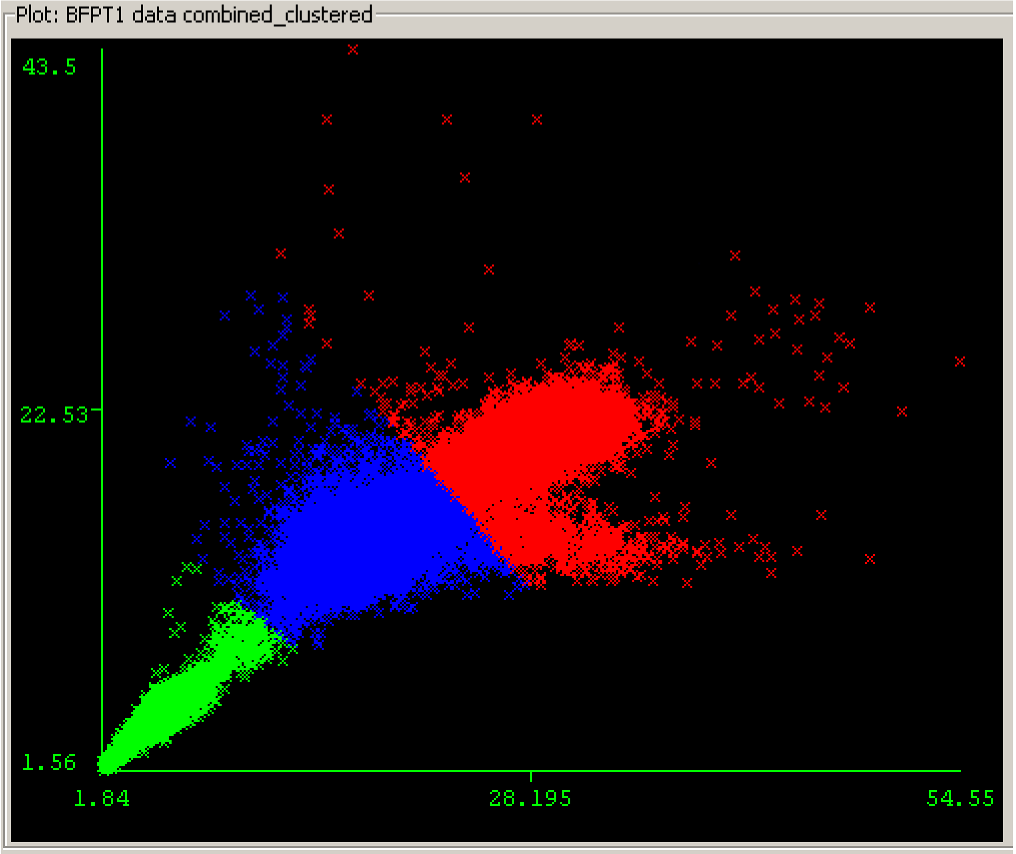


Figure 6.23. LP vs. HP data plot divided into three clusters – combined data

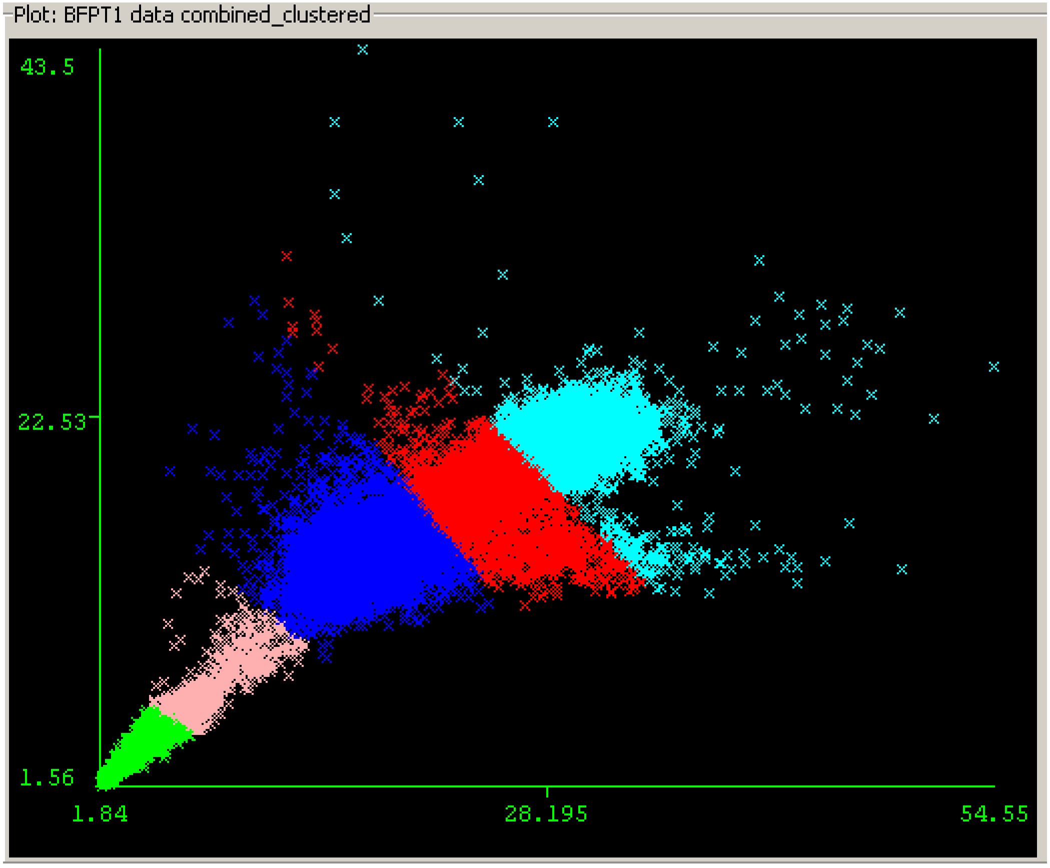


Figure 6.24. LP vs. HP data plot divided into five clusters – combined data

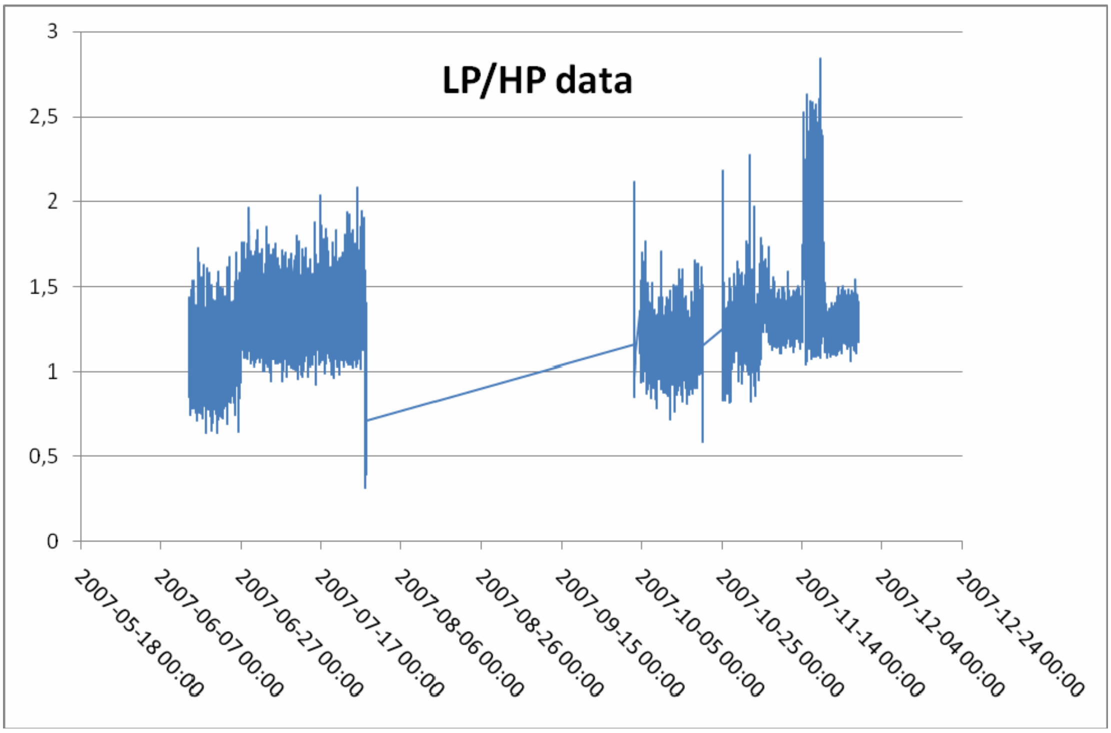


Figure 6.26. LP/HP data from file 2

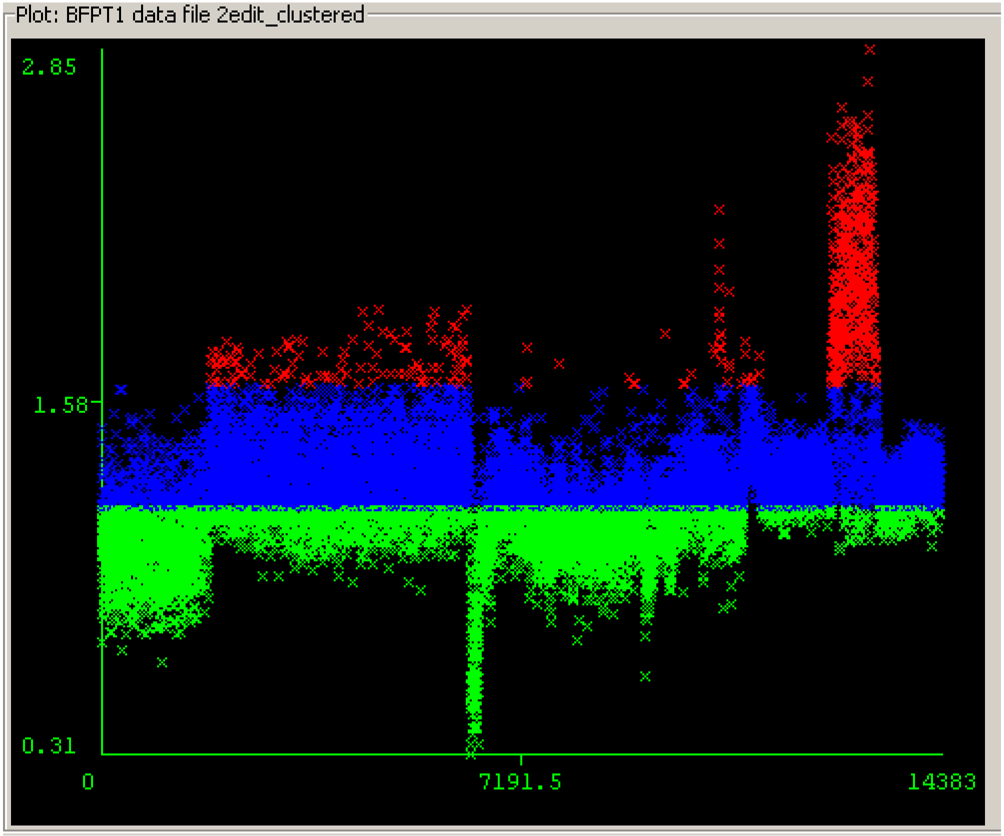


Figure 6.34. Instance number vs. LP/HP ratio divided into three clusters – from file 2

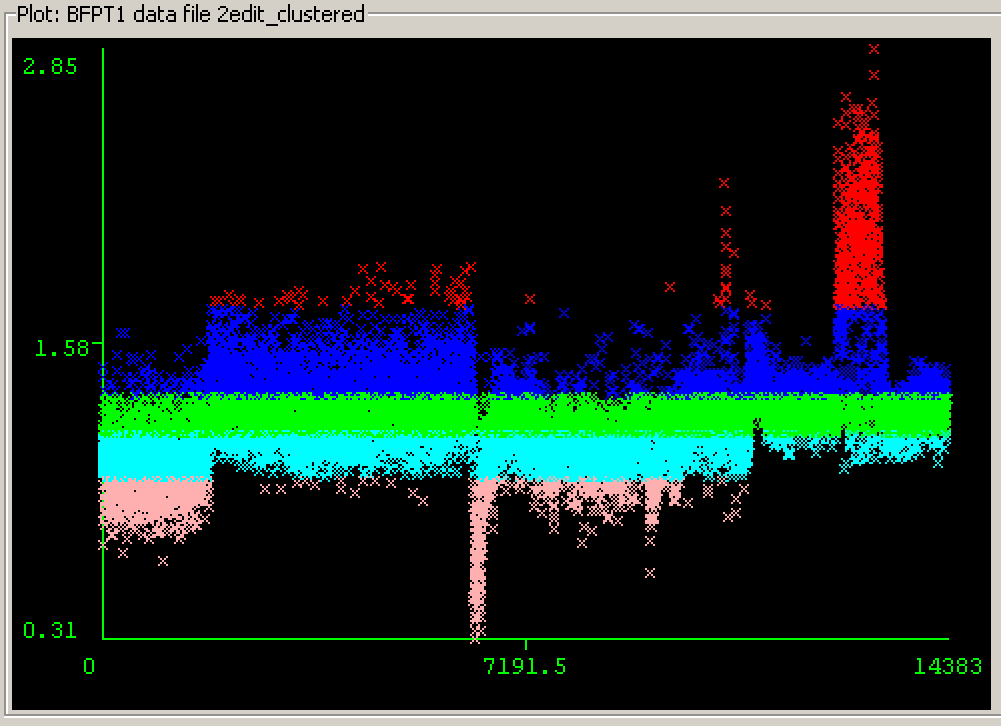


Figure 6.35. Instance number vs. LP/HP ratio divided into five clusters – from file 2

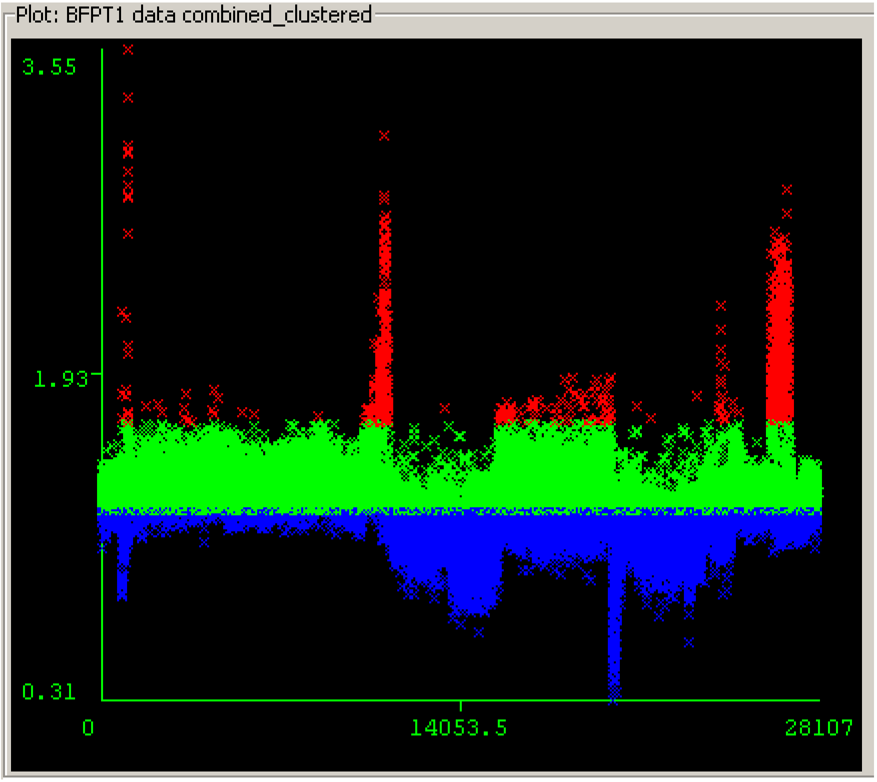


Figure 6.38. Instance number vs. LP/HP ratio divided into three clusters – combined

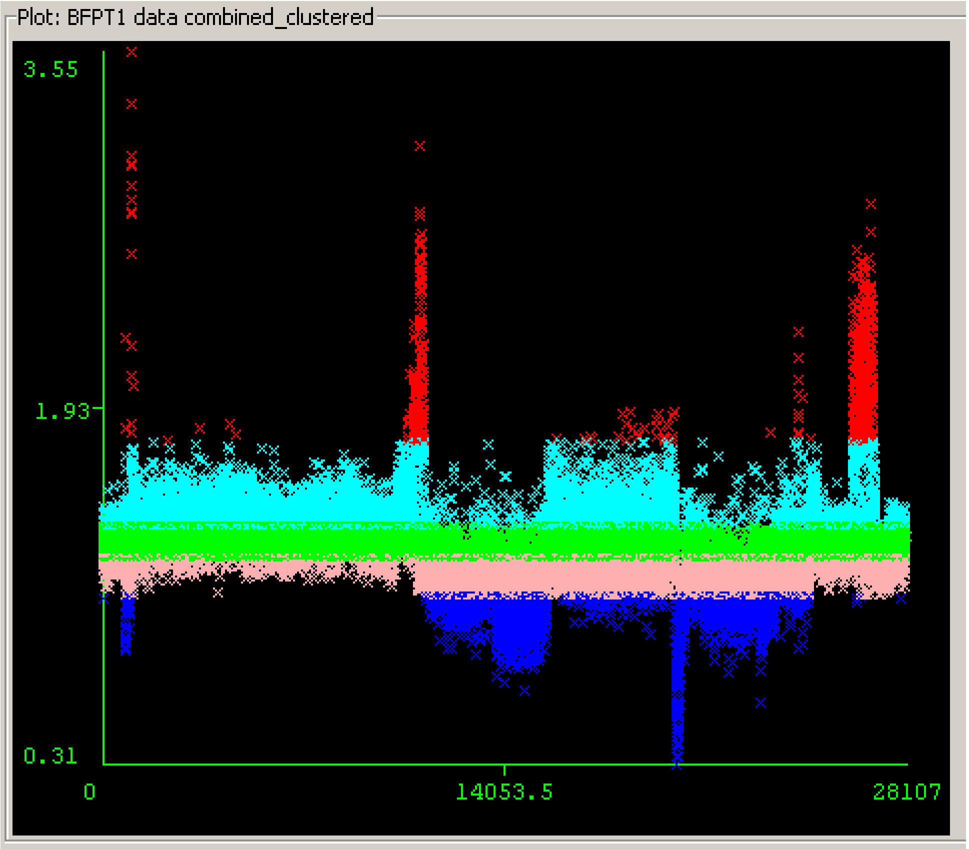


Figure 6.39. Instance number vs. LP/HP ratio divided into five clusters – combined

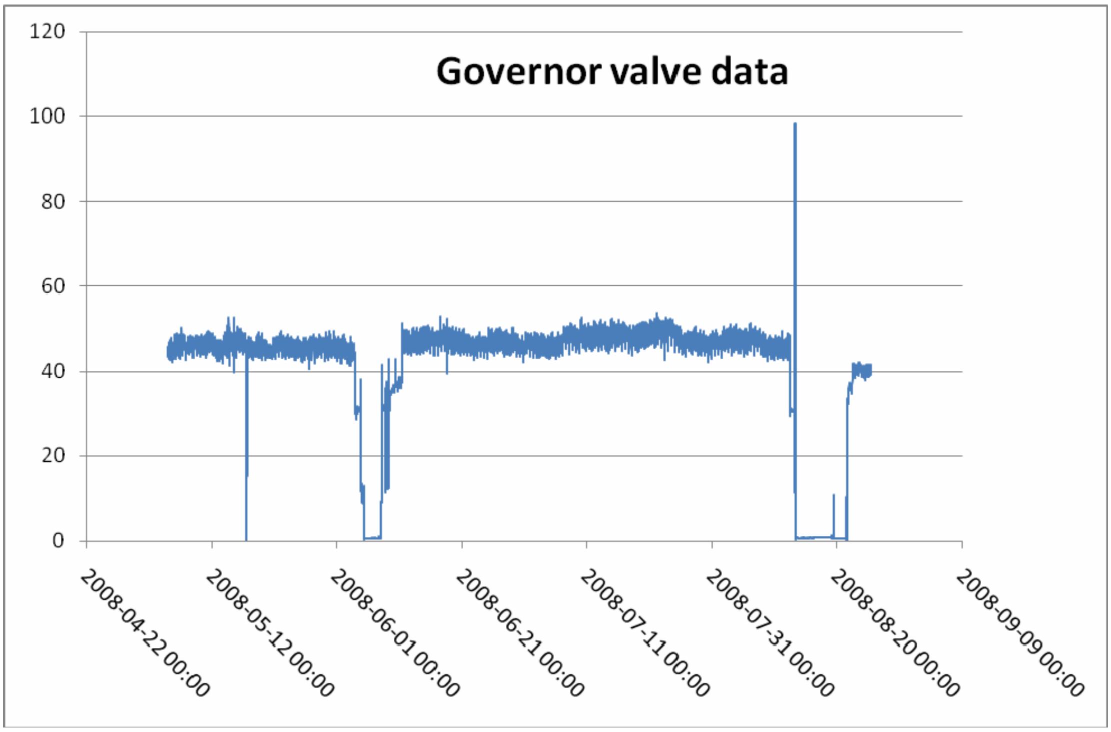


Figure 6.42. Governor valve data from file 1

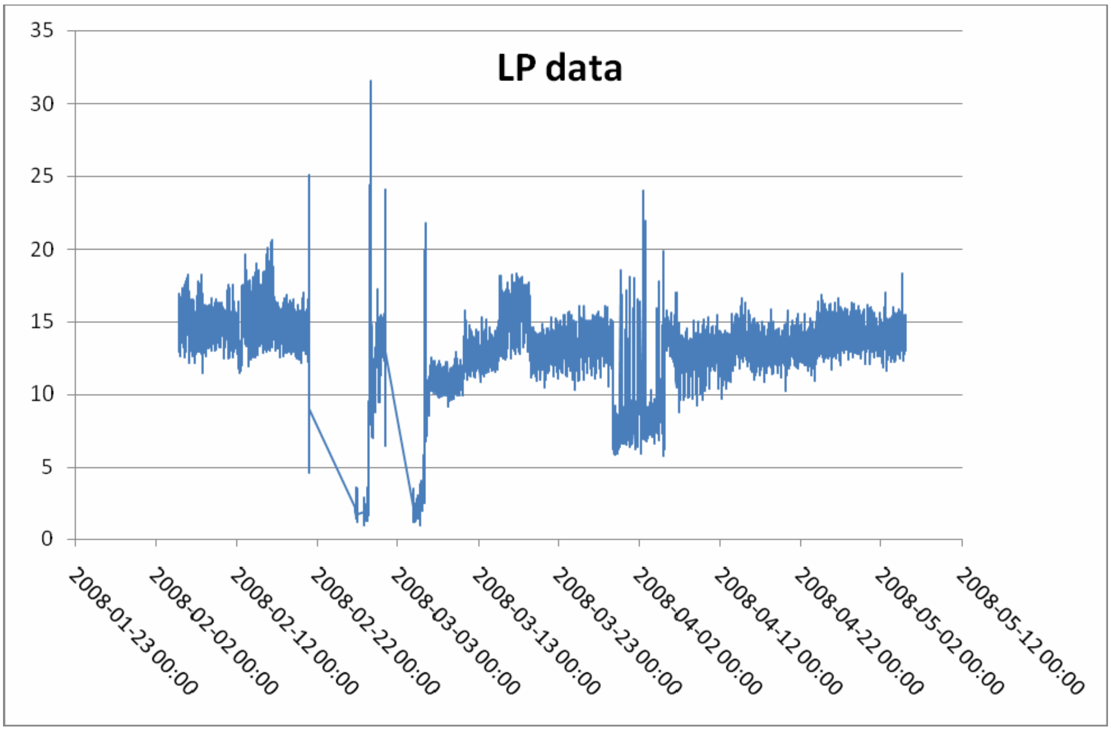


Figure 6.44. LP data from file 2

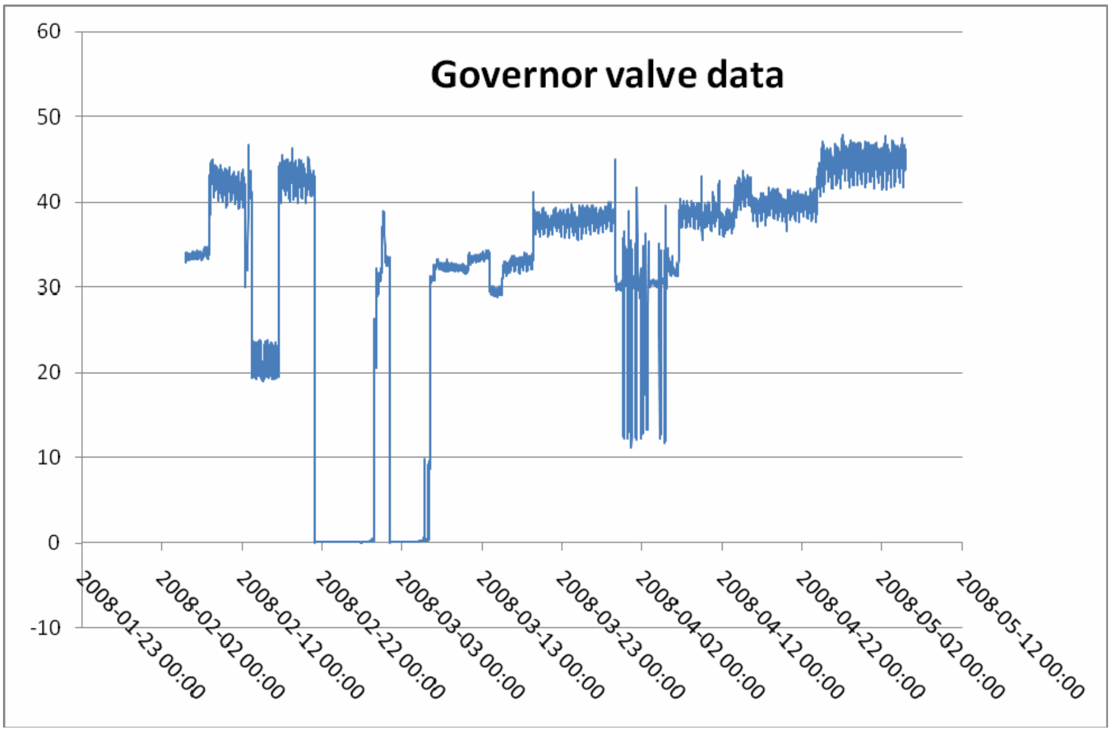


Figure 6.45. Governor valve data from file 2

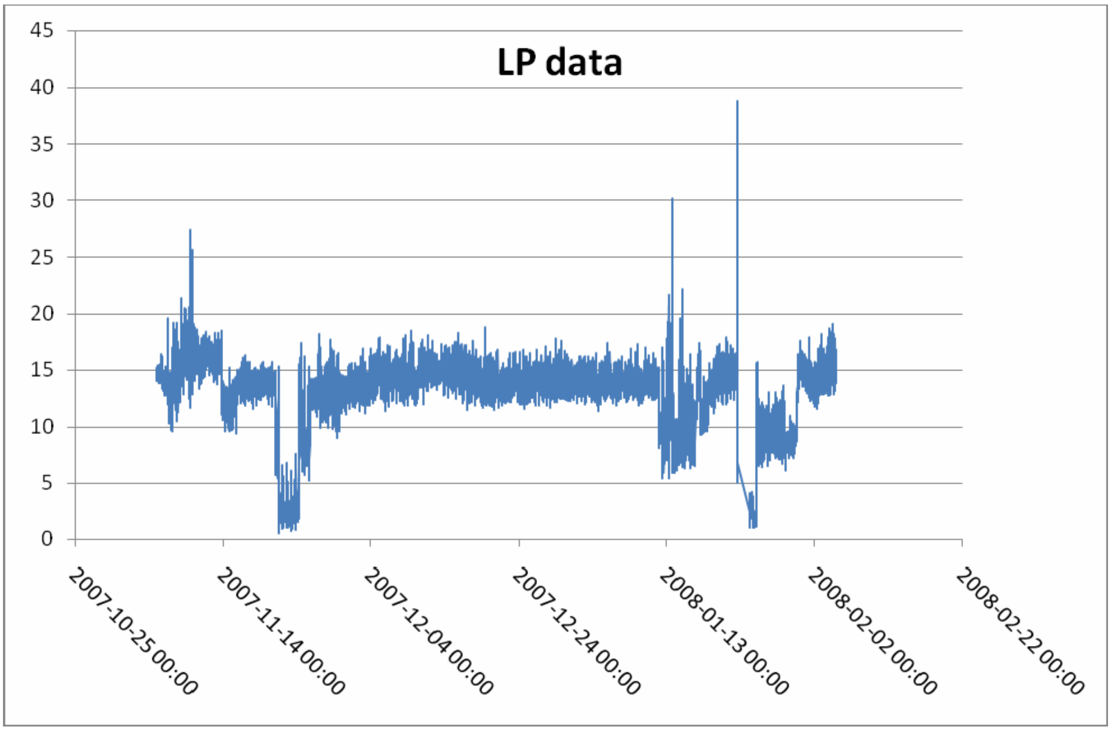


Figure 6.50. LP data from file 3

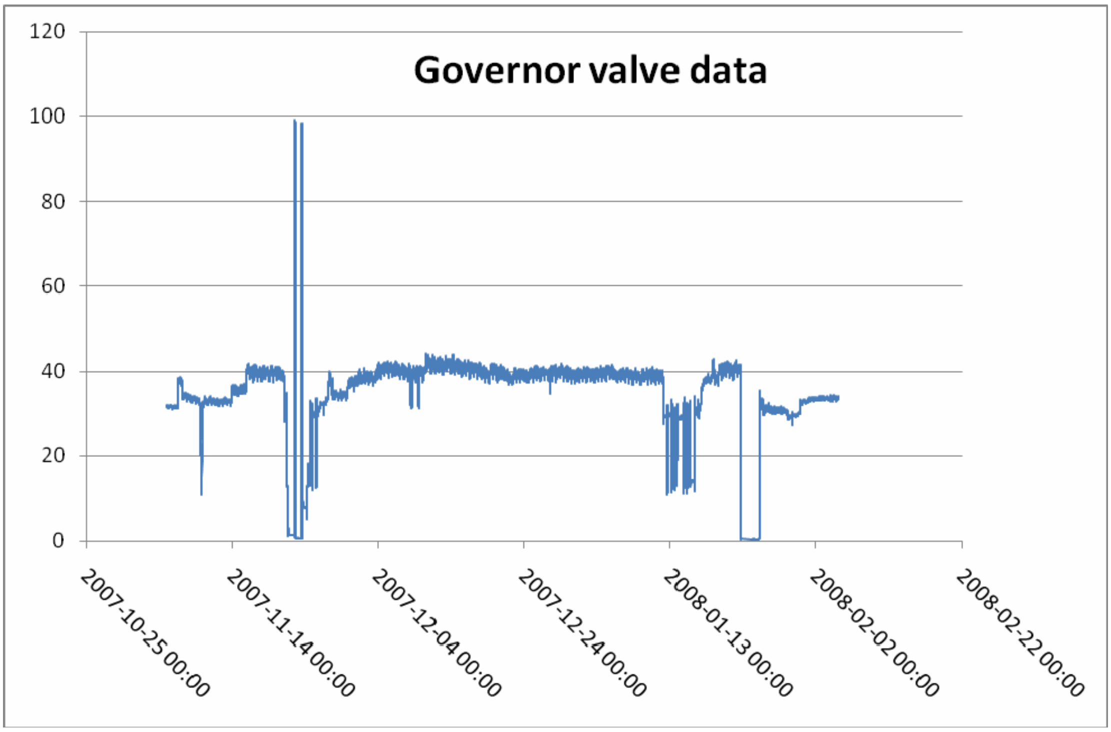


Figure 6.51. Governor valve data from file 3

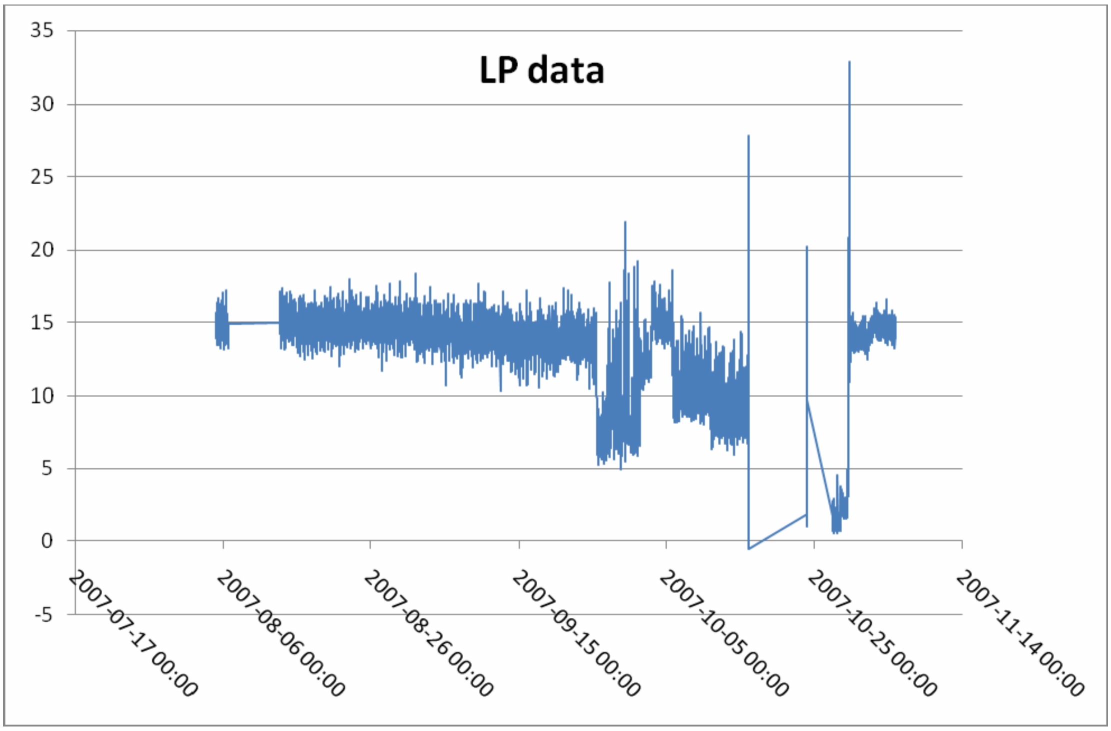


Figure 6.53. LP data from file 4

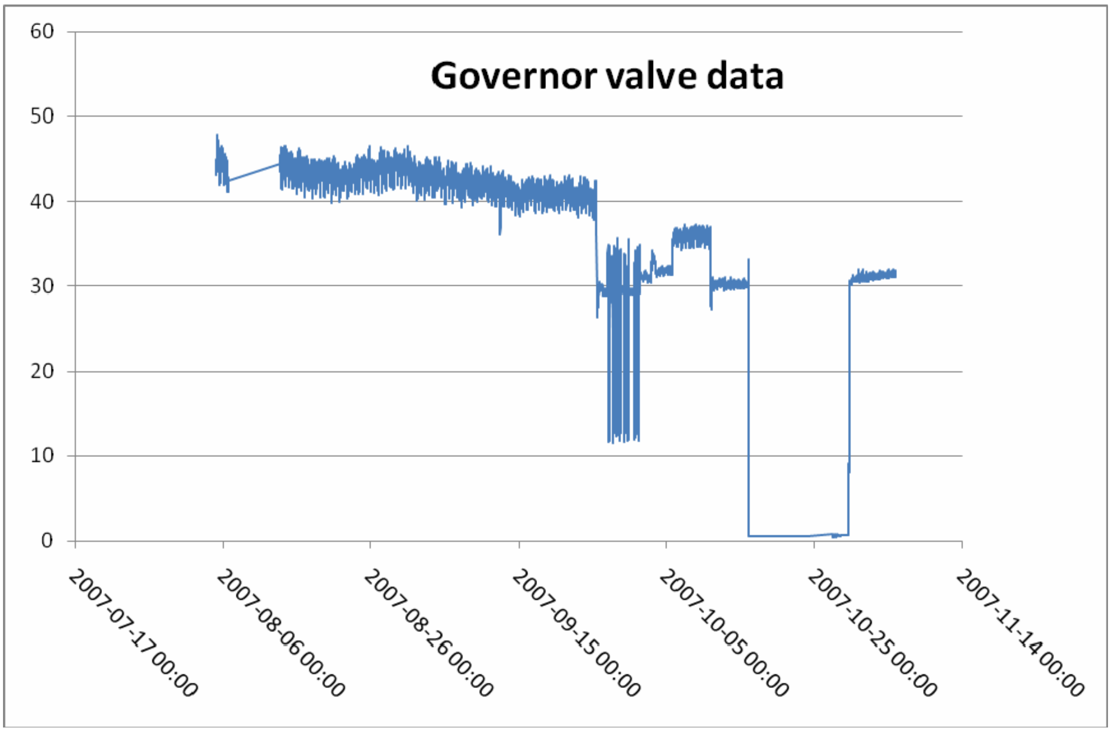


Figure 6.54. Governor valve position data from file 4