7

Success Stories on Real Pilots

**Rafael Socorro[1], María Aguirregabiria[2], Alp Akçay[3],
Michele Albano[4], Mikel Anasagasti[5], Andoitz Aranburu[6],
Mauro Barbieri[7], Iban Barrutia[8], Ansgar Bergmann[9],
Karel De Brabandere[10], Marcel Boosten[11], Rui Casais[12],
David Chico[6], Paolo Ciancarini[13], Paulien Dam[14], Giovanni Di Orio[15],
Karel Eerland[11], Xabier Eguiluz[2], Salvatore Esposito[16],
Catarina Félix[17], Javier Fernandez-Anakabe[8], Hugo Ferreira[17],
Luis Lino Ferreira[4], Attila Frankó[18], Iosu Gabilondo[2], Raquel García[1],
Jeroen Gijsbers[11], Mathias Grädler[19], Csaba Hegedűs[18],
Silvia Hernández[1], Petri Helo[19], Mike Holenderski[3],
Erkki Jantunen[20], Matti Kaija[21], Aleš Kancilija[22],
Félix Larrinaga Barrenechea[8], Pedro Maló[15], Goreti Marreiros[4],
Eva Martínez[1], Diogo Martinho[4], Asif Mohammed[12],
Mikel Mondragon[5], István Moldován[23], Antti Niemelä[24], Jon Olaizola[8],
Gregor Papa[25], Špela Poklukar[25], Isabel Praça[4], Stefano Primi[1],
Verus Pronk[7], Ville Rauhala[24], Mario Riccardi[16], Rafael Rocha[4],
Jon Rodriguez[26], Ricardo Romero[27], Antonio Ruggieri[16],
Oier Sarasua[6], Eduardo Saiz[2], Veli-Pekka Salo[19], Mónica Sánchez[1],
Paolo Sannino[16], Babacar Sarr[10], Alberto Sillitti[28], Carlos Soares[17],
Hans Sprong[11], Daan Terwee[14], Bas Tijsma[14], Tom Tourwé[29],
Nayra Uranga[1], Lauri Välimaa[27], Juha Valtonen[24], Pál Varga[18],
Alejandro Veiga[1], Mikel Viguera[26], Jaap van der Voet[11],
Godfried Webers[11], Achim Woyte[10], Kees Wouters[7], Ekhi Zugasti[8]
and Urko Zurutuza[8]**

[1]Acciona Construcción S.A., Spain
[2]IK4-Ikerlan, Arrasate-Mondragón, Spain
[3]Technische Universiteit Eindhoven, The Netherlands
[4]ISEP, Polytechnic Institute of Porto, Porto, Portugal
[5]Goizper, Spain
[6]FAGOR ARRASATE, Arrasate-Mondragón, Spain
[7]Philips Electronics Nederland B.V., The Netherlands
[8]Mondragon Unibertsitatea, Arrasate-Mondragón, Spain

[9]STILL GmbH, Germany
[10]3e, Belgium
[11]Philips Medical Systems Nederland B.V., The Netherlands
[12]ADIRA Metal-Forming Solutions SA, Portugal
[13]Consorzio Interuniversitario Nazionale per l' Informatica, Italy
[14]Philips Consumer Lifestyle B.V., The Netherlands
[15]FCT-UNL, UNINOVA-CTS, Caparica, Portugal
[16]Ansaldo STS, Italy
[17]Instituto de Engenharia de Sistemas e Computadores do Porto, Portugal
[18]AITIA International Inc., Hungary
[19]Wapice, Finland
[20]VTT Technical Research Centre of Finland Ltd, Finland
[21]Fortum Power and Heat Oy, Finland
[22]XLAB, Ljubljana, Slovenia
[23]Budapest University of Technology and Economics, Hungary
[24]Lapland University of Applied Sciences Ltd, Finland
[25]Jožef Stefan Institute, Slovenia
[26]KONIKER, Arrasate-Mondragón, Spain
[27]Tekniker, Spain
[28]Innopolis University, Russian Federation
[29]Sirris, Belgium

This chapter describes success stories. The MANTIS architecture (Chapter 3) was implemented for a number of use cases on real pilots, and the techniques described in Chapters 4, 5, and 6 were experimented with in real settings. Results on the techniques were already presented in previous chapters. This chapter, on the other hand, describes the pilots, from their objectives and context to the system integration efforts to the attained results. One of the results of the application of the techniques was the enhanced Technology Readiness Level of the techniques, which is summarized in Figure 7.1.

Each section takes care of providing details for a different use case, and as a whole the chapter proves the large breadth of the applicability of the MANTIS approach.

7.1 Shaver Production Plant

Contributors: Bas Tijsma, Paulien Dam and Daan Terwee

The goal of the shaver production plant use case is to increase the predictability of the maintenance actions through smart use of data. By actively utilizing various data sources in an automated manner, it is expected
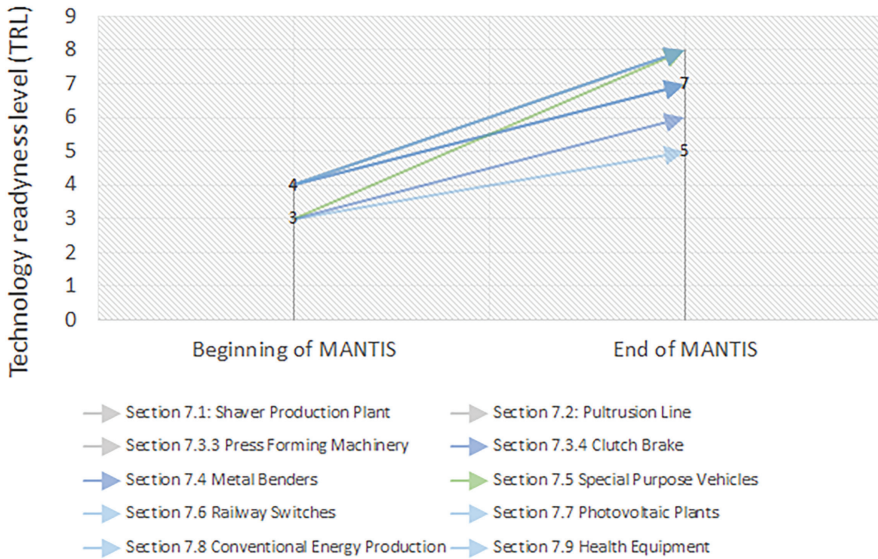
TRL improvements use-cases Mantis

Figure 7.1 Maturity Level before and after the MANTIS project for the techniques in the use cases.

to reduce the cost of the tooling, and reduce the amount of unexpected downtime, thereby reducing the total cost of production. This section provides an overview of the practical application of the several elements developed and applied to this use case. The rationale of the approach is represented in Figure 7.2.

7.1.1 Introduction to the Shaver Manufacturing Plant

The shaver manufacturing plant is one of the largest manufacturing plants of Philips. For mass production of certain parts of the shaver an advanced machining technique is used as manufacturing technology. The tools used in this manufacturing process are the focus area for this use case.

Large amounts of data are gathered in the manufacturing plant about the products and processes. These data is mostly used for manual, after-the-fact analysis of process disruptions, machine failures and quality issues.

It is expected that these data (and where necessary additional data) can be used to make predictions about product quality, process disruptions and impending maintenance actions. By actively utilizing the data in an automated manner, it is expected to reduce the cost of our tooling and
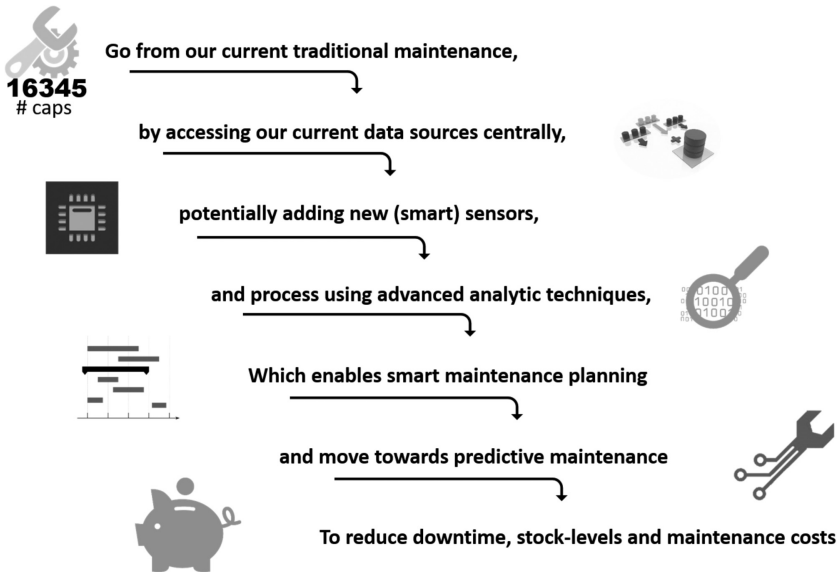
Figure 7.2   Goals and vision for the shaver manufacturing site.

maintenance. The objective is to demonstrate that impending failures can be accurately predicted by mining large amounts of data from heterogeneous databases, such that tooling maintenance can be timely scheduled to prevent unexpected downtime of the production lines and maximize tooling lifetime.

7.1.2 Scope and Logic

For the complete project, a full project scope was made. This chapter focusses on a subset shown in Figure 7.3, and is largely based on practical experience and domain expertise. The main focus, the prediction of tool failure, is put in the center of the picture.

There are three main influencing factors regarding the life-time of a tool:

- Process behavior
  The combined behavior of the process (measured by many sensors, see Section 7.1.3) during the discrete manufacturing processes, as well as process error behavior over time. Both the process sensor measures behaviors, as well as process errors which may cause damage to the tool and influence the state of the tool;
- Quality Status
  It entails the geometrical measurements of the products made by the aforementioned process, which need to comply with product
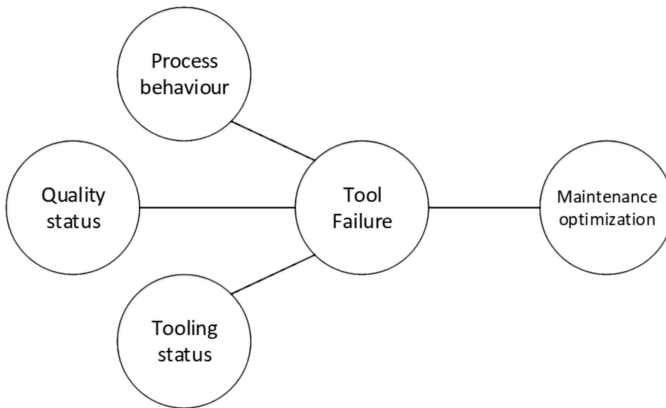
Figure 7.3 Basic interactions.

specifications. Many quality deviations on the products can be related to a change (damage, wear) in the tooling geometry, which might indicate an (upcoming) failure;

- Tooling status

   The current status of the tools with respect to wear, small damage, etc., over time. These cannot be measured by the process sensors and the quality status, and usually imply measurements performed over longer time periods.

The main goal is to use these three data sources in a combined model to predict tool failures.

With the ability to model tool failure behavior, the output can be used as an input to optimize Philips' current maintenance policy and strategy. For example, the amount of spare stock can be regulated much better, if the future failures of tools can be predicted.

7.1.3 Data Platform and Sensors

In this use case, the existing proprietary platform of the manufacturing site is used as much as possible, to be able to focus more on the analysis and application part of the project. Most of the manufacturing machines are connected to a legacy data platform, known as the Factory Information System (FIS). It consists of various relational databases, to which the machines are connected by custom developed drivers. This was custom built over the course of several decades by the internal IT department.

Despite to recent developments in the so-called industry 4.0 revolution, where much progress has been made into generic data exchange protocols

(e.g., OPC-UA) and storage architectures (e.g., Mimosa), the equipment in scope is legacy and cannot be changed easily. During the course of the project, it became clear that performance of the existing platform was sufficient. However, in the future, increasing amounts of data requests might hamper overall performance.

The main data source in the platform is the process data. The machines in scope are equipped with a wide variety of sensors. The output of the sensors is collected by a machine controller to perform several pre-processing steps, like filtering and aggregation, before the data is sent to the FIS platform. In general, a set of data is collected and sent once per production cycle (in one cycle one product is made). Each cycle contains over 100 parameters. The data is externally accessible via the FIS platform.

Tooling information is also automatically stored in a separate database. A digital log is kept on the lifecycle of each individual tool, for example on which machines it was placed, the amount of products the tool has made and maintenance actions taken by the tooling maintenance department.

The last data source is the product quality metrics. Quality data is gathered by taking offline product samples on dedicated measuring devices. These measurements are inputs for the quality system, which is also part of the FIS platform, meaning they can also be accessed externally. All measurements are geometrical, like form accuracy and thickness. Usually these data are aggregated values of a larger set of measurement points, like average, standard deviations, etc.

7.1.4 Data Analytics and Maintenance Optimization

The manufacturing process consists of several physical elements. Electrical, chemical and mechanical elements are working together in order to manufacture the products, making it a highly complex process where interactions between different signals can be easily overlooked when just monitoring every signal individually. A prediction model (soft sensor) that combines all different signals and processing them together gives better insight in these interaction effects via computational intelligence. This sensor fusion deals with disparate sources that do not have to originate from identical sensors.

7.1.4.1 Physical models and background

Before being able to successfully analyze process and manufacturing data, domain knowledge is required, which can be provided by process engineers. Without domain knowledge, it is very hard to understand the data, do the analysis and validate the results.

The machining process in scope is electrochemical machining (ECM). This is an unconventional electrochemical manufacturing process, but it is well established in niche applications like turbine blades and medical implants [McGeough, 1988; De Barr and Oliver, 1975]. This process removes material at the anode (work piece) using current controlled electrochemical process. By feeding a shaped cathode (tool or electrode) towards the work piece, the reverse shape of the tool is copied to the work piece.

It is a complicated process incorporating a number of physical phenomena interacting with each other. Common problems with this process are related to the variations of the material composition, variations in chemical conditions, as well as the influence of the geometry of the tools. All these effects may, in some form or another, change or damage tooling geometry, which in effect will lead to quality issues.

7.1.4.2 Process monitoring with Principal Component Analysis & Hotelling's T^2

The Principal Component Analysis algorithm in combination with the Hotelling's T^2 score is used to get insight in the interaction effect of all different process parameters. To train the model, the data is extracted from the FIS platform and analyzed to make sure that the historic dataset consists of data that indicates only normal process behavior, with no deviations or outliers. This is an important step, since this data will serve as a reference for future predictions.

The PCA algorithm is trained on the historical data where the dimensionality of the entire dataset is reduced, while retaining as much as possible of the variation present in the data set. This is done by transforming the data to a new set of variables called the Principle Components (PCs), which are, by definition, uncorrelated. By definition the PCs are ordered in such a way that the first few PCs contain most of the variation present in the original variables (see Figure 7.4). In this example, the red line indicates that 5 PCs explain more than 90% of the original variance thus reducing the dimensionality from 15 to 5 parameters. The PCA model transforms every observation of the dataset into a set of scores of the same size as the number of PCs.

For real-time calculations the trained PCA model is deployed on a server and data from the PLC-PMAC system is fed into this model in order to obtain the new weighted scores that indicate how close new observations are related to the historic dataset. Because the PCA algorithm is a 'white box' algorithm,
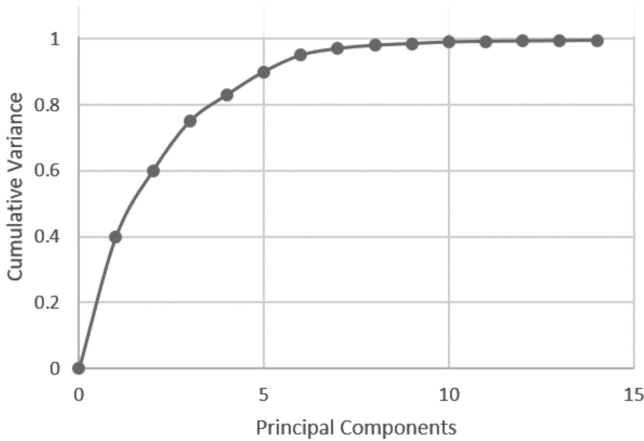
Figure 7.4   Cumulative explained variance of the principal components.

it has self-diagnostics abilities. The model can be used to determine the root causes for fluctuations, trends and outliers.

The resulting PCA model allows for real time streaming data to be transformed onto the principal component space defined during the training phase. Assuming the training dataset is a good representative of normal machine behavior, any significant deviations seen with the live streaming data can be interpreted as signs of a defective machine behavior.

The Hotelling's T^2 value can be used as a measure for how close the transformed live streaming data are to the training set. It is reasonable to start from the simpler univariate case t-test that is defined as follows:

$$t = \frac{\overline{x} - \mu_o}{s/\sqrt{n}} \tag{7.1}$$

where the null hypothesis H_0 (the historic dataset) provides a mean $\mu = \mu_0$ and a sample of n can be acquired from the population with mean $\overline{x}$ and standard deviation s. The statistical interpretation is as follows: on average the difference between the sample and null hypothesis will fall within s if normal statistical variation can explain these differences. The overall t-score is weighted by $\sqrt{n}$ since any differences in the numerator become more significant the larger the sample size becomes.

To generalize this result, it is possible to square the expression for the t-test to obtain:

$$t^2 = \frac{(\overline{x} - \mu_o)^2}{s^2/n} \tag{7.2}$$

$$t^2 = n\,(\overline{x} - \mu_o)\left(\frac{1}{s^2}\right)(\overline{x} - \mu_o) \tag{7.3}$$

This is the same as an F-distributed random variable with 1 and n-1 degrees of freedom. We can replace the difference between the sample mean $\overline{x}$ and μ_0 with the difference between the sample mean vector $\vec{X}$ and the hypothesized mean vector $\vec{\mu_o}$. The inverse of the sample variance is replaced be the inverse of the sample variance-covariance matrix S:

$$T^2 = n\left(\vec{X} - \vec{\mu_o}\right)'\mathbf{S}^{-1}\left(\vec{X} - \vec{\mu_o}\right) \tag{7.4}$$

Using the above expression for T^2 it is possible to compress a multivariate dataset into one scalar metric that can quantify the status of a given manufacturing machine. The use of a single parameter also facilitates machine status visualizations with the monitoring of a single quantity, and the calculated T^2 value can be compared with predetermined confidence limits to trigger warnings or alarms, should human interaction be needed to return the given machine to a nominal status (see Figure 7.5).

7.1.4.3 Product quality prediction with partial least squares regression

As described in Section 7.1.4.1, the product geometry is a negative copy of the tool. Therefore, the product geometry is a suitable metric for detecting deviations in the tool geometry such as damages or wear. In production, the product geometry is measured on a sample basis, once every 4 hours. This
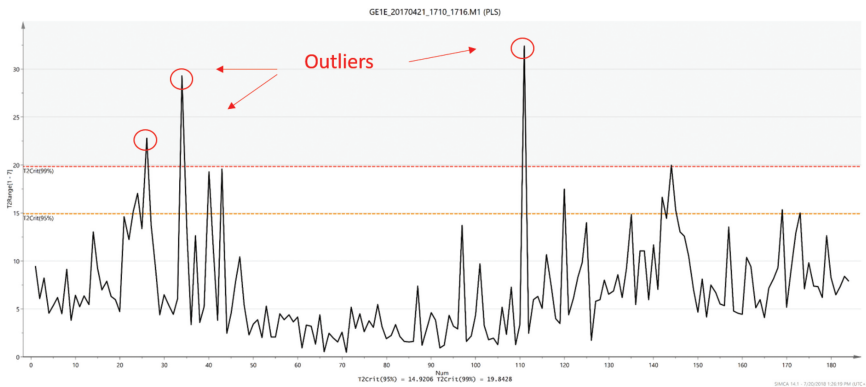

Figure 7.5 Control chart showing the Hotelling's T^2 score and tolerances to identify trends and outliers in the multivariate data.

is not frequent enough for monitoring potential tooling issues. However, by relating the process data as described in the previous section with the product quality data, a predictive model can be trained to estimate the quality of every product made.

For this predictive model, the process data, enriched with a few extra variables of the tooling is used as predictors (X). This process data has already been pre-processed, and consists of an observation for every product made. Each observation, in turn, consists of a collection of data points coming from a variety of different sensors in the production machine. After careful research and bootstrap modelling, 12 predictor variables were selected as input for the modelling algorithm.

As response data (Y), product geometry data is used, which consists of four Y-variables, which are sub selection of all the quality indicators for products made by the manufacturing process. Both data sets have a common product identifier which makes a good join between the two possible data sets. Since product quality data is only measured on a sample basis, the time interval between observations differs greatly from the process data. For this reason, the data set is filtered, where only those observations that consist of both process and quality data are kept.

The last step in the data preparation is outlier removal. The outliers (see Figure 7.5) due to short term sensor failures or miscalculations during ETL are removed from the dataset. Outliers due to physical events in the process itself remain in the dataset, since they can hold valuable information for modelling the relationships between parameters. Finally, the data is ready to serve as an input for model training.

When the number of predictors (X) is too large (10) the more traditional regression methods, such as linear regression, are not adequately effective. Furthermore, in many cases manufacturing and/or sensory data have a correlated nature. This causes the sample covariance matrix to be ill-conditioned, because it becomes almost singular, which is a problem for the more traditional regression methods. This can be solved by using linear projection methods such as Partial Least Squares Regression [Wold, 1975].

Mathematically, there are quite some advantages for PLS compared to traditional regression algorithms. Because PLS is a linear projection method, it decomposes the covariance matrix that settles the singularity problem. This gives PLS the ability to handle multicollinearity among the predictors (X). Furthermore, these linear projection methods have the advantage that they can handle missing data points in the data set. For example, if a particular sensor has a short term failure or certain data points are removed from the

data set during outlier analysis, then the whole observation does not have to be discarded, but can still serve as input to the algorithm. Moreover, PLS has the extra advantage that it can incorporate multiple Y-variables (or responses) in one statistical model. Finally, PLS is suitable for modelling and monitoring larger number of variables simultaneously.

The trained PLS model gives promising results. In Figure 7.6, the results of the PLS predictions are plotted against the actual (observed) results of a specific product quality parameter. We see that there is only a small deviation between the two trend lines; the Root Mean Square Error (RMSE) is 1.86, and the R^2 is 71.2%, which are good results given the acceptable range of the quality parameter and the complexity of the production process itself.

7.1.4.4 Computational trust

Every production line has incidents. These incidents can be related to, for example, machine failures or process errors. Due to the complexity of the manufacturing process, quite many process errors are generated in time. Some of these errors are critical, as they may cause additional damage. Other errors have minor impact. The impact level and the frequency of occurrence are both important factors in calculating the current state of a particular process, but more advanced interactions are only possible, in which a particular order or process errors can be critical.

Therefore, the concept of computational trust is researched, in order to quantify the current 'trust' in the machine being in a 'good' state, or in a 'bad' state. This allows for errors to be specified (impact-level, fall-off level) and to be combined in a specific 'trust in good machine'-metric, as input for the overall tool failure predictions. For example, refer to the particular case presented in Figure 7.7, which reports generated error codes over time. Values under the threshold of 0.6 are considered as a 'bad' state.

Figure 7.6 Graph of the predicted and the observed quality with PLS regression.
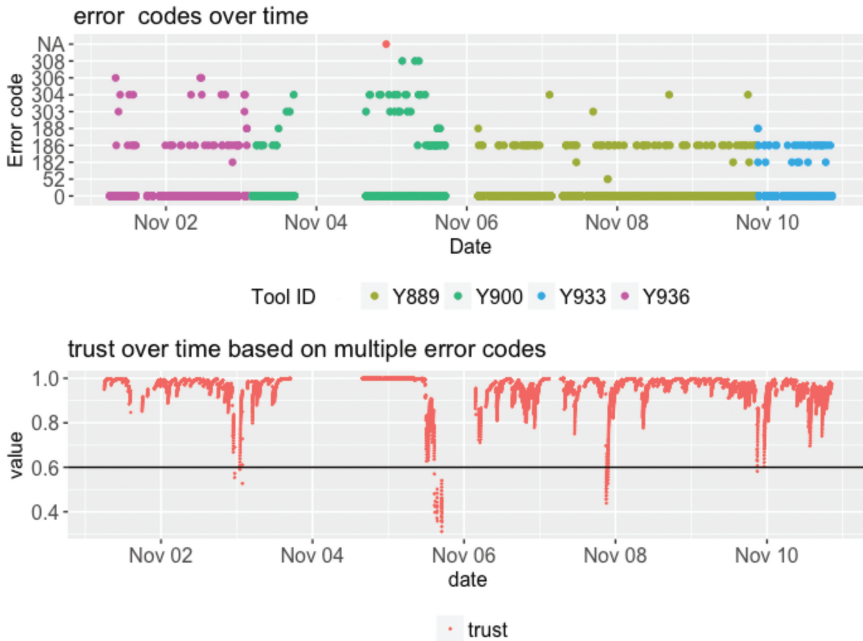
Figure 7.7   Error codes over time, and corresponding trust values.

7.1.5 Visualization and HMI

For visualization of the analytics toolset, a prototype interactive dashboard was developed and tested in production. It is inspired by a principle component analysis (PCA) score plot, but redesigned to be better usable in a production environment.

The goal is to give operators direct insight to the current status of the production process in terms of overall process stability, but also enable more detailed insight by allowing to drill down to the specific cause of potential deviations. This is a major advantage of the methods described in Sections 7.1.4.2 and 7.1.4.3, since it allows to relate aggregated scores (like Hotelling's T^2) with individual sensor values.

The main screen (Figure 7.8) for giving direct insight consists of eight plots, corresponding to eight machines performing the same process in parallel. The individual graphs consist of a scatterplot, where the first and second principal components are plotted. The green ellipse indicates the 95% confidence limit, in which the process can be considered stable. Each point represents an individual product manufactured on that specific machine (the yellow point is the latest product).
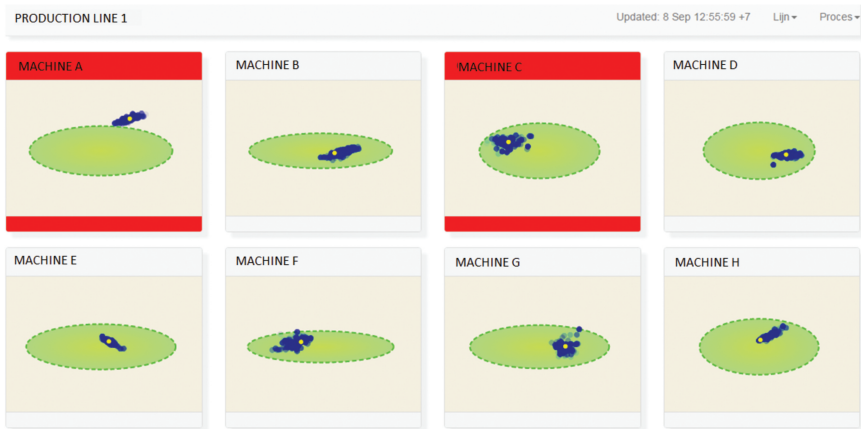
Figure 7.8 Main screen for production operators, showing process performance for eight similar machines. The green circles indicate 'good' behaviour.

When the outcome of the process is too critical, the top bar will change to either orange or red, indicating potential problems. The operators are expected to respond to the alarm, and perform pre-specified actions.

By clicking on one of the eight machines, the dashboard will show more information about this particular machine (see Figure 7.9). This screen provides additional details, such as time-series graphs, to get a
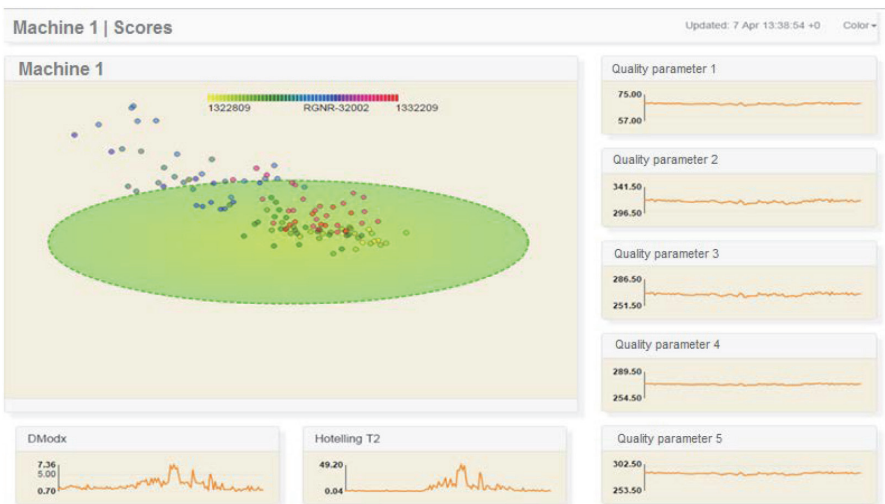

Figure 7.9 Analysis screen of operator dashboard.

better understanding of the problem. The operator (or engineer) can select individual points, can show individual plots of sensor values and can colour the scores based on a selected sensor. These are all experimental tools to help find root-causes and solve production issues much faster.

7.1.6 Maintenance and Inventory Optimization Results

Within the MANTIS project, research has been performed on both the prediction of machine process errors and failures of tools. So far, remaining useful life estimations for the tooling are still uncertain, but moderate results have been achieved with classification methods. Hence, it is investigated what the added value is by using imperfect predictions of tool failures to decide when to perform maintenance actions (e.g., tool inspections and replacements) and when to place orders for new tools. Therefore, it is assumed that a classification model (and its approximate performance) will be the predictive input to the designed policy.

Furthermore, an analysis is carried out into the amount of products that a tool can produce before being discarded. It turns out that the chance that a tool is replaced due to a defect decreases when a tool reaches a longer tool life (see Figure 7.10). In other words, tools that have a longer lifetime are more likely to be replaced due to the age of the tool rather than defects caused by the machine.

For the proposed policy, predictions of upcoming tool failures are generated for every predefined time period (8-hour shift) within a prediction horizon. With a multi-period prediction horizon, the predictions thus overlap. The aggregate of the overlapping predictions is compared to an 'inspection threshold' to decide whether to inspect a tool in the upcoming shift. This threshold is optimized by explicitly modeling the imperfectness in predictions. The predictions are also added as a data-driven component to a (R, s, Q) inventory control policy. The expected amount of tools is added to *not* fail in the coming review period (based on imperfect predictions)

Figure 7.10   Time to failure (in amount of products) for the three main tool types.

to the conventional inventory position (IP). This altered IP is compared to the reorder level *s* at each review period *R* to decide on whether to order *Q* tools.

The predictions thus allow for *postponing* orders to the next review period if enough tools are still expected to be operational by then. Implementing this joint policy for the use case study will lead to an estimated annual savings that consists of inventory savings, maintenance savings and a decrease in tools ordered. There are also one-time savings in tooling purchasing costs from lowering the current average inventory on hand level to the optimal level.

7.1.7 Conclusions

For the shaver production plant use case, quite a lot of effort has been spent on the analytics parts during the MANTIS project. This was mainly possible due to the already existing data acquisition platform, which provided a solid starting point for analytics. Some very promising results have been found for this use case.

It is proven that it is possible to use process data of the machines to make good estimations of product quality. As a result, insights into current quality performance has increased significantly, as well as a reduced reliability on (slow) offline quality measurements of the products. On top of this, it can also be used as an input for estimation on tooling status, since there is a known relationship between the quality (shape) of the product and the shape of the tool.

Another result is that with some creative thinking, concepts from the academic world can be translated to real-world use cases. One such an example is the computational trust-modeling, which looks quite promising to quantify machine performance with respect to error behavior over time. Again, this quantification is important, as it can be used as an input to the tooling status model.

Image recognition techniques applied to the tools has proven to be difficult, especially on complex shapes and glossy surfaces. Several attempts have been made to use image recognition techniques to calculate tool wear, which can also be used as an input for the wear model. None of them, so far, have given any usable results.

Last, we also looked at the promises of predictive maintenance on tooling, while keeping in mind the future applications to other production tooling assets as well. One challenge is the predictive performance of models, which is hard to estimate when the real output of the models is yet unknown.

Data driven assumptions and simulation modeling can still provide good insights in potential (financial) benefits. From a business point of view this is important, as businesses typically demand clear business cases in order to provide resources for further enhancements of these concepts.

7.2 Deploying an User Friendly Monitoring System for Pultrusion Line Production

Contributors: Rafael Socorro, Raquel García, Nayra Uranga, Silvia Hernández, Mónica Sánchez, Alejandro Veiga, Eva Martínez, Stefano Primi

This use case aims to design and develop a reliable monitoring system locating different sensors in key locations for gathering relevant data to improve the preventive maintenance for different processes included in the pultrusion process. Moreover, it allows to create an historical storage of all the data collected to identify patterns in the future, and contribute to better proactive maintenance.

7.2.1 Introduction to the Pultrusion Use Case

ACCIONA operates one manufacturing plant (with two production lines) in Alcobendas (Madrid) for production of composite structures through a pultrusion process. This process has been widely used for manufacturing highly strengthened and continuous composite structures with low weight, elevated mechanical and chemical resistance, and electrical and thermal insulation. For example, they were used for the Pajares tunnels in Asturias (Figure 7.11), for Valencia Lighthouse (Figure 7.12), and for the pedestrian bridge in Madrid (Figure 7.13). The properties of the composite structures are the main reason why this method has become essential in the development of ACCIONA's highly differentiated construction projects.

This process is very challenging in terms of production and maintenance of the equipment involved as it is a continuous process and the machines are running 24 hours a day, so it is necessary to avoid production stops or unexpected delays.

7.2.2 Scope and Logic

The production line, shown in Figure 7.14, is continuous, it stops only when the part model being produced changes. A new product to be produced entails a new configuration of the machine. The current maintenance policy

Figure 7.11 Waterproofing the Pajares tunnel in Asturias, Spain.

Figure 7.12 Valencia lightouse (Spain).

Figure 7.13   Pedestrian bridge in Madrid (Spain).

Figure 7.14   Pultrusion machine at ACCIONA workshop.

for the pultrusion line is preventive (some tasks are performed every time unit periodically), reactive corrective (if a failure is detected) and opportunistic (if the line is stopped and the deadline for the maintenance task is close), which means that once the production line stops to perform maintenance tasks, units can be replaced because they were detected as defective but also because it is an opportunity to change it because the line is stopped. Ideally the line will not stop until the type of product being produced changes.

There exist three roles involved in the production site of pultrusion line from ACCIONA; production manager, process engineer and operator. Each of them has the following responsibilities and objectives in the framework of the pultrusion process:

- *Production Manager responsibilities:*
  Study how to improve the overall process for achieve a higher level of efficiency, adapt the process to new kinds of work requests, transmit the outputs obtained from the aforementioned evaluations to the process engineer, and along with the process engineer make appropriate decisions about how to solve deviations caused by repetitive failures;

- *Process Engineer responsibilities:*
  General maintenance of the machine, designing the incoming updates from the production manager, deploying the aforementioned updates and test them, leave the equipment ready for use, develop technical instruction for the operators to explain how to proceed with the process, give training sessions to the operators involved in the process, decide how to proceed when a deviation occurs;

- *Operator responsibilities*:
  Daily use of the machine, acquire a high level of autonomy to perform the work requests without supervision, in case of any nonconformity in the process the operator must report the problem to the process engineer.

Focusing on maintenance tasks, the process engineer is the person who makes important decisions related to several aspects, such as machine maintenance or in the case of any deviation occurs. He is responsible for the equipment to be ready for use and this requires to develop technical instruction for the operators explaining how to proceed with the maintenance tasks. Furthermore, he is in charge of giving training sessions to the operators involved in the maintenance process oriented to apply knowledge in a practical way from the technical instruction.

Maintenance tasks within pultrusion line are manual processes based on visual checks or manual tasks scheduled from time to time. For this reason, these processes must be optimized in order to achieve a reliable analysis of maintenance tasks, foresee potential failures in the systems, decrease production delays and assure proper machine functioning.

7.2.3 Data Platform and Sensors

The maintenance tasks related to this machine do not only involve the machine itself. It is necessary to monitor the workshop's environmental conditions.

Data to be collected from the workshop.

- *Environmental parameters:*
  Temperature, humidity and luminosity. These parameters concern to possible change of machine configuration, which can affect the maintenance tasks. The obtained information through temperature and humidity sensors will be an indication of workshop status with a direct implication on the workshop maintenance tasks. For example, the resin used during the production process becomes solid at different points depending on temperature and humidity. The proper mixture of the components should be carried out in a controlled environment that influences on the reaction rate and the proper maintenance of these substances to ensure good criteria of quality. In turn, environmental conditions affect the cleanup and purge tasks of the machine;
- *Workshop air extraction capacity:*
  The maintenance of the ventilation system can vary significantly depending on its use and the outside environment. It is an important system to be taking into account due to the kind of substances that are used during the pultrusion process (i.e., carbon fibre, resin...) and the need to offer the best conditions in the workshop;
- *Workshop electrical consumption:*
  Currently there is no information about machine downtimes produced by electrical failures. Electrical consumption monitoring will help to provide a reliable maintenance of the machine, engines or any other electrical installation in order to find the possible causes of downtimes or malfunctions.

Data to be collected from the pultrusion machine:

- *Pull-Clamp system*:
  It is one of the main pultrusion line subsystems responsible for moving the profile along the machine and its subsystems while the different treatments to produce composite profiles are performed. The data that is missing regarding this subsystem is the one related to production speed, pull force and presses oil status:
  - *Production speed* in order to detect machine behavioural patterns and anomalies;
  - *Pull force* in order to know the appropriate amount of oil for lubricating this subsystem;
  - *Oil tank* presses system. Generally, oil contamination is one of the major causes of hydraulic system and lubricating system failures. The oil inside the hydraulic system is changed from time to time by

the operators, based on the process engineer instructions through visual check. Therefore, the continuous supervision of oil state plays a vital role in predictive maintenance systems. The status of oil humidity and temperature are parameters that influence on the quality and profitability of process;

- *Injection Chamber Resin system*:
  It is the responsible for impregnating the fiber with resin. The profile quality depends on the homogeneity of the spray pattern within this process. There are several parameters that will help to analyze an adequate maintenance of this system, such as:

  - *Injection System Temperature*: To determine possible malfunctions that can impact on the proper resin status and manufacture profile quality;
  - *Thermocouples break detection,* in order to avoid malfunctions and errors in the production line;
  - *Resin Header* (resin pressure or stream flow rate): It was an unreliable system without any sensor installed. Measuring pressure or resin stream flow rate proved to help to identify when drain maneuvers are needed to avoid breakdown/malfunction of resin injection system;

- *Compressed Air:*
  Monitoring pressure, humidity and temperature for ensure a proper maintenance of this system. Continuous monitoring reduces ongoing operation costs, cuts investment costs for new compressors and ensures availability around-the-clock. Compressed air subsystem is one of the most expensive systems in production plants. Many companies are not aware of the fact that the generation and treatment of compressed air accounts for up to 20% of their overall energy costs;

  Measurement of compressed air maintenance activities is the first important step towards a cost-conscious and efficient approach to energy consumption and to increase the life-time of the system. Detailed knowledge of the actual compressed air is the basis for reducing energy costs and is an important indicator for investment decisions;

  Dew point, pressure, temperature and flow monitoring makes a significant contribution to quality assurance in expensive systems and the products produced there. Only sufficiently dry compressed air can

reduce the risk of corrosion, machine failures and low-quality end products;

The correct maintenance of this system will assure:

- To ensure efficiency

    Permanently record, monitor and optimize the effectiveness and efficiency of compressed air generation and treatment processes;

- To assure product quality

    A change in consumption of compressed air in a production plant is a first indication of possible deviations in the production process. Sufficiently dry compressed air assures the quality of the system and the pultrusion pieces produced;

- For accounting

    Billing individual costs for compressed air according to actual consumption can contribute significantly to enhancing a cost-conscious system, and it can suggest whether the compressed air is dry enough and can thereby avoid unnecessary operating costs for compressed air treatment;

- To detect leaks

    25–40% of the compressed air generated is lost through leaks. Consumption of compressed air in a system that is switched off is a clear indication that there is a leak.

After some analysis and lab tests, the most suitable sensors were selected for gathering the data.

The main drivers and constrains that were considered are:

- An efficient sensor installation process was needed, as the pultrusion line is continuously working with a few limited stops per months. Moreover, these stops usually are very short;
- Wires installation has been limited, both for communication and for powering the devices;
- The required maintenance for the monitoring system should be minimum;
- The ambient conditions inside the workshop are far from a friendly environment, so all devices installed need to be protected.

An overview of the architecture deployed in the use case demonstrator is given in Figure 7.15, and the list of installed sensors is given in Table 7.1. The figure represents all equipment, communications and software services needed to transfer the information provided by sensors to the specific point where the HMI is going to show all the information to the user.

The system implemented can be divided in different parts or subsystems:

- Sensors;
- Local Sensors Controller
    - Zigbee Root Node
    - X86 Gateway
    - LTE Router
    - Wi-Fi Access Point
- Local HMI
- Cloud Servers
    - OpenMQ Server
    - Database Server
    - Web Server
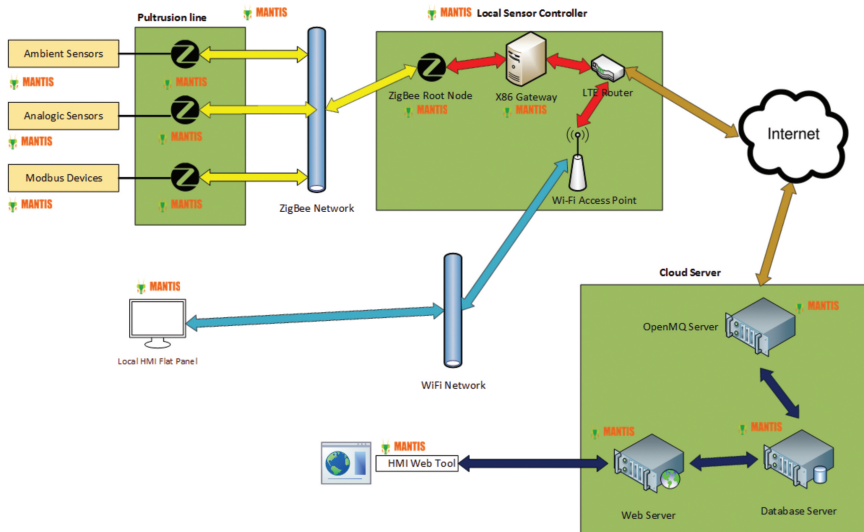
Figure 7.15 Monitoring system.

Table 7.1   List of installed sensors

Sensor	Measure	Unit	Range	Connection with Wireless Platform	Icon
Environmental parameters: Temperature, Humidity, Luminosity	Temperature, humidity, luminosity	°C, %, lux	-50 - 80, 0 - 100, 0 - 1000	Wireless ZigBee	
Workshop and Pultrusion Machines Electrical Consumption – Energy meter sensor	Energy consumption and electrical parameters of the installation.	V, Hz, A,φ	0-200A	Wireless-Modbus RTU Protocol	
Air flow sensor	Air flow	m/s, l/min, m3/h (fps, gpm, cfm)	Measuring range [m/s] 2...100 Setting range [m/s] 0...200	Wireless-4-20mA output	
Air pressure sensor	Air pressure	bar	0-35 bar	Wireless-Modbus RTU Protocol	
Air temperature and humidity Sensor	Air temperature and humidity	°C, %	Relative humidity: 0 %HR ... 100 % HR Temperature: 0 °C ... 85 °C	Wireless-Modbus RTU Protocol	
Oil quality sensor	Percentage of fine particles/ Percentage of coarse particles/	μm %	4,6,14,21μm %	Wireless-4-20mA output	
Oil Temperature sensor	Oil temperature	°C	0 - 200	Wireless-4-20mA output	

Table 7.1 Continued

Water temperature sensor	Water temperature	°C	Measuring range [°C] -20...90 Resolution [°C] 0.2	Wireless-4-20mA and frequency output	
Water flow sensor: Up Circuit/ Flow, Temperature: Down Circuit	Water flow	m/s, l/min, m3/h (fps, gpm, cfm)	Measuring range [m/s] 0.05...3Setting range [m/s] 0...6	Wireless-4-20mA and frequency output	
Coolant reservoir level	Level measurement	°C	-40°C to +125°C.	Wireless-0-5V Analogue output	
Resistors Consumption/ Rupture of wire	Consumption/ Rupture	A	0˜50mA	Wireless-4-20mA output	
Impregnation chamber temperature/ Surface resistors temperature	Surface temperature	°C	0-150	Wireless	

7.2.4 Human Machine Interfaces

Given the different roles of the professionals who use this system, there have been developed and deployed two different HMIs: Local HMI and Remote HMI.

Local HMI:

Our use case has particularities due to 24h/day production in the pultrusion line. In general, local HMI is focused on providing useful information for the machine operators and process engineers, displaying instant parameter values and alerts detected by the sensors (Figure 7.16). The display is located near the pultrusion line machine, in a very visible spot for the workers. Whenever an alert occurs in the local HMI (due to data out of range) the operators and process engineers who are near the local HMI are able to see the alert and they will act in order to face the problem when possible. The operators, following process engineer directions, get the job done and in case of any

Figure 7.16   Local HMI. PC, Tablets, mobile devices.

nonconformity in the process the operator must report the problem to the process engineer.

All the information and alarms should be shown without any interaction. There are several priorities of alert level composed by warnings and alarms.

Remote HMI:

This HMI, on the other hand, is focused on displaying historical information and high or low level alerts detected by the sensors devices (data out of range, the latter will be the same alert as in local HMI). Historical information and its representation using graphs can help the process engineer or production manager to anticipate possible failure and plan maintenance tasks.

The objective of this HMI, represented in Figures 7.17 and 7.18, is to provide a powerful tool not only for data visualisation but also for showing any analysis that the production manager (or the maintenance staff) would require.

Some of the features include:

- User management;
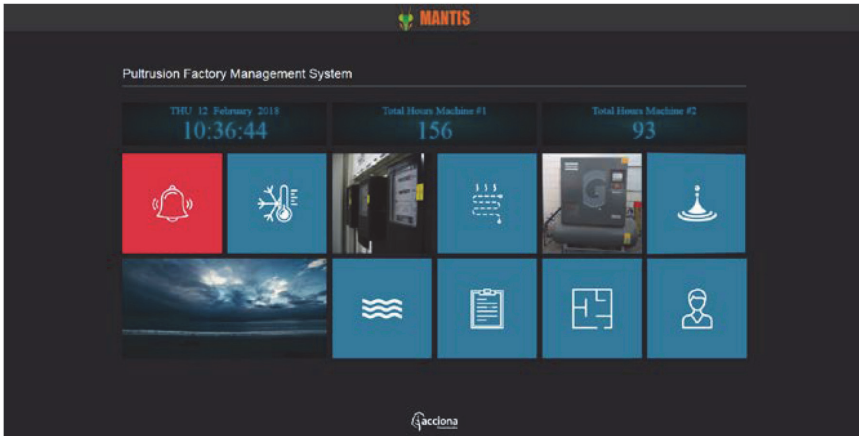- Notifications configuration;

Figure 7.17 Remote HMI. Home menu.

- Alarms configuration;
- Report generation;
- Data graphs.

The HMI is able to display alerts if values are out of range (warning or alarms) or if possible anomalies or failures are detected from historical data. Historical information includes input of time-stamped data. The detection of anomalies could have future implications in the machine with the resulting risk of failures. The system involved in the maintenance task would be identified, together with what is wrong according to the sensing devices.

7.2.5 Maintenance Optimization and Validation Results

This section focuses on the results achieved on the pilots by the MANTIS techniques and monitoring process.

7.2.5.1 Temperature control system located in the mixing area and in the storage area

It is required to keep the storage and mixing temperature of some products used during the pultrusion process lower than a specific value due to safety and quality issues. Consequently, the temperature of the storage and mixing area (Figure 7.19) must be controlled, in this case using an air cooling system.

Figure 7.18   Remote HMI. Temperature, Humidity & Light (THL) menu.

The environmental parameters are gathered through the monitoring system. The operator checks out on the local display warnings and alarms in case that abnormal values of temperature, flow, or air pressure were detected. The Process Engineer checks out the system performance through the Remote HMI according to the Control Program.

In the case that a temperature alarm has been triggered and if the maximum temperature allowed is reached (safe temperature is recommended in the safety-sheet of the stored products), explosive products should be moved to an alternative secure area before carrying out the reparation of

Figure 7.19   Mixing area at workshop.

the cooling system and the safety manager should be reported. If failure was not critical, the problem could be solved during a scheduled stop of the line according to the maintenance program.

Figures 7.20 and 7.21 show some values of temperature and relative humidity of the mixing zone during a continuous parameter registration between February and March 2018. The plots illustrate periods where the machine is not working, achieving temperatures between 15 to 23 degrees Celsius, as well as temperature ranges where the pultrusion line is working during production, achieving temperatures between 22 to 24 degrees Celsius. Relative humidity is always below 50%.

No warning and alarms (yellow and red color alarms) were detected by the operator in the local and remote user interfaces according to the obtained data. The temperature and humidity were kept within the optimal parameters in the mixing area during the pultrusion process tests.

7.2.5.2 Cooling system for the injection chamber

The cooling system (Figure 7.22) maintains the temperature of the injection chamber low enough to avoid premature gelification of the resin inside the injection chamber. The temperature of the chamber is controlled through a liquid chiller. The cooling fluid must be water, optionally mixed with a certain percentage of ethylene glycol (to prevent freezing), depending on the water outlet temperature. The water is cooled using a refrigeration circuit.
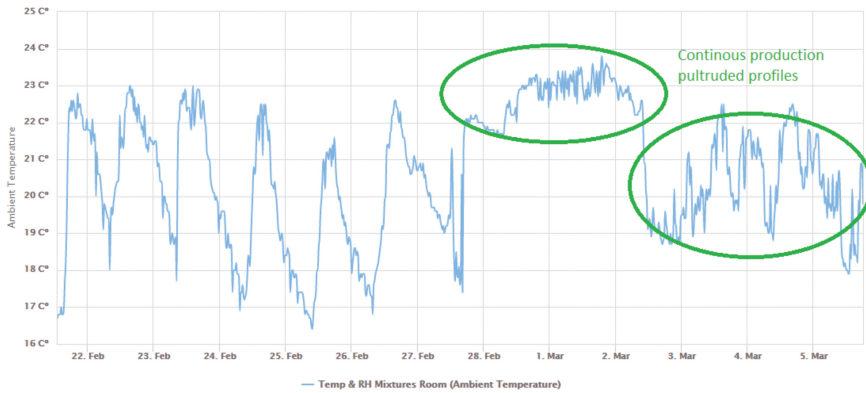
Figure 7.20   Temperature control in the Mixing Area. (Monitoring period 21/02/2018 -06/03/2018).
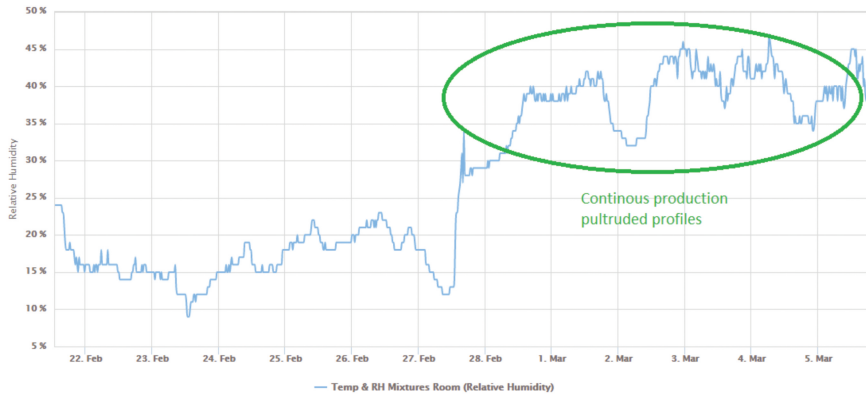

Figure 7.21   Humidity control in the Mixing Area. (Monitoring period 21/02/2018 - 06/03/2018).

Figure 7.23 shows the elementary scheme of the circuit that allows a refrigerating cycle. Figure 7.24 shows a schema of the pipeline circuit used to cool the injection chamber.

The cooling system is monitored detecting the liquid temperature and flow as well as its level in the chiller's deposit, using wireless sensors located in the liquid tank (level) and at the pipelines that connect the chiller with the refrigeration circuit of the injection chamber (temperature and flow).

The operator checks the possible warnings and alarms on the local display in the case that abnormal values of temperature, level or flow of refrigerant liquid were detected. The process engineer views the system performance

Figure 7.22   Cooling system from injection chamber.

through the Remote HMI according to the Control Program. The specialized operator, who is allowed to manage the machine in emergency cases, will be responsible for checking the local display.

If failures were not critical, the problems would be solved during a scheduled stop of the line according to the maintenance program or keeping the machine running (for example, when the chiller's deposit needs to be filled or if a leak occurs). Repair on the fly is allowed in this scenario because the chiller is not physically integrated in the pultrusion machine so it is easier and safer to solve small complications. If it was critical, a non-scheduled stop would be performed and the process manager would be reported. Then, both the process engineer and the process manager would analyse the data available and decide if it is necessary to modify the maintenance program or instead, if it is an unforeseen failure and it is preferable to contact the specialised technical service.

The limits of the temperatures were defined by means of previous manufacturing experience and taking into account that the set temperature of the chiller must be different than the ones we obtain inside the manufacturing site. During the winter the chiller set temperature is between 22–24°C and during summer it is adjusted to 16–18°C to obtain the desired process parameters. It is expected that after a complete year of monitoring a more accurate relationship between these two parameters can be obtained. The installed sensors can be seen in Figure 7.25.
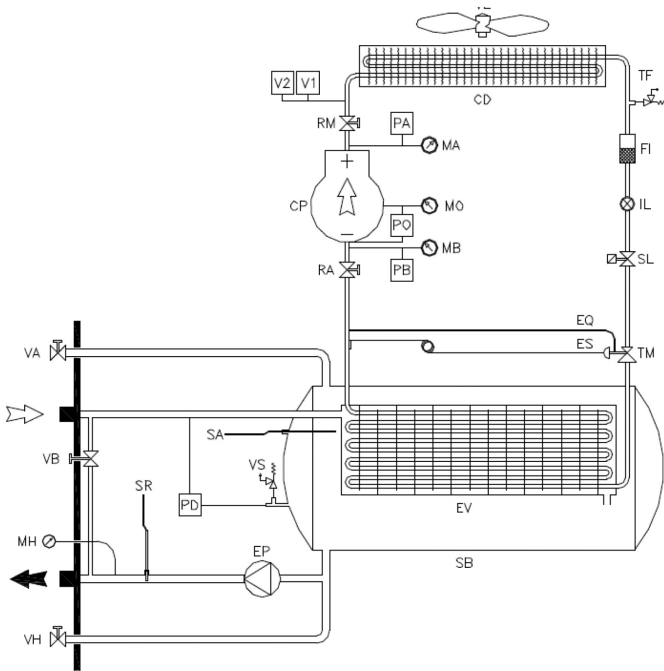
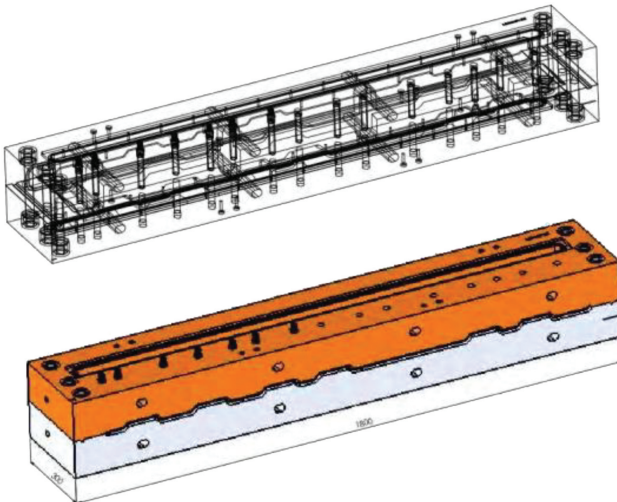**Figure 7.23**   Schema of the chiller cooling system from injection chamber.

Figure 7.24   Schema of the injection chambers cooling circuit.

Figure 7.25   Sensors installed in the refrigeration circuit.

Time series data of the liquid temperature have been recorded for 3 months with the sensors installed as part of the Mantis project. Due to some problems between the sensors and the data acquisition program it has not been possible to see the flow measurements, but it can be predicted that some of the variations seen in the temperature could be directly related with it.

The graphical display for room temperatures and the inlet circuit temperature of the refrigeration liquid is shown in Figure 7.26. It can be clearly seen that although the exterior temperature suffers large variations than the temperature in the refrigeration circuit. An important point to check is that the refrigeration circuit operates correctly and that the external factors are not directly affecting the manufacturing process.

Taking into account upper and lower limits for the alarms defined, it can be seen that some points must be checked although none of them are in the red alarm range. In the first part of the control chart the variation of the temperature is larger than in the last stage. The increase of the temperature (Early January 2018) is due to problems derived from non-constant flows, the water stagnates in the pipes and the temperature increases. On the other hand, low temperatures (middle January 2018) are due to some maintenance stops for the installation of new sensors and connections. As the room temperature is low the refrigeration system decreases its temperature.

In the chart in Figure 7.27, which corresponds to the production of the pultruded profiles, the temperature of the refrigeration circuit is more stable. It can be seen that some points are in the rage of the yellow alarm. As can be seen by comparing Figures 7.27 and 7.28, these increases in the temperature occur at the same time that the level of the liquid in the refrigeration circuit decrease.
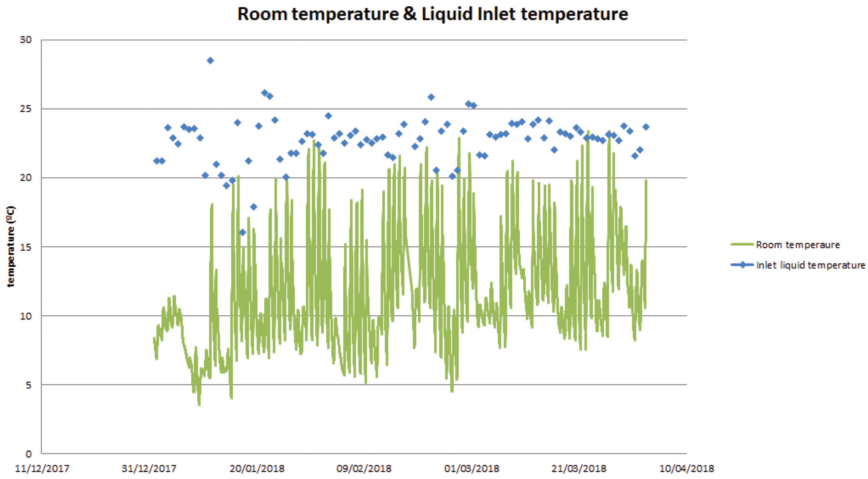
Figure 7.26 Room temperature vs. liquid inlet temperature control chart.
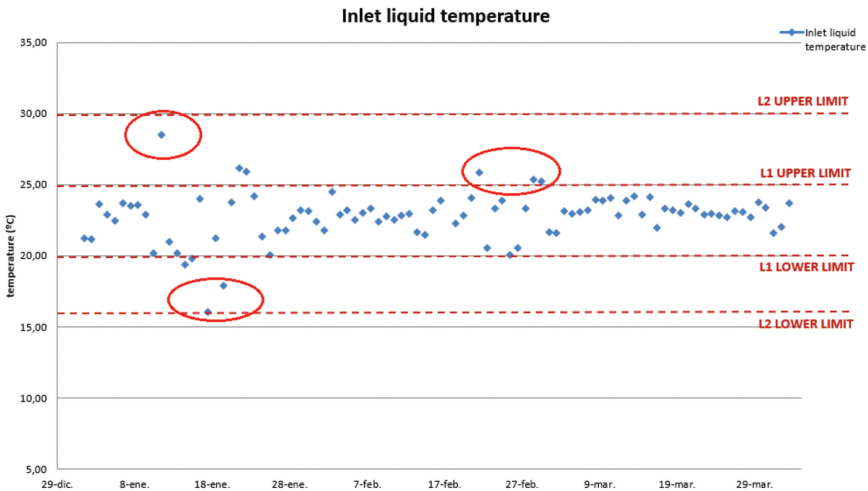
Figure 7.27 Refrigeration circuit inlet temperature control chart.

Figure 7.28 shows the evolution of the refrigeration liquid tank level. The consumption of the tank depends on different factors such as exterior temperature, working hours, number of equipment connected to the circuit, etc. so it is not easy to predict a constant behavior. But it has been seen that a reduction of more than a 35% of the level of the tank has a direct impact in the refrigeration circuit temperature, so it has been decided to refill the tank

Figure 7.28   Refrigeration liquid tank level control chart.

once a day above the 80%. The operator must be in charge of this action and also must be aware of any other alarm that can appear.

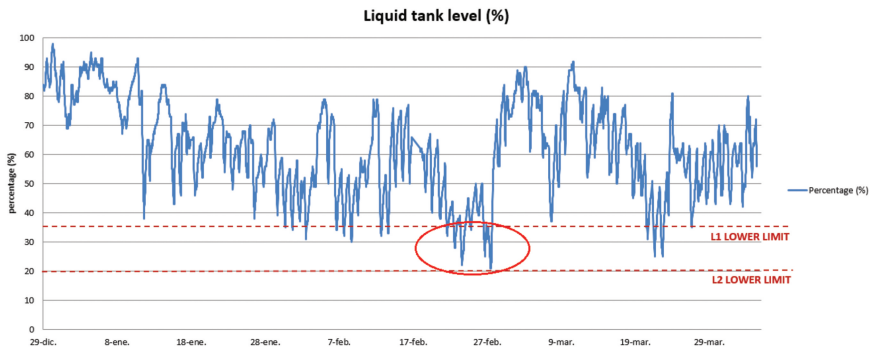All the charts in Figures 7.26–7.28 are displayed in the data recorded by the remote HMI system. Any data out of the target values will give an alarm that must be checked out by the process engineer.

It has been seen that the monitoring system of the refrigeration temperature and tank level exhibit useful information for the control of the manufacturing process. The system works correctly and in-situ checks with the control through the monitoring system could avoid production problems and advise of maintenance needs that otherwise are difficult to detect. A good control of the refrigeration circuit will allow reducing the purge needs of the system and the cleaning operations.

7.2.5.3 Compressed air system from pulling system

For this application, the goal is to assure the correct workings of the compressor (Figure 7.29) and, on the other hand to detect any anomaly in the injection circuit (Figure 7.30) (pressurized tanks, pipelines, connections, etc.).

Using wireless sensors located at the compressor outlet and at the pipelines that connect to the injection chamber, air pressure, temperature and flow are monitored. The operator checks out on the local display warnings and alarms in case of abnormal values of temperature, flow or air pressure were detected. The process engineer checks out the system performance through the Remote HMI according to the Control Program.

Figure 7.31 shows the data record of the evolution of the air pressure obtained from the Remote HMI interface for a period of two and a half months of production (Monitoring period 01/01/2018–14/03/2018).

Figure 7.29   Detail of the main compressor that supplies compressed air to the resin injection system.
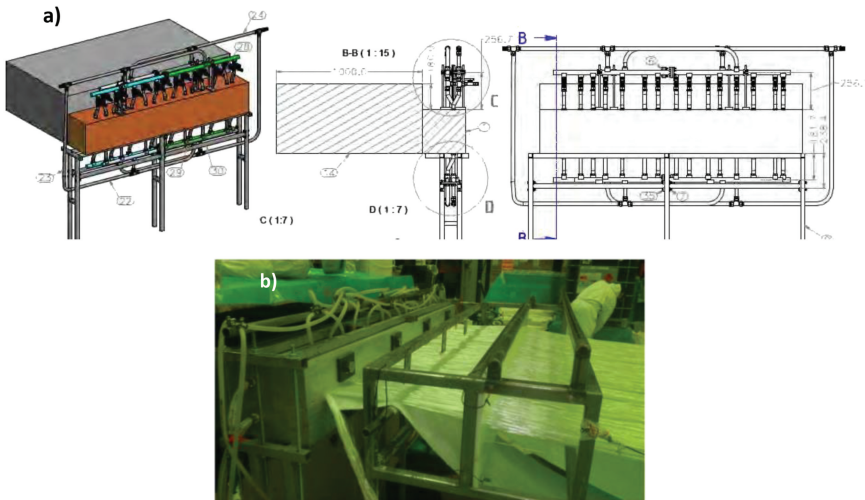
Figure 7.30   (a) Schema of the injection chamber resin injection circuit. (b) Detail of the resin injection circuit.

Figure 7.31 shows that the air pressure remained above 8 bar and thus above the operational limits fixed during the period studied. In fact, the Control Program states that:

- Value > 6 bar that corresponds to regular operational conditions no alarm;
- L1: Value < 6 that corresponds to non-regular operational conditions, but not critical alarm turns to yellow;
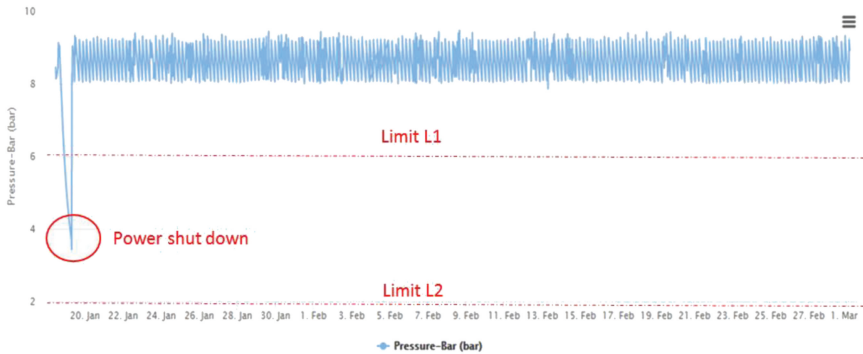
Figure 7.31 Data record of the evolution of the air pressure obtained from the Remote HMI interface.

	Green Alarm	Yellow Alarm	Red Alarm
Temperature	10–25°C	5–10°C//25–30°C	> 30°C and < 5°C
Pressure	$>6\%$	2–6%	$< 2\%$

- L2: Value < 2 bar that corresponds to critical operational conditions alarm turns to red.

The air pressure remained stable oscillating between 8–9.5 bars (Figure 7.32). This oscillation is normal within the regular operating regime of the compressor because the compressor's engine does not work constantly. It automatically turns on when the pressure inside the pressurized tank falls below a consigned value (about 8 bar).

Operators did not detect any alarm trigger in relation to air pressure during this period except 19 January when the compressor shut down due

Figure 7.32 Detail of the data record of the evolution of the air pressure obtained from the Remote HMI interface.

to a power cut. Because of this event, the pressure decreased, triggering the yellow alarm. However, the red alarm was not activated since the pressure remained above two bars until the power was recovered. The compressor was able to keep the air moderately pressurized inside the tank for few hours (Figure 7.31).

Therefore, the in situ control carried out by the operators matched with the data recorded by the remote HMI system and checked out by the process engineer.

The temperature inside the room where the compressor is placed was also recorded. Figure 7.33 shows the data record of the evolution of the temperature obtained from the Remote HMI interface for a period of two and a half months of production (Monitoring period 01/01/2018–14/03/2018).

Even though the temperature oscillates between 10 and 25°C inside the room, the effect on the air pressure control is negligible (Figure 7.32). The heat generated by the operation of the compressor causes this oscillation. The room is conditioned by means of an extraction system that brings out the room the hot air when the temperature inside the room reaches 25°C. On weekends, highlighted in red, the compressor remains off and the temperature is closer to the exterior room temperature.

During the monitoring period, the temperature values stayed within the operational limits allowed, between 5–30°C. Therefore, the current conditioning system is enough to keep the temperature within the recommended operation values.

In addition to the sensors installed in the compressor room, another sensor has been installed at the furthest point from the compressed air circuit (Figure 7.34). This sensor will help to check if the pressure in the compressor
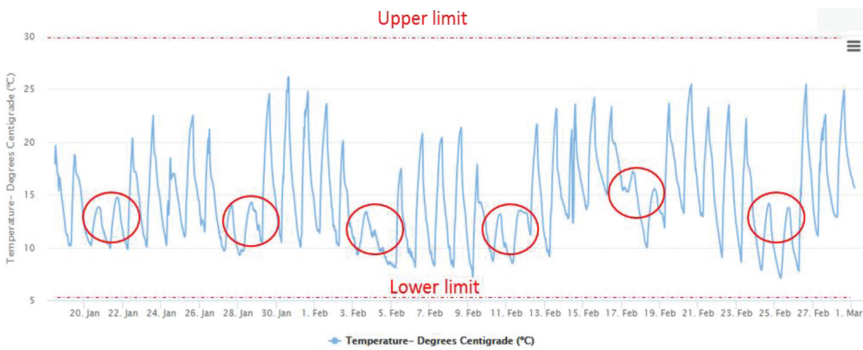

Figure 7.33 Data record of the evolution of the temperature obtained from the Remote HMI interface.

Figure 7.34   Pressure and temperature sensor installed in the compressed air circuit.

maintains the same or if there is any loss along the circuit. The installation of the sensor had to be done while there was no production in progress and complies with different security.

7.3 Maintenance in Press Forming Machinery

Contributors: Urko Zurutuza, Javier Fernandez-Anakabe, Ekhi Zugasti,
Petri Helo, Lauri Välimaa, Mathias Grädler, Mikel Mondragon,
Andoitz Aranburu, David Chico, Oier Sarasua, María Aguirregabiria,
Xabier Eguiluz, Iosu Gabilondo, Eduardo Saiz, Iban Barrutia Inza,
Mikel Viguera, Félix Larrinaga Barrenechea, Mikel Anasagasti,
Jon Olaizola and Ricardo Romero.

This use cases focuses on stamping press machines, which are metal working machines used to shape or cut metal by deforming it with a die. See Figure 7.35 for some examples of this kind of machinery.

This kind of press is built by FAGOR and, during its active lifetime, might be capable of giving more than 40 million strokes characterized by impressive force and precision, insofar as the press is used and maintained appropriately. This use case considers two scenarios.

The first scenario focuses on the press forming machinery itself. The customers expect both high quality of the pieces produced by the machine, and high availability, which led FAGOR to incorporate cutting-edge
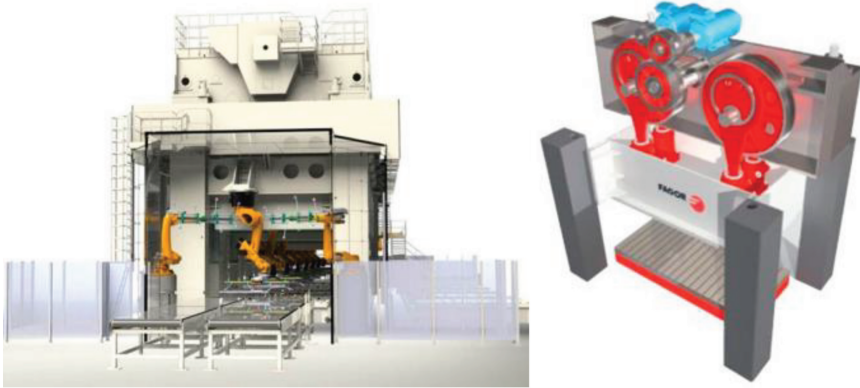
Figure 7.35   Fagor Arrasate mechanical and servo driven presses.

technologies in their products as a means of enhancing products robustness and functionality in order to facilitate proactive maintenance activities.

The second scenario considers the pneumatic Clutch Brake, which is a critical device within the mechanical press machine. The Clutch Brake is responsible of activating and stopping the tool of the press machine in order to perform different processes. In this scenario, the clutch brake became a CPS itself, able to provide data regarding its own health conditions.

7.3.1 Introduction

FAGOR ARRASATE S.COOP is a company of 800 employees specialized in designing, manufacturing and supplying sheet metal forming machine tools. Fagor Arrasate was created in 1957 and, since then, has expanded its products and business in an significant manner, being now one of the world leaders in the field. It is one of the 5 biggest manufacturers in the world in terms of turnover and the first one considering the product's portfolio.

The Company is located in the Basque Country, in the north of Spain, very close to the French border in the most industrialized area of the country and surrounded by a traditionally metallurgical and exporting environment.

FAGOR is a world leader in the design and manufacture of mechanical and hydraulic presses, complete stamping systems, transfer presses, robotised press lines, press hardening, forging; Cut-to-Length, Slitting, Combi and multiblanking lines; Processing lines as pickling lines, skin passes, reversible mills, painting, galvanizing or levelling lines; special metal part forming systems, strip roll forming, flexible roll forming, rotor/stator cutting equipment, dies and many other types of equipment.

Fagor Arrasate serves to numerous sectors, with a particular focus on the car industry, the domestic appliances industry, the Steel Industry and Service Centres. For Fagor Arrasate a key goal is the constant collaboration with its customers, so there is a close and continuous presence in order to give solutions for any process with the most adequate technology.

GOIZPER S.COOP is one of the leading technology suppliers in power transmission components, such us brakes, clutches, turning systems, gear boxes, cams or elevators. GOIZPER designs, manufactures and supplies customized power transmission components to meet market needs in sectors like metal forming, automotive, aeronautics, packaging, construction, marine, machine tools, etc.

As mentioned, Clutch Brakes and gearboxes are products consumed in automotive industry and Fagor Arrasate is currently using them within their mechanical press machines (see Figure 7.36). In other words, Goizper is one of Fagor Arrasate's current power transmission components supplier.

GOIZPER's headquarters and productive plant are located in Antzuola, Spain, and almost 80% of the sales come from exports all over the world. The maintenance of sold parts has become an issue due to the different locations of the parts around the world.

GOIZPER also has another division, totally different, focused on the design, manufacture and marketing of manual sprayers and dusters for treatments in farming, gardening, industry, construction, cleaning, pest control and vector control.

7.3.2 Scope and Logic

The final customers of FAGOR ARRASATE produce products with high levels of quality and availability seeking a drastic reduction of high cost caused by production downtimes with required maintenance-repair operations and a better delivery times' compliance. This is why FAGOR

Figure 7.36 Clutch Brake (left) and gearboxes (center) by Goizper, used in Fagor's mechanical press machines (right).

ARRASATE needs to increase the reliability of machines and components. To meet this challenge FAGOR ARRASATE is continuously incorporating cutting-edge technologies in their products as a means of enhancing products robustness and functionality in order to facilitate proactive maintenance activities.

7.3.2.1 Background information on the press machine

Mechanical and servo driven press machine elements have been analysed using sensor technologies in order to improve maintenance strategies for detecting early failures on the cranks and in forming elements of the stamping press.

A platform has been developed where the data from different components of a press machine could be captured, monitored, transmitted, stored and analyzed in order to come to reliable predictive and proactive maintenance. The data is analyzed and monitored via local or cloud level (see Figure 7.37).

The components of the press machine that require sensors with innovative CPSs are:

- Bushings (Temperature and oil condition status);
- Bolster (Relative displacements);
- Head (Structural health);
- Gear axis (Torque);
- Engine (Tension and current);
- Connecting rod (Displacement, forces).

Figure 7.37   Predictive maintenance HMI platform.

Figure 7.38 illustrates the location of the source of each component data.

The objectives of this use case are:

- Maintenance Cloud Platform development;
- Torque measurement using wireless sensors;
- Head structural health monitoring;
- Torque measurement using wireless sensors;
- Bushing status measurement;
- Gears wear measurement;
- Crank strain and force measurement;
- Press unbalances forces measurement;
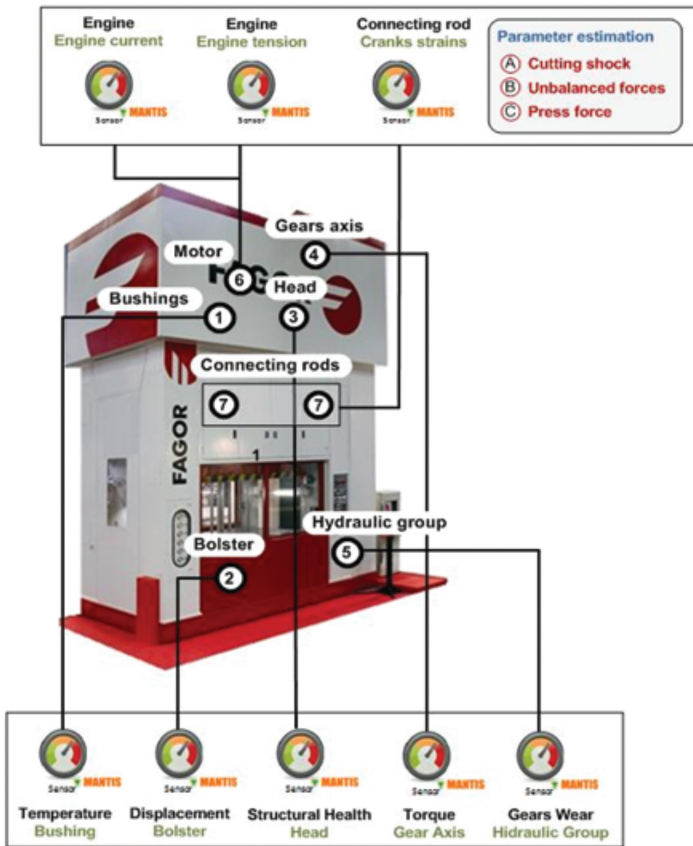- Press cutting shock measurement.

Figure 7.38   Location of data sources.

7.3.2.2 Background information on the clutch brake component

A critical device within the mechanical press machine is the pneumatic clutch brake. The Clutch Brake is the responsible component of activating and stopping the tool of the press machine in order to perform different processes. In this case, metal forming process is considered, where the clutch brake works as a mechanical commutator. The clutch brake components suffer from degradation during operation. Component degradation usually causes machine failures and downtime, generating unwanted and unexpected costs. Figure 7.39 shows how a pneumatic clutch brake looks like.

The focus lies on identifying issues related to the maintenance of the clutch brake, adopting strategies to monitor and to make decisions against those issues.

Downtimes caused by the clutch brake have been listed taking into account the number of stops. From this list, the following topics were picked in order to analyse and solve.

- Friction material slippage detection at clutching;
- Friction material slippage detection at braking;
- Friction material wear and misalignment;
- Piston chamber air leakage;
- Brake springs degradation.

For the Clutch Brake scenario, two demonstrators have been used. One of them is situated at MGEP facilities in Mondragon, Spain, and it consists of a Fagor mechanical press machine that contains a GOIZPER's Clutch Brake component. The other one is located at GOIZPER's facilities in Antzuola, Spain, and it consists of a Clutch Brake wear test bench.

MGEP Press Machine demonstrator

In Figure 7.40, Fagor's press machine demonstrator's front and back sides are shown. This machine contains a GOIZPER Clutch Brake and it is located in MGEP shopfloor for small size metal parts forming.

Figure 7.39   GOIZPER pneumatic Clutch Brake outside and inside.

Figure 7.40   MGEP press machine, demonstrator 1.

This machine is a mechanical press machine used for low duty metal forming processes. It contains a pneumatic Clutch Brake at the back side (shown in Figure 7.41) in order to activate and deactivate the ram of the press. The ram is the orange part of the press which performs the action of metal forming.

Figure 7.41   GOIZPER clutch brake in MGEP demonstrator before MANTIS.

Figure 7.42 shows the back side of the demonstrator, and the GOIZPER clutch brake next to the flywheel (orange). Sensors have been installed within the Clutch Brake in order to capture data and execute the algorithms for the preventive maintenance of the component. The electric motor, power source of the application, is located at the top side (not visible) connected to the flywheel by means of a black belt.

Within the pneumatic circuit, the electro valve is located at the right side of the picture, opening and closing the air flow into the clutch brake. This air is introduced through the black tube connected to the application axis, in the middle of the picture. Sensors have been installed in order to receive data from the Clutch Brake. These installed sensors are visualized in Figure 7.40, which indicates each sensor's location.

GOIZPER Test Bench

The second demonstrator is the test bench in GOIZPER's installations (Figure 7.43). Friction discs accelerated degradation has been forced in order to get the data from the beginning (%0 of wear) until the end of the friction discs life (%10 of wear).

The installed sensors are not giving direct information, all the captured data needs a processing stage (Figure 7.44) in order to know the actual health

Figure 7.42 Clutch Brake in MGEP demonstrator with the sensors installed.

Figure 7.43 GOIZPER Clutch Brake test bench, demonstrator 2.

of the Clutch Brake. From these calculations, different problematic scenarios have been identified. Some of the algorithms are located in a local data logger and the rest are located within the cloud.

7.3.3 MANTIS Solutions for Press Machine

This section focuses on the first scenario of the use case, and thus on the solution implemented by FAGOR for the press machine itself.

Figure 7.44   Local data logger for data capturing and processing.

7.3.3.1 Maintenance cloud platform

A new demand of technical solutions and services aiming at improving the efficiency of maintenance and repair operations is arising. In line with this need, FAGOR ARRASATE wants to offer to its customers a broad range of maintenance services based on digital technologies that allow the company to collect real-time data from the press machines installed all over the world. As Fagor and Goizper are different firms, each company is developing its own Cloud solution. However, interoperability has been taken into account for the cases that both partners work together.

7.3.3.1.1 *Solution approach*

The solution selected by FAGOR has been to develop a digital cloud platform where data from different components of a remote press machine can be collected, monitored, transmitted, stored and analysed providing services for reliable predictive and proactive maintenance. The cloud platform has an architecture divided into two different environments: **On-cloud** where data coming from the different sources are stored, processed and analysed and **On-premise** where is the data acquisition system to extract the data from the different sensors of the machines. Data format is based on the **Event Information Model** adopted for the present project and communications among both environments are secured by using VPN tunnels that guarantee the data integrity, confidentiality and availability.

In the Cloud, a Big Data architecture following the MANTIS reference architecture principles and based on different applications has been designed aiming at supporting fault-tolerance and high scalability. In this way, each part of the system is independent and loosely coupled. The general architecture of this approach is illustrated in Figure 7.45.

On-Cloud architecture consists of the following core components:

- **Elastic Search:** The data coming from the different manufacturing facilities will be persisted in a NoSQL database. It allows the storage of huge volumes of data as well as an optimised search mechanism with a flexible approach to perform a number of aggregations;
- **[Apache Kafka]:** A distributed queue message system to decouple the different applications by following the publisher-subscriber communication pattern. This technology uses a topic approach to categorise the data. In this work the different data natures are published through different topics;
- **Proxy:** This application is an HTTP proxy that receives the data from on-premise sources and categorises them in different data natures by considering their origin. Afterwards, the data is published in different Kafka topics;
- **Real Time Processing:** this application has three objectives:
    - Extracting the raw data categorised by means of the Kafka topics;
    - Persisting the data in elastic search also categorised by data nature;
    - Executing data analytics to detect possible alarms (that are published to Kafka), and performing a predictive maintenance.
- **API:** A REST API in charge of exposing the functionalities supported by the Cloud to allow the connection with a front-end (App Web).
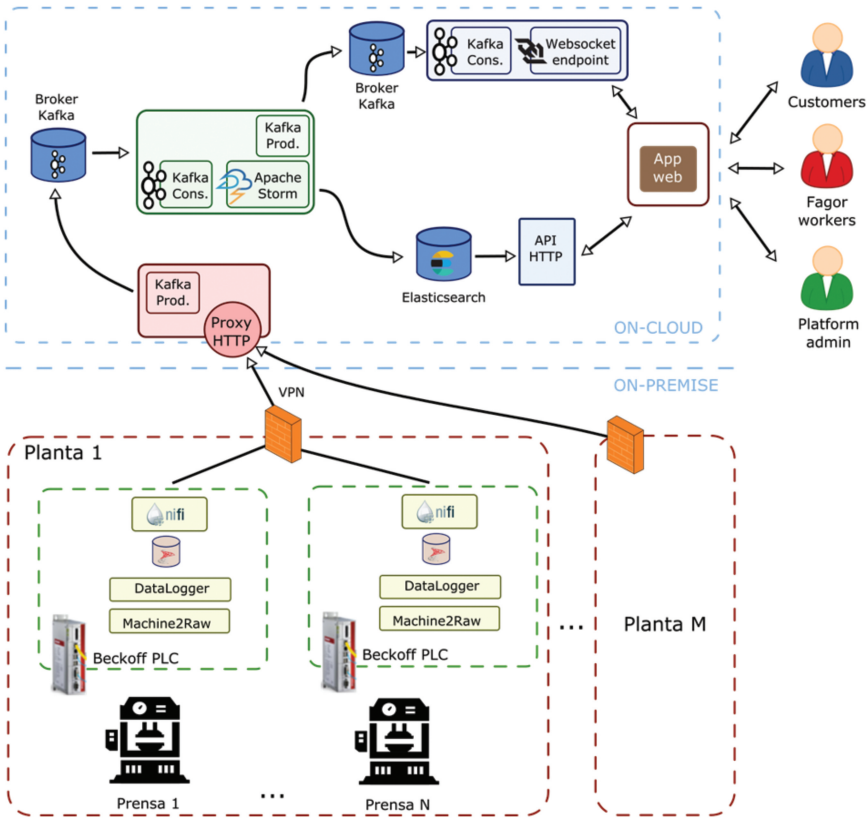
Figure 7.45 Architecture design.

Among the functionalities, they are worth emphasising the possibility of querying an Elastic Search NoSQL database to obtain historic data by applying distinct filters, the execution of [CRUD] operations over the resources required and the users' management system. This API is implemented as a [Spring Boot] application, a framework to simplify the creation and development of Java Web applications. In addition, OAuth 2.0 protocol [Aaron Parecki, 2018] is adopted in order to guarantee the security in the communications;

- **Alarm:** This application consumes from Kafka the alarms generated from real-time processing and push the corresponding values to a front end to trigger an alarm when necessary.

On-premise is considered as the technological solution deployed on the manufacturing plants of the customers where their machines are located. This environment usually has limited hardware, software and network resources. This leads to frequently delegate the high-consume processes and the exhaustive exploitation of the data to the cloud environment. Therefore, in some cases data will be directly submitted to the Cloud, while in other cases data will be normalised, standardised and persisted in a local database to subsequently be submitted to the Cloud. In the On-premise environment there is an Industrial PLC that provides resources with which the information from the automation is obtained from the sensors of the machines. This computer executes the following modules:

- **Machine2Raw:** This is the system in charge of extracting the data from the different PLCs through a PLC obtaining, in turn, the data from the different machines of the customers;
- **Datalogger:** This system is responsible for storing in a local database the data coming from the sensors and the systems that are being monitored;
- **Local Database:** A local database in SQL server to allocate the raw data structures provided by the sensors. In this database there are some triggers to centralize the information in a table that is the entry point for Apache NiFi;
- **Apache NIFI** [The Apache Software Foundation, 2018]**:** A technology that processes the raw data stored in the local database by the DataLogger. This system allows defining data-flows in a visual way. It is fault-tolerant, has a low-latency and is able to manage a high volume of data. After processing the data, Apache NiFi transfers the data to the cloud through a proxy.

7.3.3.1.2 *Results*

As a first result, FAGOR ARRASATE ended up having a cloud platform (Figure 7.46) to monitor the status of the press machine park running on their customers' premises. This platform provides several functionalities to create the network of press machines, to collect and monitor data from selected components, to analyse them and triggering alarms.

For operating the cloud platform, a control room (Figure 7.47) is set up at FAGOR's headquarters in Arrasate. This will allow the company to offer new maintenance services to its customers looking for increasing their press machines performance and availability. This way, the company aims to strengthen their market position and to create new business opportunities.
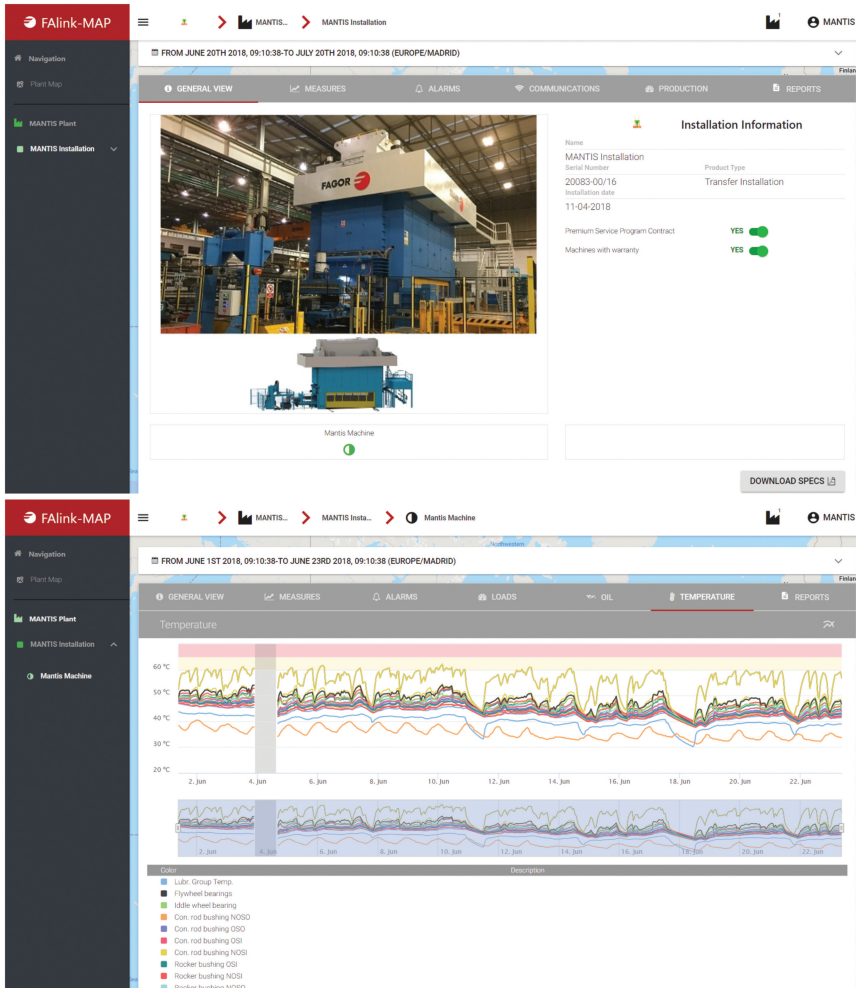
Figure 7.46   FALINK Cloud Platform.

7.3.3.2 Torque measurement using wireless sensors

Press machines manufacturers are confronted with increasing technological and cost pressure. Many customers demand faster and more precise presses. The more precisely force is applied in a press machine, the higher is the quality of the parts manufactured. Thus, increasing the press machines accuracy is one of the most important challenges for manufacturers of these assets. In addition, the market requires increasingly faster press machines

Figure 7.47 FAGOR ARRASATE future control room.

that, at the same time, offer higher bandwidth to increase production output in existing systems.

Nowadays, the torque of the press gear shaft is measured indirectly from the force that is applied in the connecting rod. This measure is quite precise but it needs to be continuously recalibrated to keep it accurate. To solve this problem, the torque is measured directly by using wireless sensors placed in the press gear shaft.

7.3.3.2.1 *Solution approach*

As a solution, IKERLAN has designed and manufactured a prototype of a shaft-adapted wireless sensor node that comprises a transducer based on torque oriented gauges, a signal conditioning circuit and a signal processing software, the latter allowing local preprocessing and treatment of the collected data by means of intelligent functions.

The design process has been made following two main phases:

- **Phase 1: Testbed validation**

Before starting the development of the wireless torque sensor, a preliminary validation step was made in testbeds both in IKERLAN and in the Try-Out press machine of FAGOR ARRASATE. This was an initial requirement to ensure the proper functioning of gauges, generic electronics and wireless communication for working in press-based conditions.

Two types of strain gauges were used: FCT strain gauges designed for torque measurement and FCA gauges chosen as a pair of strain gages oriented at 90. Also, a generic signal conditioning circuit was used where the signal

is then processed by a low power microcontroller, which transmits data wirelessly with a radio module. The data is received by a similar access point, which is controlled by a LabVIEW National Instruments Corporation interface (configuration and visualisation).

Initial tests were carried out on two testbeds scenarios. In the first testbed (Figure 7.48), different weights, which correspond to corresponding micro strains, were loaded on the bar and static measurements were performed in half bridge and full bridge configuration for gauge calibration. In the second one (Figure 7.49), the test was made using a motorized test bench. To check if measurements were suitable, the electric engine speed was slowly increased resulting in increases of the measured torque.

With regards to wireless communications, two main challenges were tested: (i) signal attenuation due to the rotation of the emitter around the shaft and (ii) multipath fading due to RF signal reflections in the metallic (steal) elements of the head of the press in which the torque sensor will be installed. Tests were successful, taking into account that depending on the angular position of the shaft, and therefore, on the relative position of the transmission and reception antennas, more or less amount of power is received periodically.

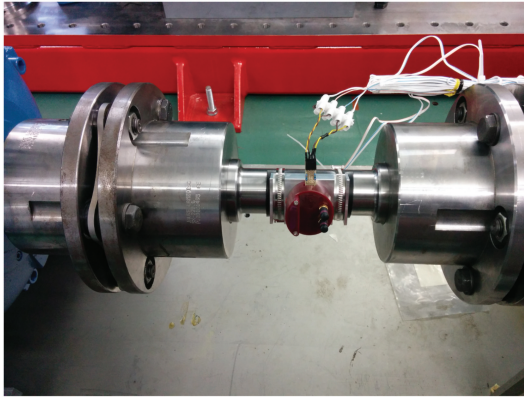**Figure 7.48**   Static testbed scenario.

Figure 7.49 Dynamic testbed scenario.

A similar test has been performed in the Try-Out press machine from FAGOR ARRASATE. In this case, both the emitter and the reception antenna have been placed in a realistic place within the head of the press machine as in can be seen in Figure 7.50.

Once the top cover is closed, creating a complete metallic case, it was observed how the received signal was not as clean as the one in the previous measurements due to multipath reflections. The statistical features obtained from this signals were used in the selection of the most suitable wireless communication technology to be used for the torque sensor.
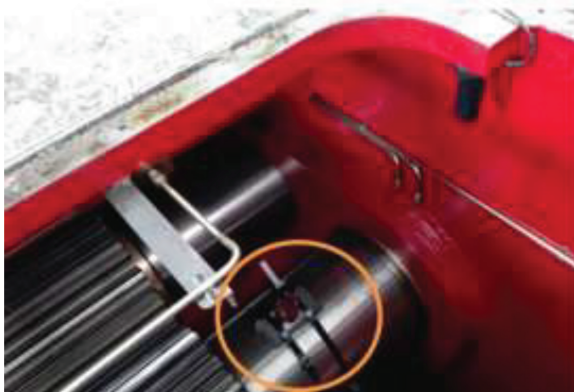

Figure 7.50 Measurement setup in the press machine.

- **Phase 2: Design and development**

Once the concept and the elements of the device (gauges, conditioning and processing, radio) were validated in a rotational environment, the system design and development was started taking into account the following Try-Out press specifications:

- Shaft material: F1140 (C45E) steel;
- Shaft dimensions:
  - Diameter: Φ 310.07 mm
  - External diameter: Φ 360 mm
  - Width: 150 mm
- System thickness: 25mm max
- Electronics & Cover:
  - No screwing
  - Speed: 88 rpm
  - Expected torque:
    - 188762 Nm
    - ∼200 microstrain
- Environment:
  - Temperature: <45°C
  - Subjected to oil: CLP-150ftesp @ 400cm3/mm
- Placement: Head of the press machine (see Figure 7.50)

From these specifications, a prototype of the wireless sensor node was designed and developed. It consists of a single PCB with the necessary interfaces to attach torque gauges, besides the conditioning, processing and wireless communication electronics. The whole system is powered by a rechargeable lithium ion polymer battery and it is encapsulated and protected by a plastic cover in the shape of the press' secondary driving shaft, which is prepared to avoid oil leakage (see Figure 7.51).

In order to configure the system and show the measured data, an user interface was designed in LabVIEW. From this interface, the Gage Factor, Poisson Ratio, Young modulus and the bar diameter can be configured. Typically the amount of received data is huge, so data values are averaged and only the average value is visualised in the user interface.

7.3.3.2.2 *Results*

Once the design and fabrication of the wireless torque sensor was finished, the sensor was installed in the Try-Out press machine from FAGOR
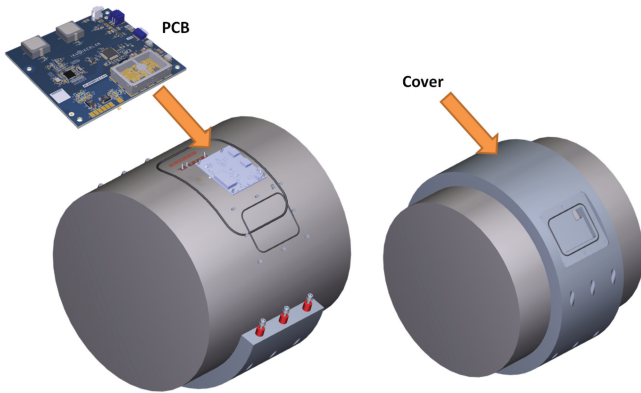
Figure 7.51   Wireless sensor node.

ARRASATE. First tests regarding the overall performance of the sensor were successful providing signals with the torque measurements were sent to an external laptop were they could be visualized (Figure 7.52).

Later, the complete validation process was carried out. This process aimed to test de accuracy of the sensor's measurement against several torque and speeds and the robustness of the wireless communication protocol employed.

15 different tests were carried out combining 30%, 60% and 87% of the nominal torque of the press, 57%, 78% and 100% of the nominal speed and several configurations of the sensing electronics. These results were

Figure 7.52   Torque measures.

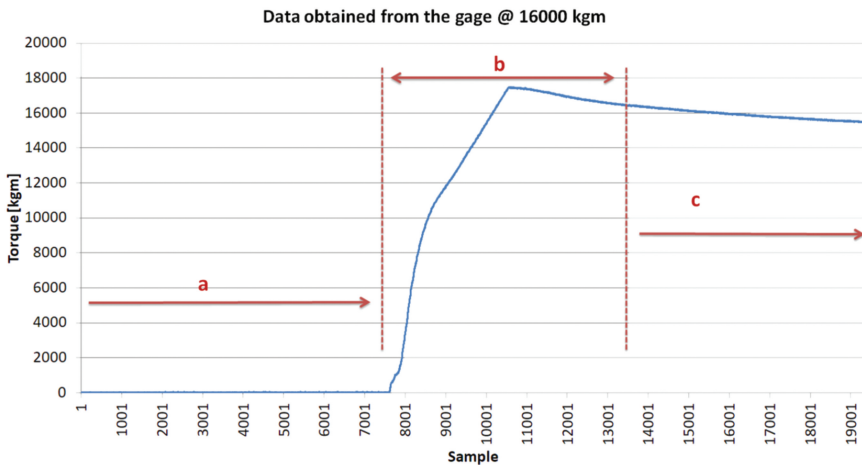compared with an estimation of the torque at the drive shaft obtained from an overload pressure evolution analysis. Besides, some measurements regarding the performance of the wireless communication were also taken.

Figure 7.53 shows the results of the test in which the maximum torque (87%) and the maximum speed (100%) were configured at the press machine.

The measured torque values at almost each stroke are close to 60 kNm, which fit the estimated torque values. Moreover, the clutch brake engage and disengage events were captured.

In general terms, it is considered that the obtained results are valid, taking into account that they are compared with estimated values and not with another measurement obtained by a commercial system. However, regarding the amount of data shown at the measured torque values, some data can be missed either on the positive or the negative peaks, as the same amplitude should be acquired for each stroke. With regards to wireless communications, in general the expected performance in terms of data throughput and network availability has been achieved. However, the loss of some data packets has been detected, which should be corrected in future versions.

As future work, the use of antenna diversity inside the shell of the press machine will improve the communication between emitter and receiver hence decreasing the number of packets lost. Besides, being the energy management of the system a key feature if it is pretended to leave it permanently attached to the press machine's drive shaft, a more energy efficient redesign will be carried out together with the development of an energy harvesting system to power up the wireless sensor node.
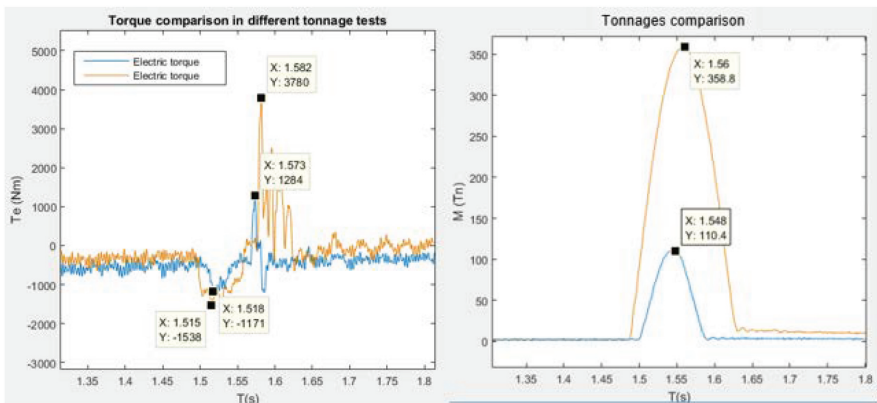

Figure 7.53 Comparison between estimated and measured torque values (87% of the nominal torque, 100% of the nominal speed and gain 1000).

7.3.3.3 Head structural health monitoring

The press structural components are welded steel structures where, on rare occasions, cracks may appear. The crack initiation is usually associated to fatigue damage in the welds and weld transitions (maximum design load is not exceeded due to the overload security devices). Fatigue is a cumulative phenomenon due to fluctuating loads, when material is subjected to repeated loading and unloading. The nominal stress for such loads may be much less than the ultimate tensile stress limit of the yield stress. If the loads are above a certain threshold, microscopic cracks begin to form at the stress concentrators. Eventually, a crack could reach a critical size, suddenly spread and provoke the fracture of the structure.

7.3.3.3.1 *Solution approach*

Physics based degradation models (see Section 3 of Chapter 5 on RUL) are implemented to detect most damaged zones of the press structural components. Based on the classical high cycle fatigue damage model, a damage indicator map is implemented. A damage threshold is set (Damage threshold = 1.3, for example), which is associated with a minimum length crack (2 mm, for example) appearance probability. Based on the measurements of real forces in the press rods and the stresses calculated at every point with Finite Element Models, 3 methods recommended by the International Institute of Welding are used to calculate the damage indicators at the welded structural components (mean stress, hot spot and notch stress).

As the real forces are being measured, dynamic and asymmetric loading effects are taken into account and accumulated over time. At the same time the damage indicator is calculated, the remaining time to reach the predefined threshold (see Section 3 of Chapter 5 on RUL) is also calculated. Identification of unexpected premature occurrences can easily be identified and analyse probable associated Root Cause (see Section 2 of Chapter 5 on RCA).

Additionally, for certain cases, a minimum crack length is supposed and a second physics based degradation model is applied to study the fracture mechanics. This is the field of mechanics concerned with the study of the propagation of cracks in materials. During the second stage of crack propagation or stable propagation stage, Paris' law is used to estimate the crack propagation under certain load. The time needed to reach a threshold critical length is calculated. This is interesting when a crack is detected and the evolution needs to be estimated in order to schedule the corrective maintenance action.

Two different crack sensors were tested to detect or measure crack propagation: a commercial local crack gauge and a conductive ink sensor at a stage of development. Once a crack has been localised, the RUL to a critical crack length is calculated. If crack length data is available, Particle Filter method is used. This method combines the physical model and the available measurements in order to improve the RUL estimation to the critical crack length.

Algorithms and sensors are applied in a testbed prototype before the final application in the press machine head, shown in Figure 7.54. As a result of this, the structural health of the head of the press machine is monitored by means of two developments: i) The setup of a testbed for structural failure prediction and simulation and, ii) the analysis of conductive inks for crack detection.

- **Testbed for structural failure prediction and simulation**

Due to the difficulty to artificially create structural failures in a real press, the press head Structural Health Monitoring scenario has two demonstrators:

- A fatigue testbed, where welded structural details are tested until complete fracture. The algorithms and sensors are applied and checked in this demonstrator prior to applying them in a real press machine;
- A real press, where a structural damage and associated RUL indicator is applied, taking into account the results obtained in the fatigue testbed demonstrator and the features of the real press.

In the case of the testbed, a fully sensor welded specimen has been submitted to fatigue loading until its complete fracture. The welded specimen is selected according to a structural detail located in the press head. The material of the specimen is the same of the press head as well as the welding procedure. The thicknesses of the sheets have been reduced according to the testbed load capacity.
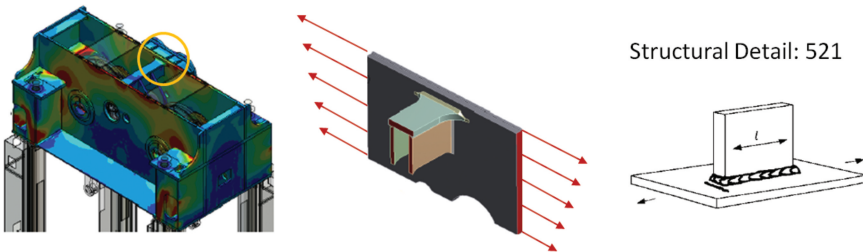

Structural Detail: 521

Figure 7.54 Selected structural detail from press head.

A Finite Element Model of the specimen (Figure 7.55) was built in order to estimate the stresses at any location of the welded specimen.

The damage and RUL are estimated every minute during the test. Complete sensor data for a period of 20 seconds were stored every hour. Some results to remark are that the complete fracture occurs at 3.1E6 cycles, 3.8 times compared to the design life for 95% (8.4E5). The macrocrack initiation (failure) is estimated to occur near 1.25E6 cycles, 50% above the estimated design life. Crack initiation occurs where predicted by the FE model and the weld damage methods (Figure 7.56).

- **Conductive inks**

The objective is to proof the concept of using conductive inks to detect cracks in a mechanical test specimen. Tests with several conductive inks have been carried out. Different circuits are painted by this conductive ink on top of a test specimen coated manually with an insulating layer (e.g., Magnesia 919). The specimen is heated at different temperatures to increase the ink conductivity. During this heating step, cracks appeared in the insulating layer, which is interesting for the application since the objective is to study the bottom crack effect in the conductive ink. It is observed that it is difficult to apply the conductive ink homogeneously along the insulating layer surface.
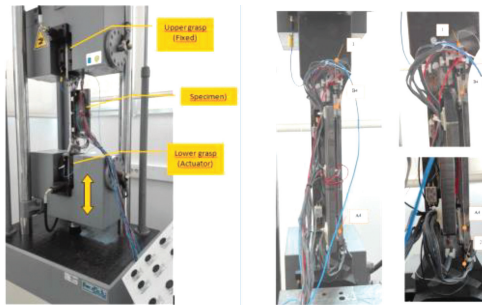

Figure 7.55   Fully sensorised welded specimen in the testbed.
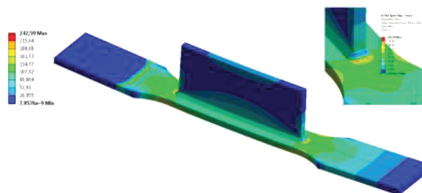

Figure 7.56   Stress plot in the testbed due to a unitary load.

After these preliminary tests, a conductive ink is deposited on top of a mechanical test specimen prior to subjecting it to a mechanical fatigue test to create a crack, which should be detected by the conductive ink.

First, the "silver conductive adhesive pro" from RS is deposited on top of a mechanical test specimen shown in Figure 7.57, next to the pre crack, in perpendicular direction. First of all, an insulating layer is bonded directly on the test specimen surface. Secondly, different insulating adhesive layers are tested: a kapton adhesive film (A) and a Teflon adhesive film (B). Then, a rectangle of silver conductive ink (4 x 40 mm2) is painted on top of each insulating layer, using the paint brush provided by the ink. Finally, electric cables are bonded in the four ends using a silver conductive epoxy adhesive.

Regarding the experimental conditions for fatigue test, a mechanical test specimen made of S235JR steel material and CT Compact Tension geometry is created. The stress is applied in a different way in two different phases: (i) 225000 cycles of 0.8–8 kN with a frequency of 10 Hz and (ii) 109000 cycles of 1–10 kN with a frequency of 10 Hz. In both phases, every 1000 cycles, a constant force is applied (F = (Fmax + Fmin)/2) during 10 seconds. The subsequent post processing is carried out with data acquired in these intervals of 10 seconds. The sampling frequency is 20Hz.

Apart from the signal of conductive inks, two other signals are acquired in the test, in order to be able to compare afterwards the results with a reference (a commercial crack detection gauge and a commercial extensometer, see Figure 7.58 for all the elements used in the test, and Figure 7.59 for the final set-up for the experiments).

7.3.3.3.2 *Results*

Structural damage and associated RUL indicator based on fatigue damage are ready to be integrated in a real press machine. Stress at critical positions is obtained from a finite element model and online experimentally measured forces.
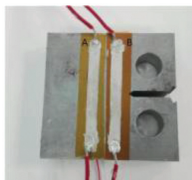

Figure 7.57   Mechanical test specimen with two sensors based on silver conductive adhesive deposited directly on top of the surface.
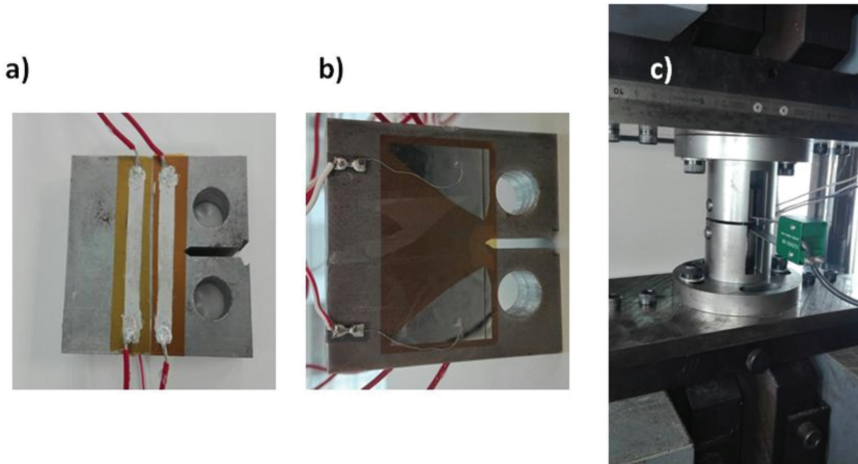
Figure 7.58 (a) Two sensors based on conductive inks placed in one side of the specimen; (b) Commercial crack detection gauge placed on the opposite side of the specimen; (c) Commercial extensometer placed in the crack, during the fatigue test.
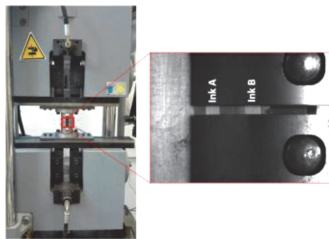
Figure 7.59 Picture of experimental set-up.

With regards to conductive inks, the following results were achieved. The blue line in Figure 7.60 corresponds to ink B, which is placed closer to the pre-crack and is deposited on top of the Teflon insulating layer. On the other hand, the red line corresponds to ink A, which is placed farther away from the pre-crack and is deposited on top of the Kapton insulating layer.

The signal increase of ink B indicats the presence of a crack. However, this increase started too late, when the tests specimen was completely broken. The behaviour of ink A is similar, where the signal began to increase later, again when the test specimen was completely broken.

It is concluded that although apparently the ink performance is correct, the bottom insulating layer behaviour is not as expected. It is too flexible and does not transfer in a correct way the crack from the test specimen to the ink.

Figure 7.60   Ink signal during the second phase of fatigue test and ink pictures corresponding to the two points indicated in the graph. Pictures of both sides of the test specimen, at the end of the fatigue test.

Both insulating layers are extended too much before cracking due to their high flexibility.

Direct crack detection methods are going to be tested (ink sensors and crack gauges, acoustic emission sensors) and an indirect model based damage detection method (Extended Constitutive Relation Error approach). This method compares the measured strain at different moments of the system with the ones expected by the model. It is a method to identify, localize and determine the severity of the damage.

Three research lines of interest are identified:

- Probabilistic approach of the fatigue damage;
- Stress estimation based on model and sensor data. Improvement of stress estimation and hence, structural damage and associated RUL indicators;
- Crack length estimation based on fracture mechanics and RUL estimation using particle filter based prognostics algorithm.

The last two will improve the estimation of the stress or the crack length by combining physical models and experimental data.

7.3.3.4 Bushing status measurement

Bushings are critical parts in press machines to reduce friction between rotating shafts and stationary support members. Depending on the working conditions and oil status, the bushing can increase its temperature getting stuck with the connecting rod. This failure forces to stop the stamping process and the time to repair it is about one working week. Taking into account that the one just described is the best case scenario when there are spare parts available in stock, due to the magnitude of the problem, it was necessary to tackle it within the MANTIS project.

7.3.3.4.1 *Solution approach*

It is considered that bushing failure due to seizure can be anticipated, and that it is possible to estimate its RUL by collecting and analysing defined measured data. Bushings were tested and run to failure at different ranges of working conditions in a test bench. A wear out model is created based on the time data obtained from the sensors and control installed in the test bench. For the analysis, regression based models are used as a first approach.

Bushing temperature and oil sensor signal based alarms are set in order to prevent seizure failure. This is done by limit and trending checking. A physical model represented in Figure 7.61 and based on DIN 31652 has been programmed in a simulator to characterise the theoretical behaviour of the bushing when it is working in hydrodynamic ideal conditions. For some given working conditions, the model calculates the expected equivalent friction, temperature rising in the bushing as well as the lubrication through the oil film thickness (Figure 7.62). This is done in two steps:

- Firstly, it is checked if the model describes the real behaviour of the bushings during the tests;
- Secondly, the model is used to define a safety working condition where bushing seizure should not occur. The main idea lies in detecting bushing abnormal behaviour when temperature alarm triggers in the defined safety working conditions.
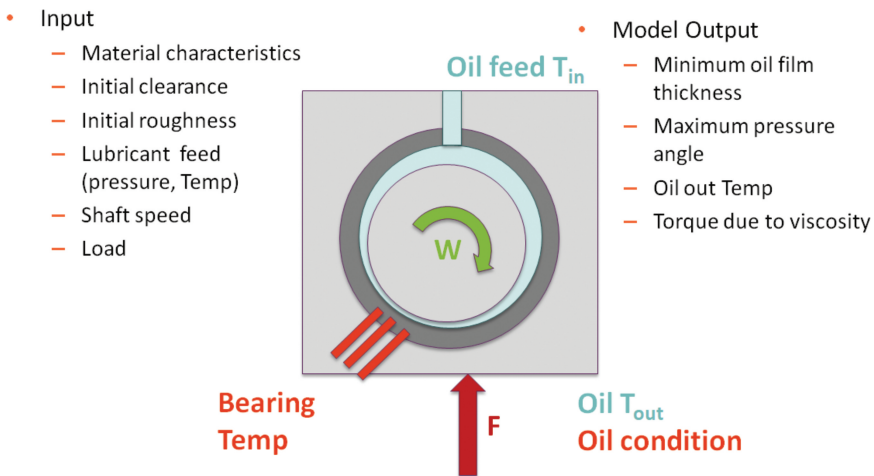

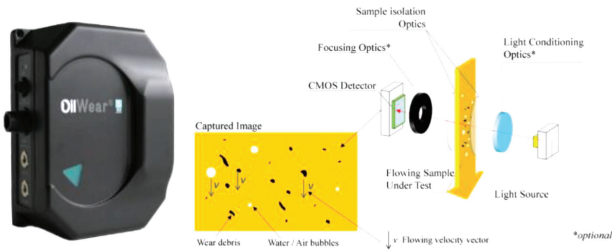**Figure 7.61** Bushing seizure model.

Figure 7.62　OilWear sensor.

7.3.3.4.2 *Results*

Some improvements were made to the test bench in order to improve the automation of the tests and lubrication conditions. Further tests are carried out in different working conditions and bushing types to complete the correlation between the measurements and the model, characterising the theoretical safe working conditions zone. Additionally, analysis is conducted to establish the safe working condition area based on real data measured from the tests.

Results of this analysis are expected to be applied in the future to press machines by installing temperature and oil sensors in critical bushings for collecting, monitoring and analysing real time data. This is in line with the company strategy of minimising these kinds of incidents.

7.3.3.5 Gears wear measurement

Another failure that happens in press hardening machinery is that, depending on the force and working conditions of the press, the gear can be damaged. Sometimes this damage is caused by wear with the passing of the years or working hours, sometimes the problem can raise much earlier caused by bad working conditions.

7.3.3.5.1 *Solution approach*

To predict that the gear is wearing out and must be replaced, it is possible to analyze the oil condition. Taking into account that the more hours the press works the more metal particles appears in the oil, it is possible to predict if the gear is wearing by analysing the oil itself.

The selected sensor for acquiring and monitoring the particles present in fluids is the OilWear S100. This sensor can classify particles larger than 20 μm according to their size and shape, to determine the root cause: fatigue, sliding or cutting.

The OilWear S100 is located in the hydraulic tube from where the oil returns to the tank.

Data is captured every scan cycle. The different data values are stored on a local server.

Once data is storage in the local server, data pre-processing is done in local mode, taking into account requited parameters such as the maximum and minimum values as KPIs. After data pre-processing, particles data that FAGOR ARRASATE considers are critical for their press machine working conditions will be analyse using specific algorithms for that.

Apart from acquiring oil particles characteristics data, some other data is acquired, in order to know exactly the main reason of the gears wear:

- Die reference;
- Press machine Total Strokes;
- Die Total Strokes;
- Press machine Speed: Stroke Per Minute;
- Press machines maximum force;
- Stamping force.

7.3.3.5.2 *Results*

The OilWear S100 sensor is now installed in FAGOR ARRASATEs press machine. During next months, data will be captured and preprocessed locally tacking into account defined KPIs.

The next step will be to integrate the data sources within the cloud platform.

7.3.3.6 **Press forces measurement**

The ram force is a fundamental parameter that affects the final quality of the produced workpieces. Furthermore, its deviations could cause damage to the press machine's components. Besides, in presses with multiple cranks, an unbalanced forces could appear due to an imbalance of the cranks or other components, affecting both the quality of the produced workpieces and the integrity of the press. The cutting shock effect is another undesired effect that has to be taken into account during any process of metal forming. If a big enough cutting shock is exerted, many components of the press can suffer damages.

These force measurements usually are carried out by hardware sensors located throughout the press structure and tooling, whose calibration loss are caused by the strongest forces the press experiences during its life cycle.

7.3.3.6.1 *Solution approach*

In order to overcome the limitations associated to the hardware sensors, indirect measurements are proposed by means of model based soft sensors that leverage the existing signals and the knowledge about the process that the system performs.

The servo driven press machine is modelled as slider - connecting rod - crank mechanism, considering also the gearbox between the connecting rod shaft and the servomotor shaft. Along with the servo driven press machine model signals are measured in order to use them as inputs for the model, such as the servomotor current signals, voltage signals and rotor position signal. A *Prediction Error Method* based soft sensor is used for estimating the coefficients of a friction model that completes the model of the system.

As visible in Figure 7.63, the servo driven press machine model is formed by three sub-models, each one generating a torque that interacts with the rest of the system.

After estimating the coefficients of the friction model, the whole model is utilised for estimating important coefficients that are:

- Press ram/slider force;
- Cutting shock effect;
- Unbalanced forces.

Model

Initially the system model mathematical representation is developed applying Euler-Lagrange function considering all the mechanical elements of the system. The shortened mathematical model is reported in Equation (7.5):

$$M\left(\theta\right)\ddot{\theta}\left(t\right) + N\left(\theta\right) + O\left(\theta\right) = \tau_e + \tau_{lb} - \tau_{fric} \qquad (7.5)$$
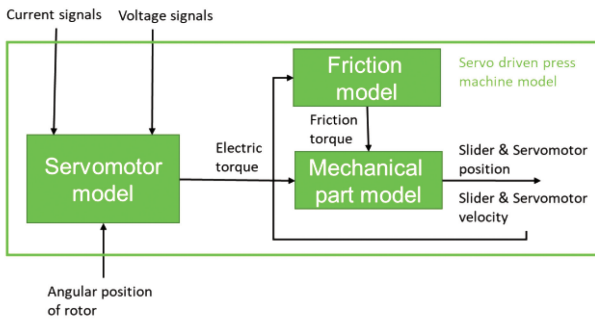

Figure 7.63 Servo driven press machine model diagram.

Where $M(\theta)$, $N(\theta)$ and $O(\theta)$ represent the inertia and mass of the gearbox, crank, connecting rod and the slider of the system. Respectively, τ_e is the electric torque, τ_{lb} is the load balancer torque and τ_{fric} is the torque exerted by the friction of the system. At this point, the *Prediction Error Method* based soft sensor is applied for estimating the aforementioned friction model coefficients, yielding a friction related torque that acts against the electric torque. For this purpose, unladen tests (without strokes) are carried out at different press speeds.

Once the model is fitted, the estimated angular position is plotted against the measured one, tracing a path similar to the measured one, as shown in Figure 7.64.

Estimation of the objectives

The estimation of the previously mentioned objectives is performed adding another stroke related torque to the Equation (7.1), which yields the Equation (7.6):

$$M(\theta)\,\ddot{\theta}(t) + N(\theta) + O(\theta) = \tau_e - \tau_{fric} - \tau_S \qquad (7.6)$$

where τ_S represents the torque generated by the slider stroke. This new term collects the applied force during a cycle, and thus, many metal forming processes are monitored through the measured signals, the model (Figure 7.65), and a soft sensor.

In order to estimate the states (position and velocity of the slider, process force) of the model, a step by step Bayesian soft sensor is used, which can perform real time estimations of the system states. On a first stage,
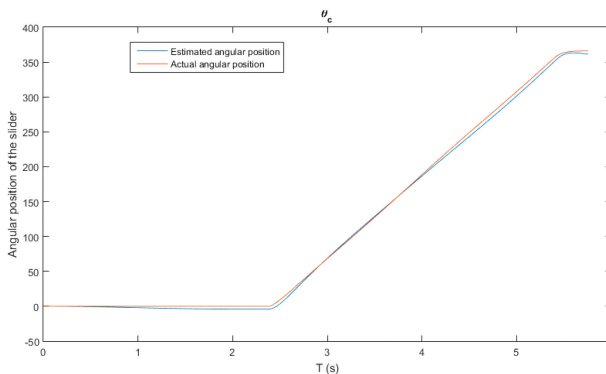
Figure 7.64 Estimated and actual angular positions.
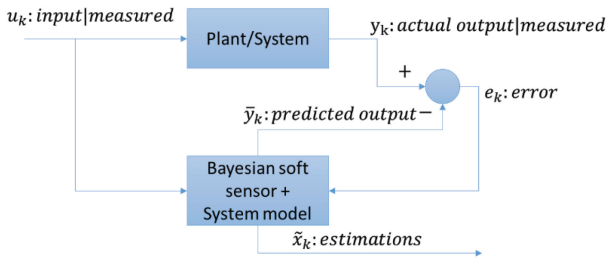
Figure 7.65   Step by step Bayesian soft sensor estimation performance diagram.

all the system inputs are recorded in advance and are used in a computer environment. The Bayesian soft sensor takes system inputs and measured outputs and is able to estimate unmeasured system states as shown in Figure 7.65.

The three estimation objectives of the use case are related to the slider stroke force.

7.3.3.6.2 *Results*
Preliminary results of measured electric magnitudes of the servomotor are discussed in this section. Some of those test results are displayed in Figures 7.66 and 7.67 where the relation between electric magnitudes of the servomotor and the applied ram force is shown.
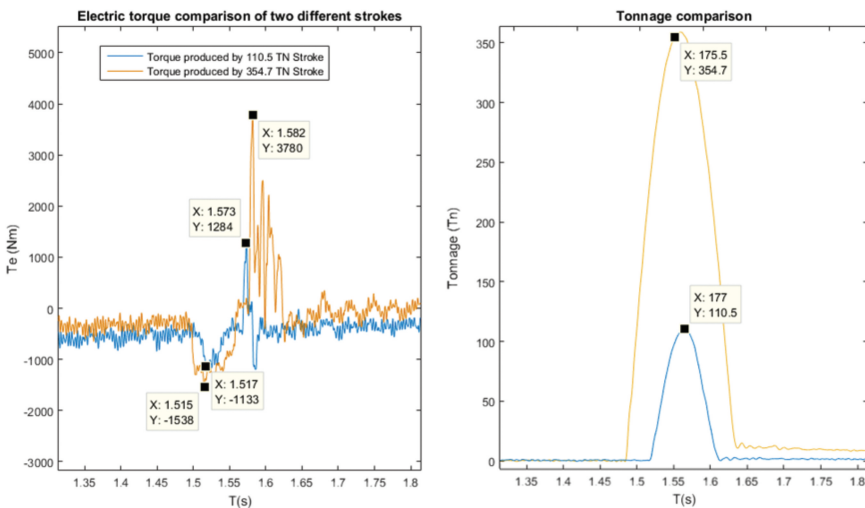
Figure 7.66   Estimated electric torque for different tonnage tests.
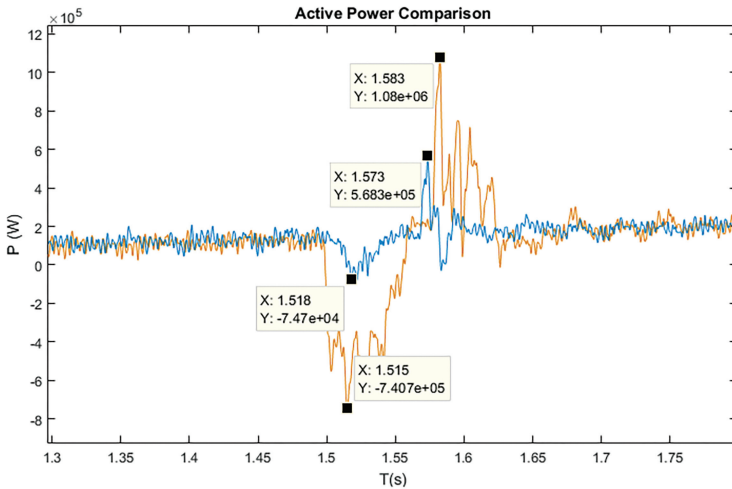
Figure 7.67 Calculated active power for different tonnage tests.

These tests reveal that the servomotor magnitudes (electric torque and active power) change accordingly with the applied process force, feature that is used for estimating the applied real force directly from measurement of the servomotor magnitudes.

Regarding the cutting shock effect, Figure 7.68 shows the estimated electric torque and its corresponding tonnage for two cutting shock tests. In the test where the narrowest metal sheet is cut, the cutting process finishes earlier comparing to the widest metal sheet cutting test and besides, the generated cutting shock compensation electric torque is smaller than in the other test. The image on the right side of Figure 7.68 depicts the applied force and the cutting shock effect quantification which are of 5 and 21 tons for the 300mm and 500mm wide metal sheets respectively.

Figure 7.69 displays the consumed active power by the PMSM. As in Figure 7.68, the cutting process and the cutting shock effect compensation shape the PMSM power consumption.

With respect to the unbalanced forces, tests have not revealed any difference looking at the analysed electric magnitudes between the balanced and unbalanced process. Figure 7.70 shows a similar electric torque for the unbalanced and balanced processes. Estimated electric torques for unbalanced and balanced processes are also unable to determine differences between the two processes, as displayed in Figure 7.71.

Figure 7.68   Estimated electric torque for 300mm and 500mm wide tough metal sheet cutting process.

Figure 7.69   Calculated active power for 300mm and 500mm wide tough metal sheet cutting process.

Figure 7.70 Calculated active power of centred and off-centred tests.

Figure 7.71 Estimated electric torque for a centred and off-centred tests.
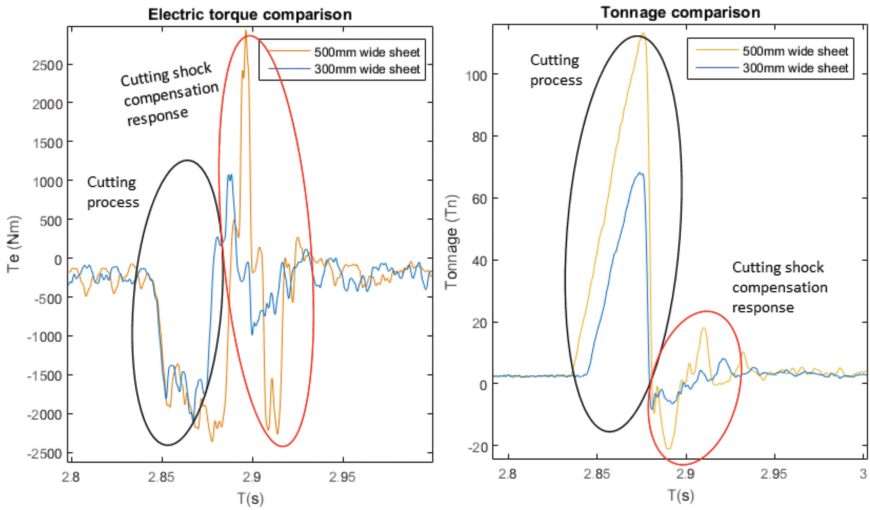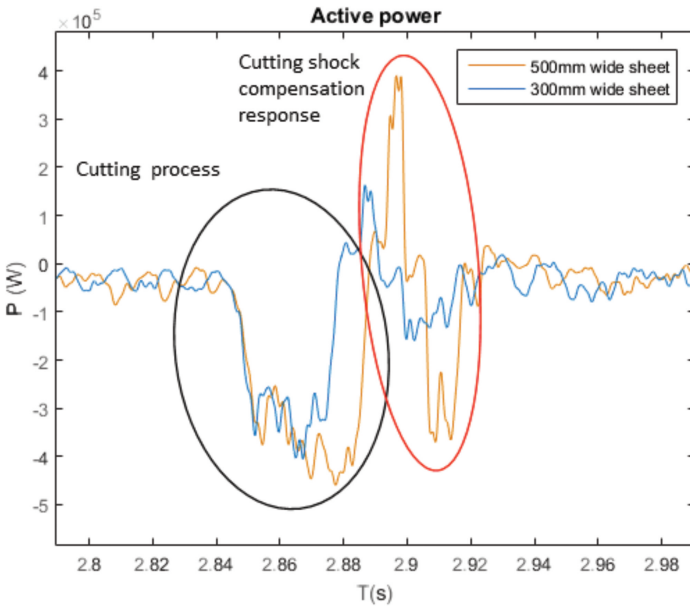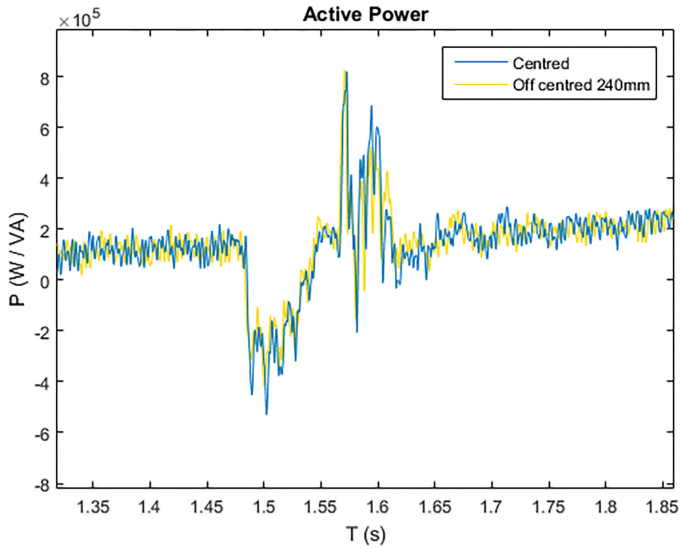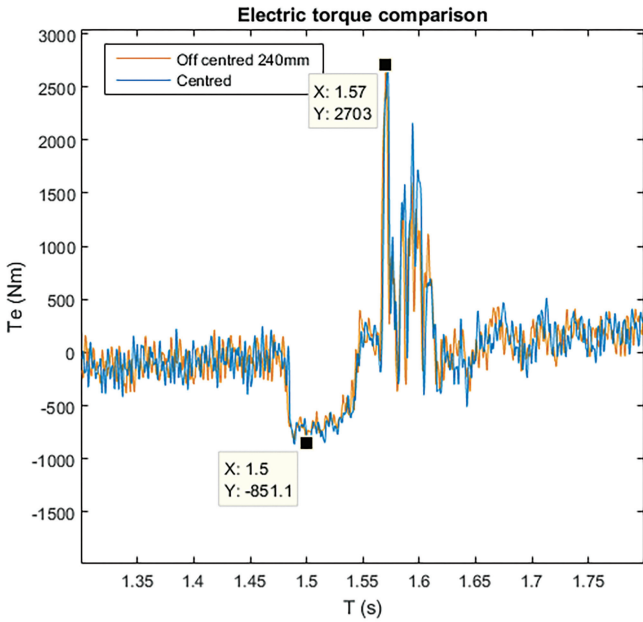
As future work, the soft sensor approach will be employed once the final test is done in order to measure indirectly (estimate) the applied ram force.

7.3.4 MANTIS Solutions for Clutch Brake

GOIZPER considers critical to increase machines and components reliability. To meet this challenge GOIZPER decided to incorporate cutting-edge technologies in their products as a means of enhancing product robustness and functionality in order to facilitate proactive-predictive maintenance activities.

GOIZPER decided to investigate different ways of sending data of its components to the Cloud. The main objective is to find the most robust and reliable architecture, and for this reason, two different data fluxes were developed in order to send the Clutch Brakes information to the cloud. One of the architectures is coordinated and developed by MGEP and the second architecture is coordinated by TEKNIKER. These two different approaches (at edge/sensor, and at platform level) are explained in detail in the next sections.

7.3.4.1 Maintenance cloud platform by MGEP

The platform presented in this section is concerned with analysing a clutch brake system and its components in press machines to detect the most important failure sources and be able to perform predictive maintenance in those press machines. Analysis techniques and algorithms, to be used on the assets data, were implemented in the platform with the aim to support predictive maintenance of clutch-brake. These technologies are (1) Root Cause Analysis powered by Attribute Oriented Induction Clustering and (2) Remaining Useful Life powered by Time Series Forecasting. The implementation of that platform was previously published in a conference paper [Larrinaga et al., 2018].

7.3.4.1.1 *Background*

The overall objective sought by GOIZPER is to early detect internal wear of a clutch-brake. To do that, the moving parts of the clutch-brake were sensorized. By continuously monitoring the system conditions proper operation of the clutch-brake can be ensured. Moreover, the most critical operating variables are registered in the platform in order to analyze the working process and prevent misuses. The data is uploaded enabling the holistic analysis of the clutch-brake system, with the aim to determine/detect the main causes of failure and the components' remaining useful life.

7.3.4.1.2 *Solution approach*

The architecture implementation agrees with the three-tier architecture presented in [Hegedus, 2018] and in Chapter 3. There are three levels: Edge tier, Platform tier and Enterprise tier. The **Edge Tier** is concerned with the technological solution deployed on the sites where the Press Machines (including the Clutch Brake component) are located. At this level, data acquisition systems to extract the data from the different sensors and SCADA systems connected to the machines are deployed. Figure 7.72 depicts the elements of this tier. An Industrial PLC based on the B&R X20CP1382 module was connected to the sensors attached to the Clutch-Brake (including the intelligent soft-sensors). The module collects all the measurements from the sensors and runs local code to pre-process the signals and produce a set of parameters that are able to characterize the overall status of the Clutch Brake. The module stores these parameters in a local file to act as a Datalogger for the cyber physical system. A second embedded computer (Edge Gateway) is attached to the Datalogger, and it retrieves the parameter files and creates IoT-A CEP Events [Internet of Things Architecture] that are sent to the cloud platform as messages.

At platform level, data coming from the different sources is persisted and different applications that allow analyzing of this data are available. The specific modules for the **Platform Tier** are presented in Figure 7.73, and are:

- **Edge Brokers:** It maintains the connection between the edge devices and edge tier, and it includes a data distributor. The distributor is a message-oriented component to collect and redistribute the in- and outbound messages between components. In this use case, this module is a publish/subscribe system that receives the data from edge tier in different queues and publish the message received to the modules in the
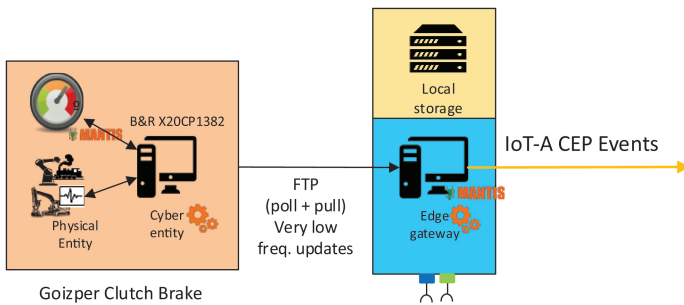
Figure 7.72   Edge Tier.
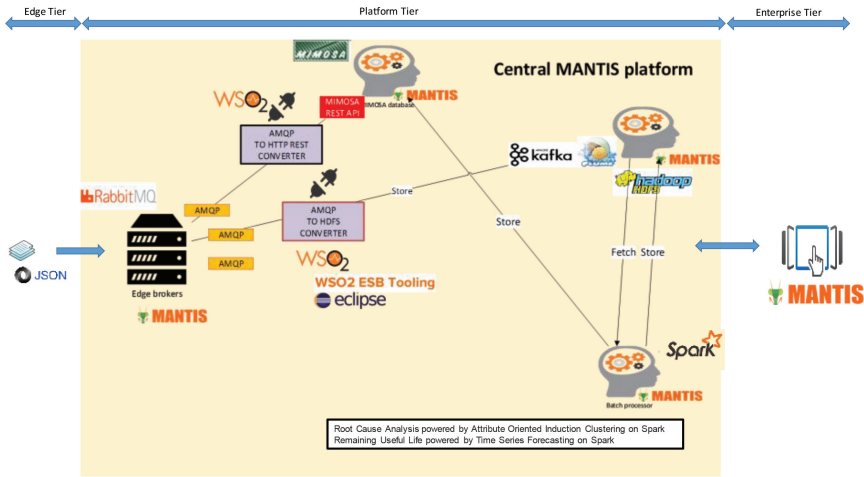
Figure 7.73   Platform Tier.

platform that have subscribed to each queue. The technological solution for the Edge Broker is RabbitMQ. RabbitMQ [RabbitMQ, 2018] is an open source message broker that supports multiple messaging protocols [Albano et al., 2015] such as Advanced Message Queuing Protocol (AMQP) [Vinoski, 2006];

- **Converters:** Software components or modules to translate edge cloud interface (IoTA CEP Events) into a database interface (MIMOSA API Rest [MIMOSA Consortium, 2003]) or files system interface (HDFS API). The converter is implemented using the translation capabilities of an Enterprise Service Bus (ESB) named WSO2 [WSO, 2018];
- **Data Storage systems** store the information coming from Cyber Physical Systems (CPS) and results obtained from data analysis maintenance actions and algorithms. Two storage systems are employed:
  - MIMOSA DB: This is a database compliant with the ISO-13374 Standard (Condition Monitoring and Diagnostic of Machines) [ISO, 2018]. One of the main objectives of the MIMOSA CBM architecture is to standardize the information flow between the various blocks, so that equipment from different vendors could be interoperable. The MIMOSA database is deployed in SQL Server and API REST is used to access data from applications;

- Hadoop Distributed File System (HDFS): This is a distributed file system designed to run on commodity hardware. Designed to be deployed on low-cost hardware, HDFS is highly fault-tolerant and provides high throughput access to application data, which makes it suitable for applications that have large data sets;

- **Batch Processing:** data analysis and processor mechanisms to enable the management of large volumes of data, fetched from storage systems and process on demand. This is implemented over Apache Spark [Apache, 2018]. The batch processor units implement the offline analytics capabilities of the platform. These technologies are (1) RCA powered by Attribute Oriented Induction Clustering and (2) RUL powered by Time Series Forecasting;

- **API or WS:** To interact with the Enterprise Tier an API offering services is provided. This component provides information and functionality (for example RUL) to components external to the platform such as HMI, applications (ERP) or even other platforms that lack certain maintenance algorithms.

The **enterprise level** is concerned with the applications that integrate information from one/several sites to enhance the global decision-making process using monitoring through Human Machine Interfaces (HMI) and data aggregation and analysis.

In relation to CBM-based PM the following aspects have been addressed for this scenario:

Equipment Failure Root Cause Analysis: The RCA is the first and necessary step to identify the main equipment failure causes. An AOI algorithm is used as the principal RCA algorithm. AOI is considered a hierarchical clustering algorithm, it is considered a rule-based concept hierarchy algorithm, and it was first proposed by [Han et al., 1992] Jiawei Han et al. as a method for knowledge discovery in databases. The representation of the knowledge is structured in different generalization-levels of the concept hierarchy with IF-THEN rules. The execution of the algorithm AOI follows an iterative process in which each variable (also referred as attribute) is generalized based on its own hierarchy-tree. This step is denoted as concept-tree ascension [Cheung et al., 1994]. To ensure the correct functioning of the algorithm, it is necessary to establish background knowledge, which specifies attribute generalization levels.

Equipment Remaining Useful Life estimation: The main objective of the RUL estimation process is to estimate the useful life of an asset before

a catastrophic failure occurs. The RUL estimation process is performed as a combination of AOI algorithm outcome and Auto Regressive Integrated Moving Average (ARIMA) statistical time series forecasting models. A common objective of Time Series Forecasting methods is to learn from previous data in order to be able to make predictions of future behaviours. In order to estimate the RUL, the first step is to evaluate a new variable to represent the machine behaviour correction factor, denoted as Normality Factor. The Normality Factor quantifies the extent of the damage of the machine. By applying ARIMA time series forecasting models, the Normality Factor evolution is modelled. As a final result, the Normality Factor model allows to predict the wear of the Normality Factor, providing the machine RUL in terms of clutch-brake cycles. Finally, clutch-brake cycles are translated into days, by combining the number of cycles the clutch-brake system does per day.

7.3.4.1.3 *Results*

Regarding the implementation of the reference architecture, a platform that accommodates different industrial processes and assets data for CBM analysis was built. The platform integrates an interoperable data model for CBM. Additionally a data/protocol converter that enables translations between most common data formats and protocols was developed.

Regarding data analysis, preliminary results performed as a proof of concept show the capability of the proposal. For the experiment, several features of the clutch-break machine have been used (trigger, angular position, application pressure, line pressure and flywheel speed). Once the knowledge-base has been created applying AOI and the most significant cluster-appearance order for the working cycles was calculated, the anomaly detection step is processed using the Normality Factor as threshold (value of 0.70). The Normality Factor evolution signal shown in Figure 7.74, is the result of applying ARIMA model over the training data utilized to generate the knowledge-base. In this experiment, around two hundred and fifty 'break' working cycles have been predicted. As it can be observed, there are five different work cycles cutting the established Normality Threshold; thus, it can be inferred that five different anomalies were detected. The next step is to analyze the characteristics of the anomalies, inspecting the reasons of their occurrence. For example, if there is any cluster in the abnormal work cycle that is not registered in the knowledge-base, it is recommendable to check the features or the grips of features in which the new values have occurred in order to establish the reason of the failure; if the order of the clusters inside

Normality Threshold: 0.70 — Value of Normality Factor of Working Cycles

Figure 7.74 showing evolution of the Normality Factor over work cycles, with y-axis "Percentage of correct cluster-appearance order in each work cycle" ranging 0,00 to 1,00, and x-axis "Work Cycles" from 1 to 246.
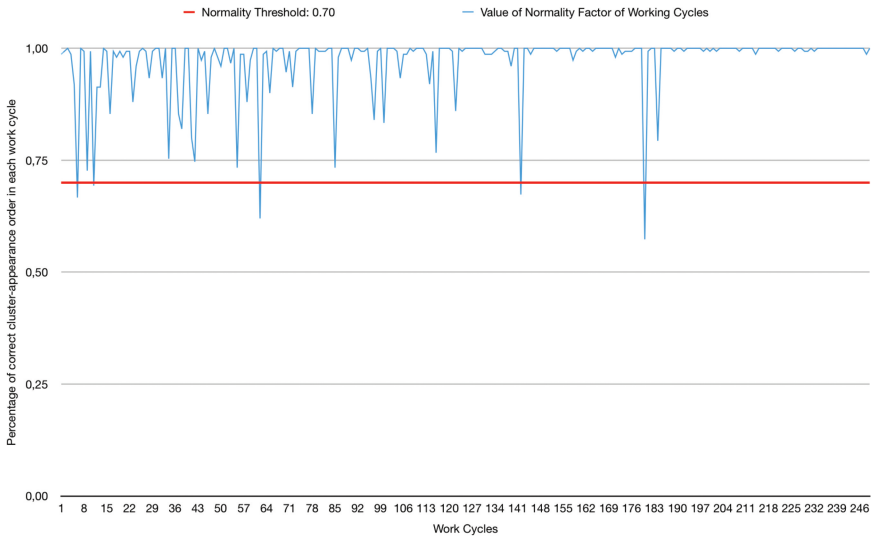
Figure 7.74   Evolution of the Normality Factor over work cycles.

the abnormal working cycle is significatively different respect to the ones registered in the knowledge-base, it can be reasonable to check the evolution of the values of the features in order to specify the reasons of the failure.

7.3.4.2 Maintenance cloud platform by Tekniker

As it is stated at the beginning of Section 7.3.4, the main objective of Tekniker's platform is the analysis of clutch-brake systems in order to detect failures. The platform supports Smart-G, a cyber-physical system that compiles critical process values and condition-related parameters, performs pre-processing based on algorithms specifically designed for this purpose, and offers a first level of monitoring and decision support directly back at the edge tier.

In addition, this valuable information recorded locally can be sent to the cloud platform where the user can access the entire historical information related to the use of the component. Therefore, the objective of this platform is to support the knowledge of GOIZPER and give predictive maintenance capabilities using different algorithms integrated in the system.

7.3.4.2.1 *Background*

The main objective is to understand clutch-brake wear in order to give services and advices to the customers. All the critical signals are acquired,

stored and processed in a PLC-based device called "Smart-G", and then sent to the cloud platform, where they are stored and processed to predict failures and give visual decision support capabilities. Therefore, information on the condition of machine components and operating processes is recorded locally on the Smart-G devices, and transmitted remotely to reduce unforeseen downtime and increase equipment availability.

7.3.4.2.2 *Solution approach*

The platform is built using the Microsoft Azure cloud services, and is represented in Figure 7.75.

The system is designed using a typical pattern in big data scenarios known as "lambda architecture", with uses three layers to solve the computing problem: speed layer, batch layer and serving layer. The batch layer holds an immutable, read-only master database, and it pre-computes a batch view with indicators and aggregated data. The speed layer deals with recent data only and executes quick algorithms (rules and machine learning algorithms) to produce a speed view with alarms and predictions. The serving layer is composed of the batch view and the speed view mentioned before.

Exploitation and visualization of data relies on the Microsoft Power Business Intelligence capabilities to show aggregated information, indicators and transient raw data of the monitored assets.

Application of big data techniques, combined with machine learning for pattern identification, and complex event processing for the detection in real
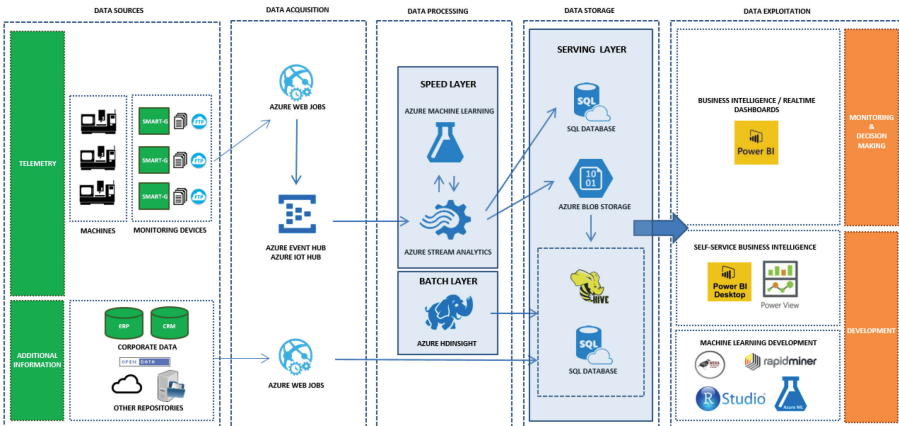

Figure 7.75   Azure-based Maintenance Cloud Platform.

time of the learned patterns, is the approach for reaching a high ratio of availability and operational performance.

The Azure services used to build the platform are described below:

- Azure Event Hubs: It is a highly scalable publish-subscribe message broker for event ingestion, with a partition-based approach to support the ingestion of millions of events per second;
- Azure Stream Analytics: This is an event data processing service for real-time analysis of streaming data. It uses a SQL-like language to create rules, and can sent requests to the Azure Machine Learning Service to execute algorithms in real time;
- Azure Machine Learning: It is a cloud service for the implementation of predictive analytical solutions. It provides a big number of built-in packages, and allows the customization of new ones;
- Azure HDInsight: This is a highly scalable solution used to prepare the data in the batch layer. It allows the combination of data and statistical equations providing many possibilities for data enrichment;
- Azure Blob Storage: It is a cloud service to store unstructured data as blobs. A variety of data files can be stored, for example binary, text, documents, multimedia files, etc.;
- Azure SQL Database: This is a relational database that is used to store alarms and aggregated values.

Additionally, a business intelligence tool (Microsoft Power Business Intelligence) is used for reporting and visualization of data. This tool facilitates the representation of information in attractive panels and reports that can be customized in a very flexible manner.

7.3.4.3 Friction material slippage

Two kind of slippages can arise during operation, clutch side slippages and brake side slippages. Each of them are caused by different reasons. Clutch side slippages cause a transmitted torque loss to the output shaft, which is dissipated as heat. Brake side slippages also cause a transmitted torque loss and in turn, a delay of the shaft's braking time.

7.3.4.3.1 *Solution approach*

In the proposed solution, many factors have been considered in order to identify slippages causes for each case. For the slippages that come up during clutching, air leakages and clutch side friction material degradation have been analysed. On the other hand, brake side slippages are produced due either to brake springs degradation or brake side friction material degradation.

Slippages can be detected directly from encoder velocity and acceleration signals as shown in Figures 7.76 and 7.77.

7.3.4.3.2 *Results*

Figure 7.76 shows velocity and acceleration profiles when no slippage is generated. As it is noticed, clutching and braking velocity and acceleration curves are continuous.

In Figure 7.77, one can notice an interruption in the velocity and acceleration rising curves generated during clutching. This interruption is provoked by slippages that have emerged due to the different reasons mentioned above.
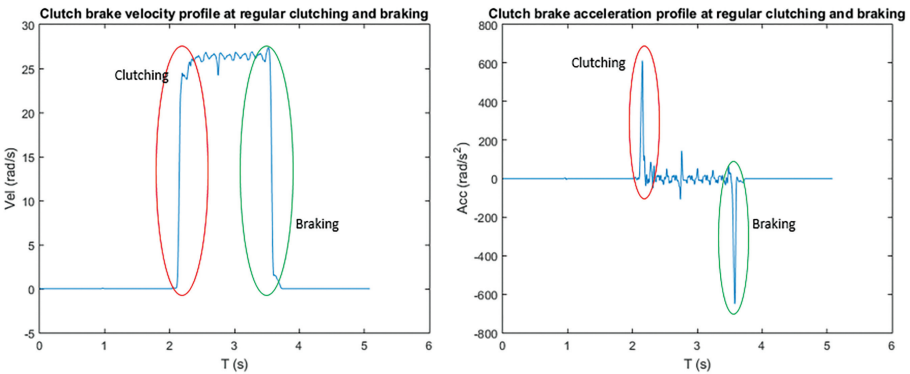
Figure 7.76 Clutching and braking velocity and acceleration without slippage.
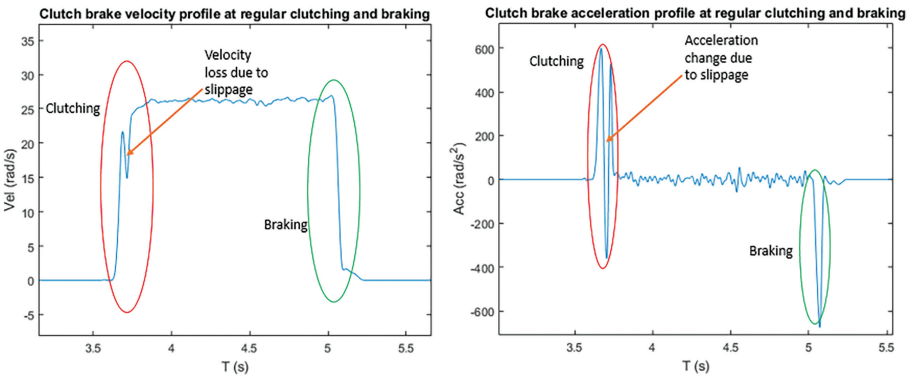
Figure 7.77 Clutching and braking velocity and acceleration with slippage at clutching.

7.3.4.4 Brake spring degradation

Brake springs degradation reduces braking torque, increasing braking time. This effect may put the integrity of the produced work pieces at risk, as well as operators' integrity.

7.3.4.4.1 *Solution approach*

A soft sensor approach is developed for estimating the brake springs stiffness. Several test were done with many brake spring combinations for simulating degradation. The set-up is shown in Figure 7.78. The pressure sensor P1 measured the line pressure, the pressure sensor P2 measured clutch brake input port pressure and the pressure sensor P3 measured the chamber pressure.

The developed soft sensor is able to estimate several clutch brake states and a parameter by measuring only the line pressure P1 and the clutch brake input port pressure P2. The estimated states were piston displacement, velocity and acceleration, and the inner chamber pressure evolution over time. The estimated parameter was the brake springs stiffness.

7.3.4.4.2 *Results*

For each test the inner brake springs were changed and the soft sensor is able to estimate the brake springs stiffness with an error less than 5%. The estimation results are depicted in Figure 7.79.

In the case of the estimated stiffness, the estimation line converges quite well with the actual value of the brake springs. The estimation of the chamber pressure does not converge so well due to the leakages that are not yet taken into account in the model.
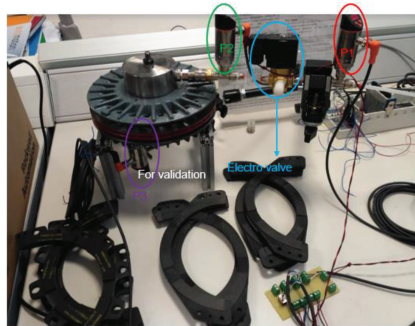
Figure 7.78   Prototyping clutch brake set-up.
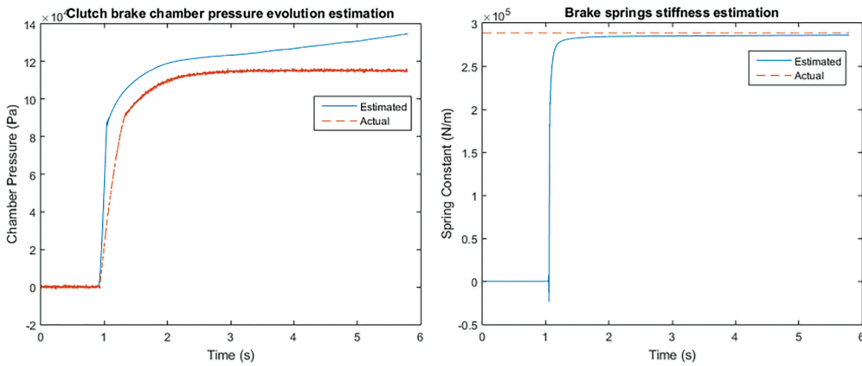
Figure 7.79   Estimated vs actual inner chamber pressure and brake springs stiffness.

7.3.4.5 Friction material wear

During operation, clutch and brake sides' friction material suffer from wear. These issues could cause some problems such as delays at braking and clutching and, at the same time, they could also cause the degradation of other components.

7.3.4.5.1 *Solution approach*

The wear of the clutch side and brake side friction material is monitored by means of a soft sensor that takes advantage of the already installed pressure sensors signals. Some metrics are defined in order to relate the air mass flow and the instantaneous pressure level with the wear of both friction materials. As in all cases where soft sensors have been applied, this solution needs a model of the analysed system.

Many tests have been carried out combining different wear levels for friction material for both sides. Analysed magnitudes or metrics have revealed a similar behaviour for an identical friction material wear in the same side, either in clutch side or brake side. Figure 7.80 shows air pressure vs air mass curves shapes for different friction material wear combinations. The percentages that appear in the figure legend represent the wear level of both sides, being left side percentage brake side wear and right side percentage clutch side's.

7.3.4.5.2 *Results*

Figure 7.81 shows a zoomed-in view of the brake and clutch related curve sections and shows how the curves track the same path for an identical wear level for both sides.
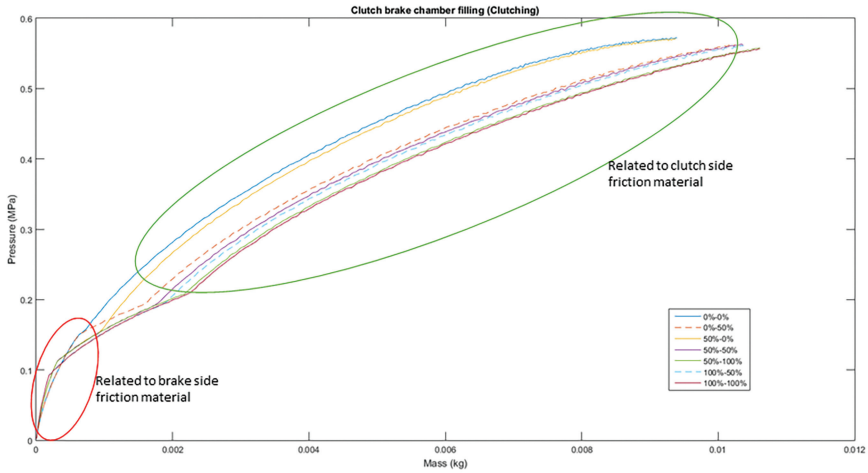
Figure 7.80 Air pressure vs air mass representation.
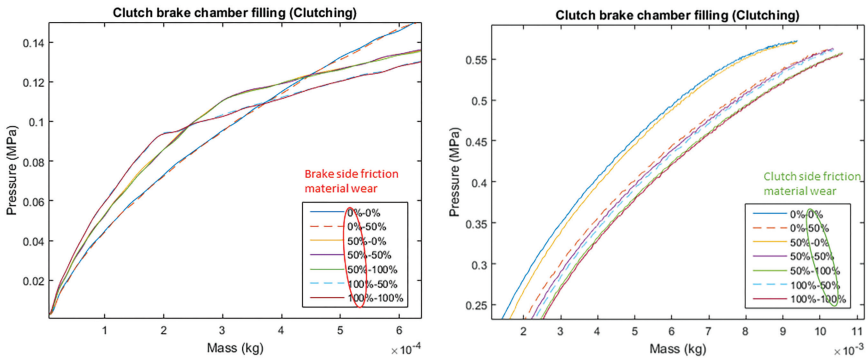
Figure 7.81 Left side, brake friction material related curve section. Right side, clutch friction material related curve section.

From those results, the soft sensor was able to estimate the wear level of the friction material attached to the friction discs, establishing some thresholds for different friction wear levels.

7.3.4.6 Piston chamber air leakage

Air leakages during a clutching operation imply an engaging force loss, which in turn is associated to economic losses since more compressed air must be provided to the clutch brake in order to compensate those leakages.

7.3.4.6.1 *Solution approach*

The air leakages are detected and measured by the air mass flowmeter. The mass flow of air reveals directly the air leakages during the press machine operation while the clutch brake is clutched.

7.3.4.6.2 *Results*

Figure 7.82 depicts the evolution of the air mass flow during a single stroke operation of the press machine. The portion surrounded in purple quantifies the air leakage that the whole system has experimented.

7.4 Fault Detection for Metal Benders

Contributors: Rafael Rocha, Michele Albano, Luis Lino Ferreira,
Hugo Ferreira, Catarina Félix, Carlos Soares, Goreti Marreiros,
Diogo Martinho, Isabel Praça, Giovanni Di Orio, Pedro Maló,
Asif Mohammed, Rui Casais

The objective of this use case is to apply anomaly detection algorithms to the data from the the CNC of metal benders and additional sensors in order to detect failures. The idea is that the machine tool will have an expected
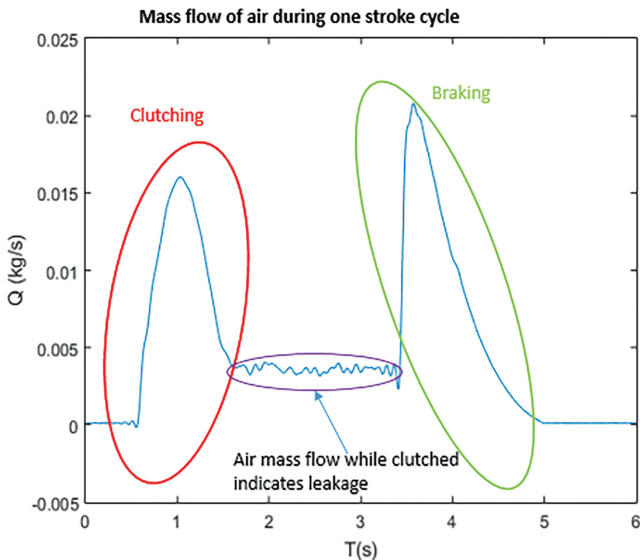
Figure 7.82   Air mass flow profile during a single stroke cycle.

behaviour given a set of environmental and production parameters. If at any point in time this behaviour deviates from the expected, then we can assume that something is wrong. In such cases, it is necessary to flag these states and warn both machine operator and maintenance personnel of this event. More concretely an analytics module that processes the data generates alerts or alarms according to the deviation detected from normal operation. These alerts are sent to a monitoring system were they can be viewed and analysed by the machine operators. This can save on downtime by both supporting the diagnosis of the potential malfunction, collecting spare parts in advance, and repairing the machine before the malfunction occurs [Ferreira et al., 2017].

7.4.1 Introduction to Press Braking

Press braking (brake forming) is the process of deforming a sheet of metal along an axis by pressing it between a clamp (tool) (Figure 7.83), performed by *metal sheet bender machines*, such as the one in Figure 7.84. A single sheet metal may be subject to a sequence of bends resulting in complex metal parts. Such operations can be used to produce a wide variety of products ranging from electrical lighting posts to metal cabinets.

In the case at hand, brake forming can bend sheet-metal (thus the common name of metal sheet benders) from 0.6 to 50 mm thick and lengths from 150 mm to 8 m long. The sheet metal bender machine considered in this section is a top of the line model and pertains to the Greenbender family [Ferreira et al., 2017], manufactured and commercialized by ADIRA. The machine (Figure 7.84) is able to exert a force up to 2200 kN using 2 electric motors of 7.5 kW each, and it is able to bend metal with high precision while saving a considerable amount of energy in the process, as per the EcoDesign (2005/32/CE) European directives.
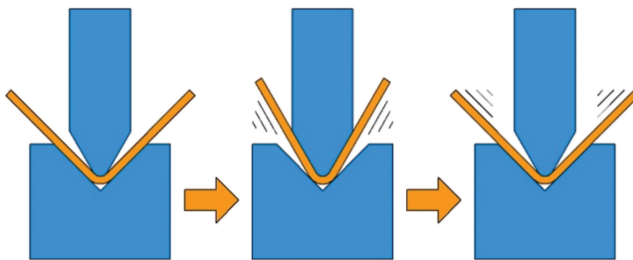
Figure 7.83   Bending process.

Figure 7.84   Sheet metal machine.

The angle and type of the bend are determined by the shape of the punch and die and the depth with which a punch penetrates a die. The dies can have "U", "V" or channel shapes. A movable ram is attached to the beam and is covered by a shroud. The punch is attached to the bottom of the ram and the die to the top of bed (covered by the lower shroud). When the ram descends on the table, a bending force is exerted on the sheet-metal between the punch and the die. The bending force and bending speed must be carefully controlled in order for the material being used to maintain their physical characteristics and insure the required bending precision. Additionally, the machine structure deforms due to the forces involved and those deformations have to be compensated in order to guarantee the machine precision.

Although this may seem simple enough, these machines require very accurate control to ensure the required bending precision (in the order of tens of microns). This accuracy is critical when the bending axis is long. The success of the operations depends on many variables including for example the tensile strength and thickness of the work piece, the type of tools (punch and die) used and the type of bend required.

To ensure the quality of the final product, the bending process comprises of several sophisticated control methods that include:

- Calculating the deformation of the workpiece based on the metals characteristics, tool geometry and the desired bend;

- Compensating for spring back by measuring the deviations and repeating the bending process;
- Compensating for frame deformation by measuring changes in the machine tools structure and adapting the pressing accordingly;
- Compensating for deflections in the bed by changing the shape of the bed during the pressing of the workpiece.

The machines used for this process consist of a hydraulic system that applies pressure to sheet metal thereby deforming the workpiece to the required specifications. The hydraulic and mechanical systems (pistons, valves, tubes, pumps) are subject to high pressures and may fail. Mechanical wear and tear also occur, which may result in damaged axis, shafts, bearings and tools (punch and dies). The hydraulic systems depend on the correct function of the electromechanical valves and motors. These elements are also subject to electrical failure. Many supplies are consumed periodically, such as hydraulic fluid, air filters and oil filters. These elements may also be subject to failure.

These machines have stringent safety requirements that also impose certain restriction on its operation. In addition to this, the production efficiency is also a very important factor in its operation. Moreover, since these machines are used for very costly manufacturing processes, downtime is extremely hurtful to the company that bought the machine. All these requirements mean that the various components such as the hydraulic system, tools (punch, die and bed) and back gauges, which are subject to extreme pressure, must operate in the best of conditions. It is therefore important to predict, detect and correct any failures that will either generate scrap, put an operator's life at risk, or cause downtime.

Next section describes the PM platform implemented to support this use case, while Section 7.4.3 provides insights regarding the employed data analysis techniques and their results.

7.4.2 Design & Implementation

Proactive maintenance strategies are implemented on the sheet metal bender machine by means of a distributed platform compliant with the MANTIS architecture described in chapter 3, and tailored to the work at hand. In particular, the focus of the platform is on sensor data acquisition, data transmission and storage, and forecasting and machine learning techniques.

The deployed platform is represented in Figure 7.85 and is described in the rest of this subsection. In particular, the components are put into relation with the three tiers the architecture is divided into (edge, platform, and enterprise).
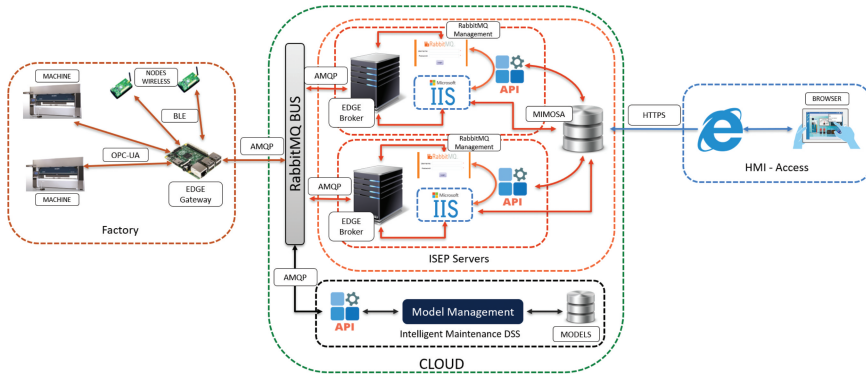
Figure 7.85   Design of the Proactive Maintenance solution.

7.4.2.1 Data collected by the machine's sensors

Data on the machine are collected by means of sensors that are part the machine's control systems or from sensors which were added specifically for maintenance purposes. In fact, the machines under study are advanced and heavily sensorized. The existing sensors can be grouped into three different data sources: the *Programmable Logic Controller* (PLC), the machine's *Computerized Numerical Controller* (CNC) that controls the machine, merging data from the PLC-connected sensors and actuators, and the *Safety PLC*.

The sensors of the PLC are used internally to control its operation. These range from buttons and pedals to advanced electric motor drives, with positioning information. Although used primarily for control functions, these sensors can also be used to determine anomalous events or states, to diagnose problems and even to infer the root cause of problems.

The PLC works in close cooperation with the CNC controlling all automation functionalities and, at the same time, it can send information from its sensors to the CNC. The Safety PLC handles only safety-related functions for the machine, such as preventing humans from being too close while the machine is working, detecting critical conditions, etc. Data from these sensors is mainly used to distinguish between component failures and safety-related events.

Data are collected indirectly from the CNC of the Green Bender Press Brake machine, which in turn collects information from the PLC control system of the machine and from its Safety PLC. An application on the CNC stores data regarding raised alarms, machine configuration and ERP-related information (e.g.,: production related data such as type of metal and

bend) on a Microsoft Access database. Note that the CNC is based on a Windows machine. The same application stores data collected from existing machine sensors (e.g.,: extensometers, pressure sensors, oil temperature and oil quality), which is collected from shared memory and to a file in the CNC filesystem. The information stored is then sent to the Edge Local node using the OPC-UA protocol. The ideal solution would be to access the shared memory directly by a single software module, but this solution was a compromise in order to ensure the safety and certification of the machine control system.

The application can be tailored and configured to different machines and applications, but the current pilot collects data from 50 machine sensors, with a periodicity of 20 ms, and from the MS Access database, which is scanned every second. The amount of data generated and transmitted to the cloud depends on the machine operation cycles. However, according to the data collected so far, we can extrapolate that it averages 300 MB per working day.

7.4.2.2 Wired nodes: The oil sensor

The application installed on the CNC also receives data directly from some of the sensors that were installed for maintenance-specific purposes and integrated with the PLC module. In particular, an oil sensor was the only wired sensors installed explicitly for Proactive Maintenance operations.

Oil condition sensors have the capability to detect ferrous particles, water, viscosity changes, etc., to detect lubricant related engine wear and lubricant quality degradation, among other problems. The installed sensor monitors the oil that lubricates the machine's hydraulic circuits, both in terms of its temperature and its quality, the latter being related to presence of contaminations like water, particles, glycol and other impurities in the oil.

The system that analyses the oil consists of two parts, the sensor unit (Hydac Sensor AS1008), and the data acquisition and computation board. The sensor reads temperature from –25 to 100°C, and saturation from 0% to 100%. Both signals are reported using a 4–20 mA interface. The data acquisition/computation module receives the signals, convert them, and exports the data through an analogic voltage signal with a range from 0V to 10V to the machine's CNC. The CNC digitalizes and sends the data through a communication middleware to the cloud for storage and processing, the latter being the comparison with custom thresholds.

7.4.2.3 Wireless nodes: The accelerometer

Each machine is equipped with two wireless accelerometer sensors, represented in Figure 7.86. The sensors monitor the blade that actually performs the bending of the metal sheet. The sensors collect the data from the own movement of the blade in the press, especially from the vibration patterns that are caused by the hydraulics. In fact, given the fact that the vibratory pattern can be associated to the condition of the machine's bending motors, the collected data can be used to perform PM of the machine.

The wireless protocol for the communication with the accelerometers is Bluetooth Low Energy (BLE), which enables data collection while maintaining energy consumption low. BLE is optimized for low power use at low data rates, and was designed to operate from simple lithium coin cell batteries.

The sensors are based on the Arduino 101 platform, which provide a 3-axis accelerometer with a maximum amplitude range of 8g. They are powered by two 9V batteries, and the sensors are configured for a lower measurement range (between 0 and 2g), aiming to attain a better accuracy. The sensors are able to perform self-calibration, synchronization and security, and the CurieBLE library is used to support communication between the sensors and the Edge Gateway by using of the Generic Attribute Profile (GATT). According to some preliminary experiments, the maximum distance for this technology is 30 meters, which corresponds with the BLE specifications.

7.4.2.4 Edge gateway

The Edge Gateway used for this deployment is located in the factory and it isolates the latter from the outside world, at the same time providing some functionalities at local level. From the security point of view, the Edge
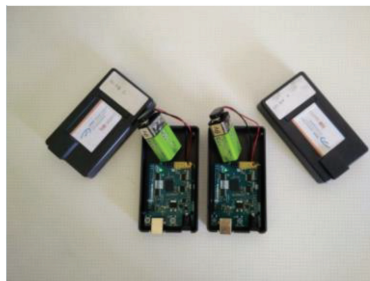

Figure 7.86   Sensor component hardware

gateway creates a DeMilitarized Zone (DMZ) in the sense that it is the only module in the factory premises that has network access, and thus concentrates all the security requirements on itself. Communication with the cloud is mediated by the AMQP protocol, and in particular through a RabbitMQ bus.

On the other hand, communication with the wireless sensors in a factory is done through BLE protocol, and the rest of the systems (CNCs of the machines and the wired sensors) is done by means of the OPC-UA protocols, which allows for a number of capabilities, including node discovery, data caching, and some degree of security.

Node discovery is used to enable the fast configuration of new machines in a factory. The CNC and the wireless sensors act as OPC-UA server and are discoverable by the Edge Gateway, which then provides the servers with mechanisms to support communication and management of the data acquired across multiple heterogeneous and distributed data sources. This is accomplished by providing an abstraction layer that detaches the application development from the intricacies of the lower level details. It acts as a virtualization platform and as data broker that connects the Machine logical block to the Cloud Middleware, capable of extracting, collecting, distributing/sharing, pre-processing, compressing, and semantically enhancing the data produced in an efficient manner. Therefore, the one of the fundamental goals of the Edge Gateway is – from one side is to support the data integration of multiple data sources and – from the other side is the provisioning of data to the cloud where more complex and resource consuming data processing takes place.

Finally, the Edge Gateway of this pilot comprises a database to cache collected data, and a local HMI service responsible for visualizing all the necessary information generated within the factory, such as list of machines available and their conditions, and data readouts.

7.4.2.5 Communication in the cloud

A few components on the cloud (Messaging Bus, Management Panel, Edge Broker and Database) manage the data, by storing and transporting them between the Edge Gateways of the factories and the Data Analysis and HMI modules. The communication mechanisms are implemented on top of a message-oriented bus and allow the interaction of the factories, mediated by their Edge Gateway, with the rest of the Platform tier, and the Enterprise tier. Communication is performed on top of a RabbitMQ bus, which is the most popular implementation for the AMQP protocol.

The basic elements of the message distribution system are the exchange and the queue. The exchange is the recipient of a message from a message producer, and its duty is to deliver the message to one or more queues, the latter being buffers from which the message consumers will pull the messages. An exchange can be connected to multiple queues, and the exchange can be configured to treat messages in different ways, such as relaying the messages to the queues in a round-robin fashion or broadcasting the messages to all the queues. Finally, the decision on which queue(s) receives each message from the exchange, is done by means of a routing key, which is a meta-datum assigned to each message. The messaging system can also implement Remote Procedure Calls (RPC) mechanisms, see for example Figure 7.87.

In the pilot at hand, Edge Gateways send data to the RabbitMQ server, where the routing process is used to deliver specific messages to the other components, and in particular the Edge Broker and the Data Analysis components. The RabbitMQ Bus is configured using a RabbitMQ Management panel (which is part of the Enterprise tier as far as the MANTIS reference architecture is concerned) and that obeys the REST architectural pattern. The component respects the reactive programming properties, namely, Responsive, Resilient, Elastic, and Message Driven. For example, the RabbitMQ platform is fault tolerant since, if a message delivery fails, the queue buffers the messages and retransmits them when the message consumer is back online. Moreover, if the broker malfunctions, messages in the queues are not lost since they are saved in the persistent memory of the broker.
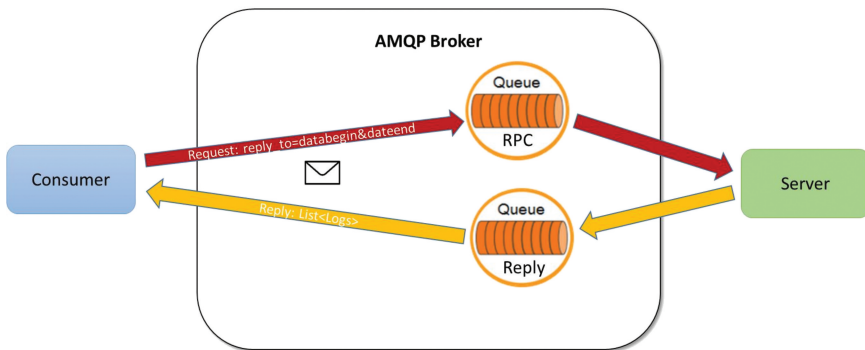

Figure 7.87   Consumer Remote procedure call (RPC).

The Edge Broker is the main peer that receives messages from several Edge Gateway devices through the queues and saves the data to a database module, which is structured according to the MIMOSA standard, which is described in Chapter 3. Current implementation of the DB is based on Microsoft IIS, even though an alternative design based on No-SQL Mongo is available. The RabbitMQ Bus provides queues for generic RPC connectors for handling database queries, to expose stored data to other components. The HMI is allowed to make queries trough RPC patterns to the database. Anyway, when data from a device is received in the middleware, the Data System Module component and HMI component will receive asynchronous messages.

7.4.2.6 Components for data analysis

The Data Analysis techniques and results for this pilot are described in next subsection, while this subsection is focused on the implemented components that support the techniques.

The main component for Data Analysis is the *Intelligent Maintenance Decision Support System*, which is used to manage the models (model generation, selection, training and testing), for example on reception of training data. The *Intelligent Maintenance Decision Support System* is composed of a Knowledge Base that uses diagnosis and prediction models and the data sent by sensors. On top of this Knowledge Base there will be a Rule based Reasoning Engine which includes all the rules that are necessary to deduce new knowledge that helps the maintenance crew to diagnose failures.

In addition to the data and algorithms, expert knowledge has been encoded as a set of rules that are used to detect and flag possible failures. Each rule indicates what sensor and CNC signals need to be acquired, how they are segmented, the type of analysis to be executed and what failure is associated with these signals.

As an example, let us consider when the brake press is working in automatic mode, terminates its bend cycle and has parked the ram on the top position waiting for the next task. If no failure exists then the ram must remain still in the same position where it stopped. Because the hydraulic system is constantly losing pressure, the CNC compensates for any deviation. Normally, such deviations are minor (imperceptible to the naked eye) and occur at very low rates. However, if a hydraulic pump fails or a hydraulics tube ruptures, leaks will cause large deviations as the CNC compensates for this.

In order to detect such problem, the positions of the pistons are recorded when the control signal indicates that the ram is at top dead center (segmentation). Statistical tests are used to check that the deviation is within a specific tolerance threshold. This threshold is determined via the machine learning algorithm (stream based) and is tweaked in order to reduce the false positive and negative detection rates.

7.4.2.7 Human machine interface

This module is part of the Enterprise tier, and it provides a Human interface for the proactive maintenance system. The HMI follows a web-oriented design and therefore can be accessed from anywhere, at any time and through all sort of electronic devices with the only requirement being the use of the Internet to do so. This allows both remote (administrative) and on-site operations such as analyzing the machine's state or view its past performance. It has two main modes, one for data visualization and another for data management.

In the visualization mode, it is possible to view historical and live data, which is collected from specific machine sensors (e.g., machine status, speed, positioning and pedals state). It is also possible to show the results generated by the data analysis module, more specifically the alarms for unusual sensor data and the warnings regarding impending failures. It is possible to match the warnings from the Data Analysis block with historical data collected from the sensors. All data retrieved from the machine and the Data Analysis logical block can be inspected through graphs and tables, provided by the HMI module (see Figure 7.88). Also, each user can be promptly notified of any event on the machine through a text notification.

The HMI also displays the results of data aggregation and calculation of statistics. Several descriptive statistics provide useful (albeit simple) indicators that support the decisions making by those responsible for maintenance and design. These indicators are therefore available to the Data Analysts. The results of the machine learning algorithms are displayed when they generate alerts and alarms, but they can also be visualized as historical data.

The Management mode allows for all the administrative operations, like users and roles management, as well as factories and machines setup. Role management allows to dynamically assign specific permissions to each type of user, which can be Operator, Data Analyst, and Maintenance Manager.

Figure 7.88 Checking the "Bend" variable on a GreenBender machine.

Operator can view historical and live data only. The Data Analyst role allows inspecting live streamed data collected from the machine, such as oil-flow and temperature sensors. The data is displayed in near real-time (Figure 7.89). The Data Analyst role also allows the visualization of historical data by selecting the data variables to be shown as well as the desired time-frame. The Maintenance Manager role allows to view some statistics, e.g., the type of components substituted and the frequency rate of the replacements. This should be specified for each monitored parameter according to the current number of cycles performed and to the maintenance actions of the machine tools. The user can also choose to display the results of the Data Analysis logical block, in the form of alarms, alerts and reports, which are displayed highlighting the relevant information for the maintenance manager and allowing the consultation of details on their provenance. These notifications include alarms that indicate unusual sensor data (for example based on simple statistics) and, unexpected behavior (for example, using outlier detection algorithms). These notifications only allow the detection of failures (corrective maintenance), but in the future may also be used to plan preventive maintenance tasks.

Security is implemented by means of SSL/TLS, for both the communication with the Cloud Middleware and the access to the HMI webpage (HTTPS). It is also important to note that the web-based HMI is running in the same node as Cloud Middleware, in order to reduce system complexity and to be able to access the same Database.
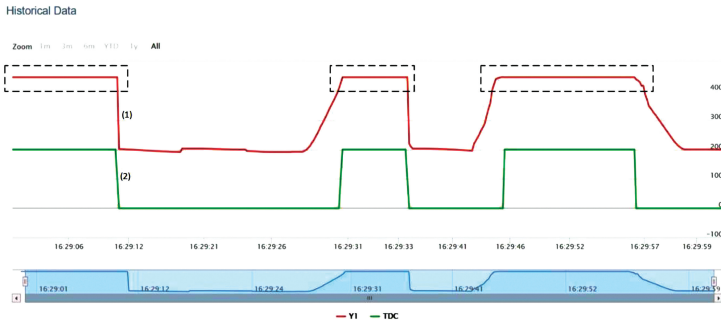
Figure 7.89   Example data from the controller.

7.4.3 Data Analysis

The analysis on data collected in the pilot is limited to a set of signals that can be divided into 3 groups: a) movements of the mechanical parts via hydraulic or electric power, b) temperature of the hydraulic fluid and c) extensometers that measure the deformation of the press brake machine structure. Additional signals are also available such as error codes and from the machine's soft numerical controller and alarms generated by the safety system. However these are not processed statistically and are only forwarded as alerts to a monitoring application.

The machine's designers and maintenance engineers determine the signals that need to be analysed. They are compiled as a set of rules that indicate not only the signals that must be monitored, but also when these signals should be collected and analysed and how the testing should be performed.

The pilot applies anomaly detection algorithms to the data from the machine tools controllers and sensors in order to detect failures. The idea is that the machine tool will have an expected behaviour given a set of environmental and production parameters. If at any point in time this behaviour deviates from the expected, then we can assume that something is wrong. In such cases, it is necessary to flag these states and warn both machine operator and maintenance personnel of this event. More concretely an analytics module that processes the data generates alerts or alarms according to the deviation detected from normal operation. These alerts are sent to a monitoring system were they can be viewed and analysed by the machine operators. Anomaly detection done in this work uses simple statistical testing (See for example Figure 7.89). The control signals are

those signals generated by the machine controller that are used to activate components and/or inform of the various process phases. Examples include: the *Top Dead Centre* signal that is active when the press ram is parked at the top position, the pedal signals (indicating if the operator pressed the up or down pedals), the axis in position signal (indicates when the backgauge is in position), the bending signal (that indicates if the ram is deforming the workpiece in manual or automatic mode), the speed change points (when the ram slows down for a press), the pinch points (indicating the start and end of the press phase).

Domain knowledge was elicited from brake press machine experts, which was then encoded as a set of rules that allow for the detection of possible failures. These rules indicate how to use the control signals to sample the process signals, thereby sampling data only during valid periods of time. It also substantially reduces the amount of data that is processed and analysed by the analytics module (for example, certain signals need not be collected when the machine is in standby or parked).

The machine tool manufacturer has provided a list of components that may be the cause of the failure for a given rule. For example, any deviations in the ram's position when parked may be indicative of a failure in hydraulic system. This list includes leaking tubes or oil sump (due to ruptures), malfunctioning valves, broken oil pumps and clogged tubes or pumps. Note that this information does not indicate the root cause of the failure. For example, a ruptured tube may be due to the hydraulic fluid being below the required temperature or a broken pump may have either electrical or mechanical malfunctions in other subcomponents. Currently this information is not used but could conceivably be combined with the alerts to help in diagnosing problems.

7.4.3.1 Data pre-processing

Data related to the machine's behaviour are received from the machine in chunks and are sent to the Data Analytics module, which has two functions. The first is the off-line learning and tweaking of the statistical anomaly detection models and the second is the on-line use of the models to detect failures and generate alerts. The combination of the alerts generated by the analytics module and the signal data collected from the machine tool facilitate the diagnosis of failures by machine operators and maintenance personnel (data selection and visualization).

An initial pre-processing phase of data processing will first segment and collect only those signals that, according to the rules we have previously

described, are required. For example, only the movement data is collected for the "parking rule" when the *Top Dead Centre* signal is on or we only analyse the ram speeds and synchronization when the pedal down control signal is on.

A second pre-processing phase will transform data in order to be able to either learn the statistics or use the statistical tests to check for anomalies. These transformations depend on the type of test to perform, which is determined by the rule. For example, this can involve calculating the difference between the two signals of piston distances when testing for synchronization between those pistons.

A third transformation of the signals is the calculation of the temporal difference of the piston displacement signals. This is used in the rules that use speed as a basis for comparison and checking. Note that usually such difference introduces significant noise into the signal and this may require additional filtering (smoothing).

From the experiments, it appears that the signals from the numeric controller and the oil quality sensor are clean. Even though the accelerometer and extensometers data are relatively noisy, no additional sophisticated pre-processing was used because the related failures could not be evaluated.

7.4.3.2 Failure detection

Statistical hypothesis testing allows one to compare two processes by comparing the distributions generated by the random variables that describe those processes [Stuart et al., 1999]. If the distributions are not equivalent, then we assume a failure occurred. A p-value is used as a threshold in order to detect any deviations from the expected process with the goal of reducing the number of false positives and negatives. In the case of the formal statistical tests, if the p-value does not allow us to reject the null hypothesis we cannot infer that the machine has no failure (type 2 error). However, here we assume that this is true, and alerts are only sent when the null hypothesis is successfully rejected. A distribution of a given process may be described by one or more random variables. Here we limited our analysis to the use of univariate statistics only using both parametric and non-parametric models.

7.4.3.2.1 *Parametric models*

In the case of the parametric models two basic tests were performed: univariate signal that should be close to a constant (within an unknown threshold), or two signals must not diverge from each other (no more than an unknown threshold). In both cases we can use the signals that indicate velocity, distance, heat and acceleration. In either case, if we can perform a

parametric Gaussian test on the mean (using a t-Test) or deviation (directly compare deviations, F-test, Bonett's test and Levene's test).

In addition to the tests described above, additional naïve statistical tests were used. In both cases an online Gaussian model is obtained via the calculation of a mean and variance. These means and variances are then directly compared to the continually sampled signals from a working machine. If a significant divergence is found, alerts are generated. Due to the high false positive and false negative rates (type 1 and type 2 respectively), an additional multiplicative threshold (in respect to one standard deviation) is used when comparing deviations. The initial values of this threshold are set automatically by selecting the lowest possible threshold that reduces type 1 errors.

7.4.3.2.2 *Non-parametric models*

As with the case of parametric statistic, the process of defining the hypothesis, sampling data and establishing a significance level as a threshold were the same. However, not all of the non-parametric methods provide a p-value for a significance level comparison. Three types of statistical tests were used in our work: the Kolmogorov-Smirnov test (K-S test), the Mann-Whitney U test and the use of a kernel density estimator (KDE). In the case of the U statistic, which is approximated by a normal distribution, a p-value is available to establish a threshold for accepting or rejecting the null hypothesis.

In the case of the KDE, an online algorithm was used that generates a dynamic number of (Guassian) kernels. The kernels and respective parameters of two different distributions cannot be directly compared. Experimentation shows that the estimated densities may be visually very similar, but the kernels themselves differ significantly. However, because we can use the kernel to sample the underlying estimated distribution we used the parametric statistical tests to compare the samples (both for the parametric cases and non-parametric cases using the Kolmogorov-Smirnov test and Mann-Whitney U test respectively). Here we could have also opted for the use of alternative algorithms such as the earth movers distance but did not do so because the naïve parametric tests seemed to be working well [Levina and Bickel, 2001] (see Section 7.4.3.2.1).

7.4.3.2.3 *Evaluation and interpretation*

For anomaly detection the positive labels are indicative of failures. Expected failure rates (as reported by the machine tool manufacturer) are very low. We therefore need to deal with data-sets that will be highly skewed (very few positive labels). This brings with it two challenges.

The first is that false negatives are more important than the false positives. This is due to the fact that if we incorrectly predict that the machine is failing (false positive), the alert/alarm message will be sent and the operator, which will verify that the machine is not in fact failing. On the other hand, if we incorrectly predict the machine is not failing (false negative) it could have serious consequences, since the operator will not receive any warning message.

The second challenge is related to the selection of the appropriate metrics used in the evaluation of the model's performance. Accuracy is not a viable metric because a biased prediction of no failure will always result in high accuracy values. We considered the following metrics: the AUC-ROC (Area under the Curve for the ROC Curve) and a set of relations involving a combination of true positive (TP), true negative (TN) and false positive (FP) and false negative (FN) counts:

$$\text{Precision, } Prec = \frac{TP}{TP + FP} \tag{7.7}$$

$$\text{Recall, } Rec = \frac{TP}{TP + FN} \tag{7.8}$$

$$\text{Accuracy, } Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{7.9}$$

F0.5, F1 and F2 measures (F_β for $\beta \in \{0.5, 1, 2\}$), $F_\beta = \frac{(1+\beta^2) \times (Prec \times Rec)}{(\beta^2 \times Prec + Rec)}$

Mathews correlation coefficient,

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7.10}$$

Exploratory work was done using precision, accuracy and the MCC metrics. Final comparisons were done using MMC. Future work will consider the metric F_β and how to establish an appropriate value of β.

Due to lack of data, initial exploratory work used artificial (synthetic) data. This data was generated using normal (Gaussian), Bimodal (composed of two joined normal distributions), Pareto and Weibull distributions because we had no way of checking the signals distributions. We assumed a failure rate of 5%. Each of these distributions was tested for differences in the mean and/or standard deviation using both the parametric and non-parametric statistics referred to above (T-test, Kolmogorov-Smirnov Test, Mann-Whitney U test). We also generated non-parametric statistical models of the synthetic data using the Kernel Density Estimation (KDE). We then

used this model to generate a reference distribution that was used in the statistical test. In addition to this we also modelled the data as a Gaussian distribution and naively compared the means and deviations to determine if the processes are different.

We found that statistical methods tend to be very brittle and usually result in many false positives. Increasing the threshold would result in an unreasonably high rate of false positives. The reason is that the formal statistical tests make important assumptions about the distributions. For example, the t-Test is only valid if the distributions have the same standard deviation. The best results were obtained using the Mann-Whitney U test and the naïve Gaussian tests.

In the next phase of the work we opted to use the naïve Gaussian tests because it allowed for the easy generation and update of the model. The data we got from the machine seems to be Gaussian (we say seems to be because at the time of writing, data with failures had not been obtained and cannot therefore confirm this). During this period we collected data from a single machine tool executing a preprogramed sequence of operations under optimal conditions for several days.

The pre-processing steps described above are applied to this data (each segment collected for a given rule) and the mean and standard deviation are calculated using a robust algorithm. The Gaussian model is initially generated using the first 10% of the data stream. During the calibration we use the same 10% of the data set to establish the threshold so that no false positives are detected. The threshold indicates by how much a mean or standard deviation must differ from the base models' mean or standard in order to flag a failure. This is not an appropriate way of setting this threshold, but because no failure data exists, using a ROC curve to establish a good compromise between false positives and negatives is not possible. We then tested the failure detection models of each rule calculating the Gaussian parameters of the segmented data and comparing those parameters to the base model. We expected a false positive rate to be close to 0 but got the results in Table 7.2.

As referred above, a multiplicative factor can be tweaked to increase or decrease the false positives rates, but there is no way to measure the false negatives and hence use the MCC metric (Equation (7.3)). However, these test serve as an important sanity check and allow us to detect and correct several issues.

The first issue is that there is a lag between the control and measured signals (due to sampling delays and (mostly) due to mechanical inertia). This means that, for example, when a control signal indicated the start or stop of the ram, the corresponding sensor reading of the ram displacement does

Table 7.2 True negatives (OK) and false positives (ALERT, ALARM) detected

rule	trained	OK	ALERT	ALARM
1	112	1004	0	7
2	83	749	0	7
3	112	1011	0	0
4	146	1310	0	5
5	187	1631	0	152
6	187	1631	0	52
7	112	1010	0	0
8	90	813	0	0
9	120	1027	53	0
101	152	1368	0	1
102	95	855	0	3
11	117	0	0	1058
12	8109	8109	0	0
13	14938	74690	0	0
14	95	734	0	124
15	184	1665	0	0

not show up immediately. This also means that the control signals are not perfectly aligned with the ram movement (for example ram movement occurs after a stop signal is sent). To solve this, in certain rules, only the final signal samples are used to generate and compare the models.

Another issue is that oil temperature varies widely during the operation of the machine. This depends not only on the load, but also on the rate of the ram movement and the environmental temperature. In addition to this, several brake press machines have a heating element to warm up the oil to acceptable operational levels. The only way to truly solve this is to increase the sampling population to several machine tools operating in very diverse conditions. More important however is the fact that the machines' oil temperature will initially rise significantly compared to its initial operation. This means that we cannot limit our sampling of the oil temperature (or any other signal) to the start of the data stream when generating the models.

7.4.4 Conclusions

The pilot described in this section is used to experiment with building a PM platform able to collect data in an effective and efficient manner from a metal sheet bender machine, and with machine learning techniques applied to collected data to find misbehaviors.

The implemented platform is used to collect data from a single machine in a factory, and transport data to a cloud for processing. It is straightforward

to align all of the pre-existing and new components with concepts from the MANTIS reference architecture, and the design of the platform is used to validate the solution described in Chapter 3. The platform uses different protocols in the different tiers of the platform, and in particular OPC-UA in the Edge tier, AMQP between the Edge Gateway and the Edge Broker and between the components in the Platform tier, and HTTPS for the components in the Enterprise tier since these latter components are web-based.

A naïve Gaussian model is used to identify failures in a brake press machine using signal readings of position, speed, temperature, and acceleration. The type of failures and the respective signal analysis is selected and encoded by domain experts. Due to a lack of failure data (positive labels) it is not possible to evaluate the effectiveness of the models using standard metrics such as the F score and MCC (Equation (7.3)). On the other hand, it is possible to partially validate the solution based on the measure of false positives. The results seems to indicate that for this specific case, the naïve Gaussian model may be a viable solution. We are able to determine how to implement and test the experts' checks by applying a simple set of pre-processing steps and using Gaussian means and standard deviations. More importantly, it enabled us to identify and resolve some issues regarding the sampling and use of the signals (delayed signals due to inertia, time series with very high variability of the oil temperature).

7.5 Off-road and Special Purpose Vehicles

Contributors: Ansgar Bergmann

Off-road and special purpose vehicles include lorries (trucks), buses, agricultural machinery, construction machinery, and forklift trucks. These types of vehicles share many common characteristics, and offer the possibility of the development of related technical solutions and technologies under technically similar challenges. In addition to the property as an investment and a working machine in production and value-added processes, there are facts like high complexity, low-volume, high variety and high quality and reliability requirements over a long lifetime.

7.5.1 Introduction to the Use Case on Vehicles

STILL supplies customized internal logistics solutions and implements the intelligent management of material handling equipment, software and

services worldwide. With over 7000 employees, four production facilities, 14 branches in Germany and 20 international subsidiaries as well as a global dealer network, STILL is a successful international player. Today and in the future, STILL fulfils the requirements of small, medium-sized and large companies with highest quality, reliability and innovative technology. STILL's forklift trucks are operating in a variety of areas and conditions with often totally different application profiles, which differ not only in the temporal use (one to multi shift) as also in the environmental conditions (from easy hall operation up to use in the heavy industry or fishing industry). This results in high demands concerning ensuring the availability of vehicles, recognition of special types of damage and an optimized maintenance scenario, which fits to the special needs of this types of usage and thereby minimizes potential and also safety-critical damages resulting from this.

STILL currently collects its data only for internal processes and customer applications. A key problem lies in the mobility of the machines, which does not allow large amounts of data to be transferred on a wireless way without high costs. In addition, the systems are operated in a wide variety of environmental conditions. Within this project, the aim is to determine whether existing data collection mechanisms are already sufficient for the desired objectives or whether new solutions have to be chosen, to be an enabler for new service solutions and other maintenance based products. Basing on these fundamental analyses, options for business actions can be derived.

Smart services (see Figure 7.90) are interesting for a company as they:

- Enable higher added value (optimization of service in combination with intelligent products) and better service quality, e.g., through shorter reaction and repair times;
- Increase user-friendliness;
- Open up new markets for services and data-driven business models;
- Enable a drastically increased efficiency of service-based operating models;
- Result in higher machine availability;
- Guarantee a more detailed planning of machine operation and downtime times;
- Improve component design through monitoring.

7.5.2 Scope and Logic

Currently, STILL has established systems to support all standard issues of maintenance, where actions and processes are mainly based on existing
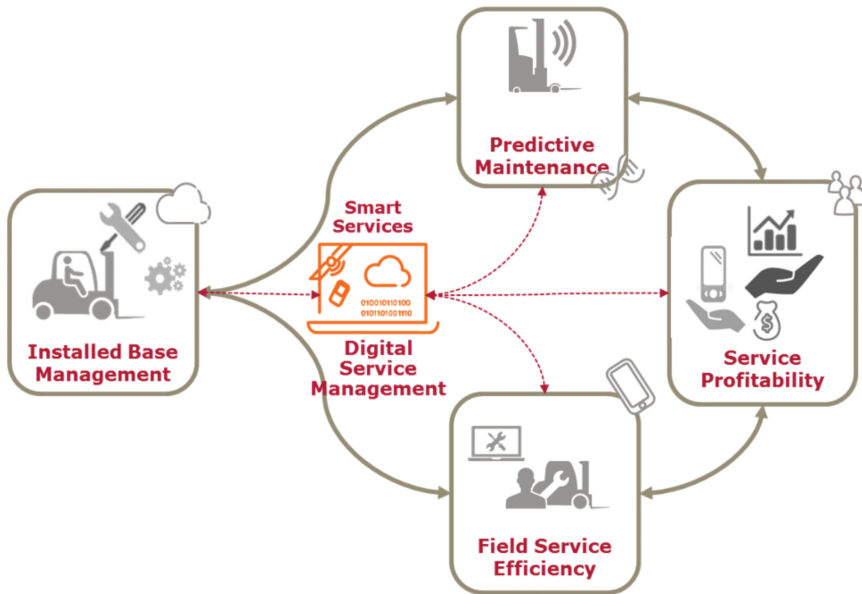
Figure 7.90 Smart service landscape as a possible result of the Mantis basis work.

technical and historical know-how and supplier specifications. There is no off-the-shelf solution for such brand specific processes on the market, all the tools which are used to day have been custom-made for the company. This is also due to the fact that this is property and essential technical expertise of the company, so there is a need to protect this intellectual property. But because of the increasing demands of the market and the increasing price pressure, processes have to be shortened and optimized. In addition, the complexity and variability of forklift trucks is increasing, so solutions have to be implemented to support the service technician in his work. In order to ensure the next steps in the service evolution, clear statements in the process chain must make fault detection clearer. First time fix is one of the most important goals of the future. By analyzing the internal system values of machine components and all other existing databases related to the service process (master data, repair databases, customer information) a fundamental base for this steps will be generated. For this reason, both the technical know-how from the examination of defective parts and the big data analysis will be used to identify specific patterns in the application and the environmental conditions that lead to breakdowns or high service costs. STILL GmbH is focusing on two main topics - Wear and Root cause analysis. Both topics

are too extensive, that a global solution found in one research project, so the expectation lies in creating first demonstration cases on this topics, which are to be refined over a longer observation period.

So concerning wear, STILL is finally focusing now on the subject of tires, because there are measurable conditions. All other relevant topics of wear of filters, mechanical and electronic parts have been discussed in the first phase of this project, but the necessary measures go beyond the possibilities of MANTIS. The study focuses in particular on the relationship between usages of the forklift by the driver and wear, since this knowledge can also be used in other business models (e.g., pay per use). Due to the lack of other environmental information like temperature, humidity and quality of the ground (this information is not recorded by the truck automatically) which are also important on wear, their influence has to be estimated.

In the topic of RCA, STILL uses the existing error messages of the forklift trucks with regard to the cumulated service reports and internal knowledge. At present, most error messages have no clear reference to the existing error screens in the field, since they are created as developer knowledge under laboratory conditions. Influencing conditions such as the environment or faulty interaction with other damaged components can be difficult to simulate in the laboratory. Under real conditions, however, the causal chains can differ considerably, so that errors can have other causes or effects. A broken wire can cause the display of a device defect, although the control unit is fully functional. The aim is therefore to achieve useful results through pattern finding and cooperative decision-making. The project will initially focus on some electronic errors to validate these results. However, these analyses are made more difficult by the fact that the error reports are freely formulated and do not show any clarity either.

The illustrated example (Figure 7.91) shows the complete range of components required for a powerful future concept in the field. Most of these modules are only listed for the purpose of being complete, but will not be considered in the following context.

7.5.3 Data Platform and Sensors

As mentioned above, most industrial companies, as well as STILL, tend to have a grown data infrastructure. For these reasons, data are neither harmonized nor centralized. The data used can therefore be based on platforms whose technologies are up to 10 years apart. The demands on the merge are enormous, especially since many data have not been checked for
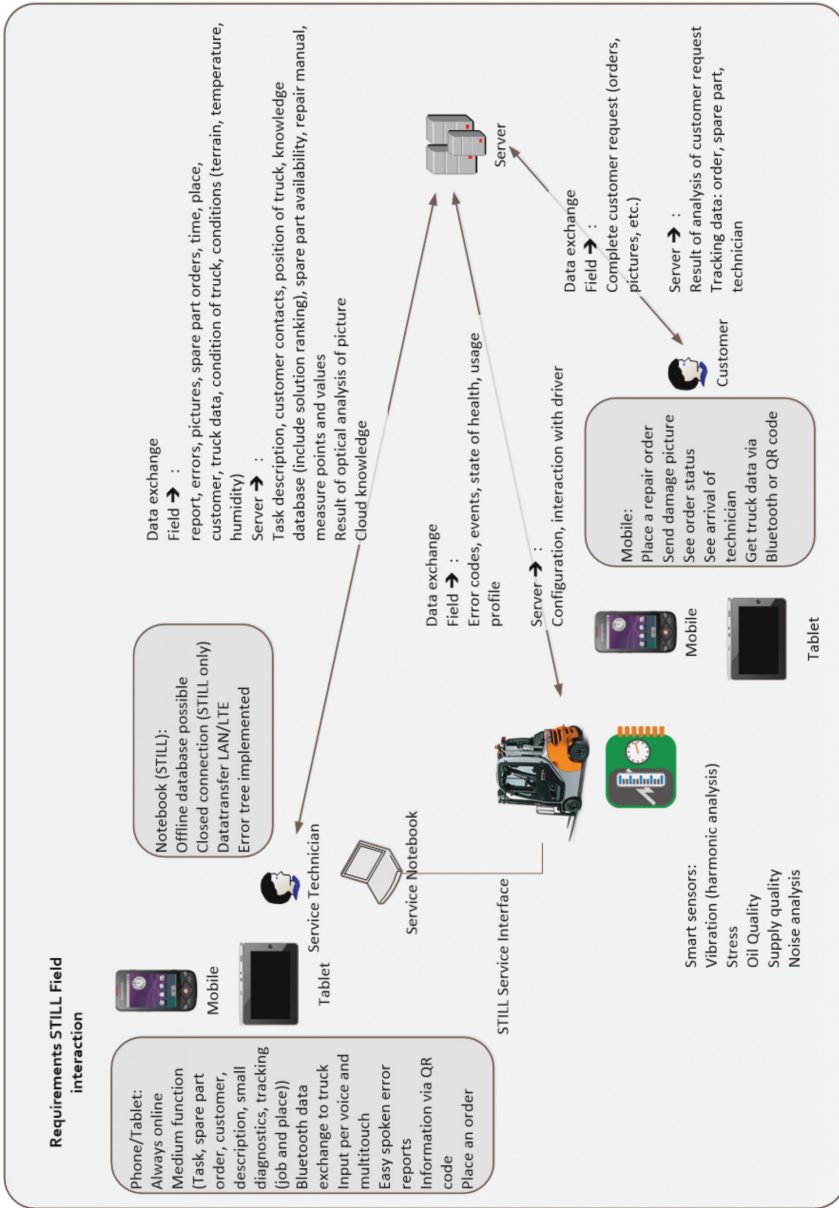
Figure 7.91 Overall scenario of a resulting field configuration and interaction.

their electronic processing capability in the past. It must therefore be checked whether it is meaningful and effective to integrate these data via interface, or whether one prefers to collect the data in a new and future-proof way. This is also due to the fact that 80% of the time spent on data analysis flows into data preparation. The following graphic (Figure 7.92) shows the necessary data sources for the following analytic steps.

Since the data landscape of STILL GmbH cannot be completely replicated for MANTIS, a demonstration solution on MS Azure was chosen (see Figure 7.93). The main reasons for this decision is the great flexibility of the step and the simple scalability. All Elements can be selected without having a negative impact on normal business processes. In the initial phase, a small number of internal forklifts (up to 5) are connected to this system. The advantage of this approach lies in the traceability of the use of these forklifts. The results can thus be validated directly. Once usable results are achieved, the integration of forklifts can also be extended to rental forklifts, for example.

The presented architecture serves as a basis for the validation and further development of the resulting knowledge from the project. As data sources, the architecture includes vehicle data as well as data from company databases and from results of the interaction with the service technician. It is deliberately designed in such a way that individual blocks can be extended or replaced as required. Communication with the forklift trucks is via Microsoft's IoT Hub[1]. The industrial protocol MQTT is used for data transmission. The incoming data can be pre-processed via various stream analytics blocks and thus either reformatted or already evaluated in parts for further steps. The incoming data is stored in the data lake for the next analysis steps. This enables the possibility of offline processing. Modified analysis methods can then be applied to vehicle data several times and the results can be compared directly with each other. This part is mainly used for RUL. For RCA additional the event hub is used, which can convert online analysis results directly into an action. This will be used for pattern recognition. To make the results available for the service technician in a special smartphone app, STILL uses the API app from Azure. Company data from the SAP system is coupled in via the Azure data factory element.

As mentioned in the previous paragraphs, forklift truck data in particular are used for the analysis. The forklift truck itself has a large number

---

[1]An IoT Hub is a site focused on the connectivity between software, the cloud and the devices used in everyday business operations.
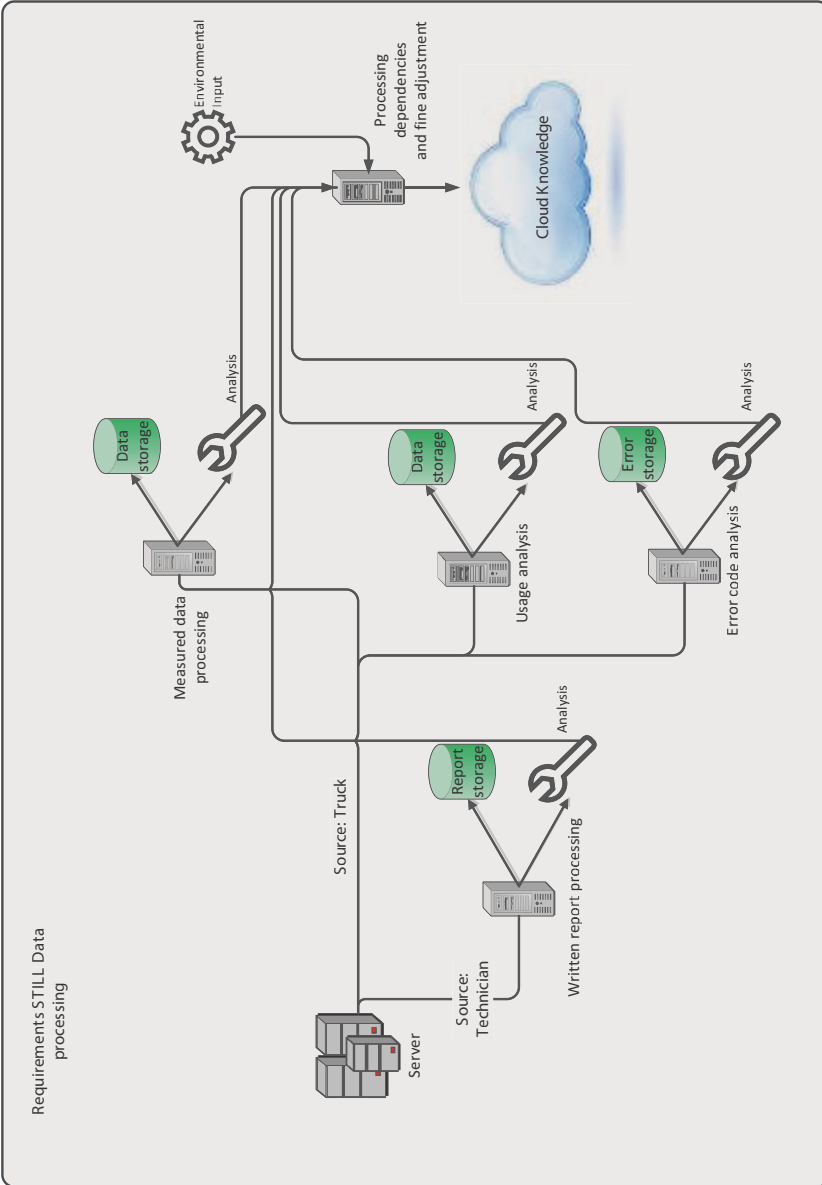
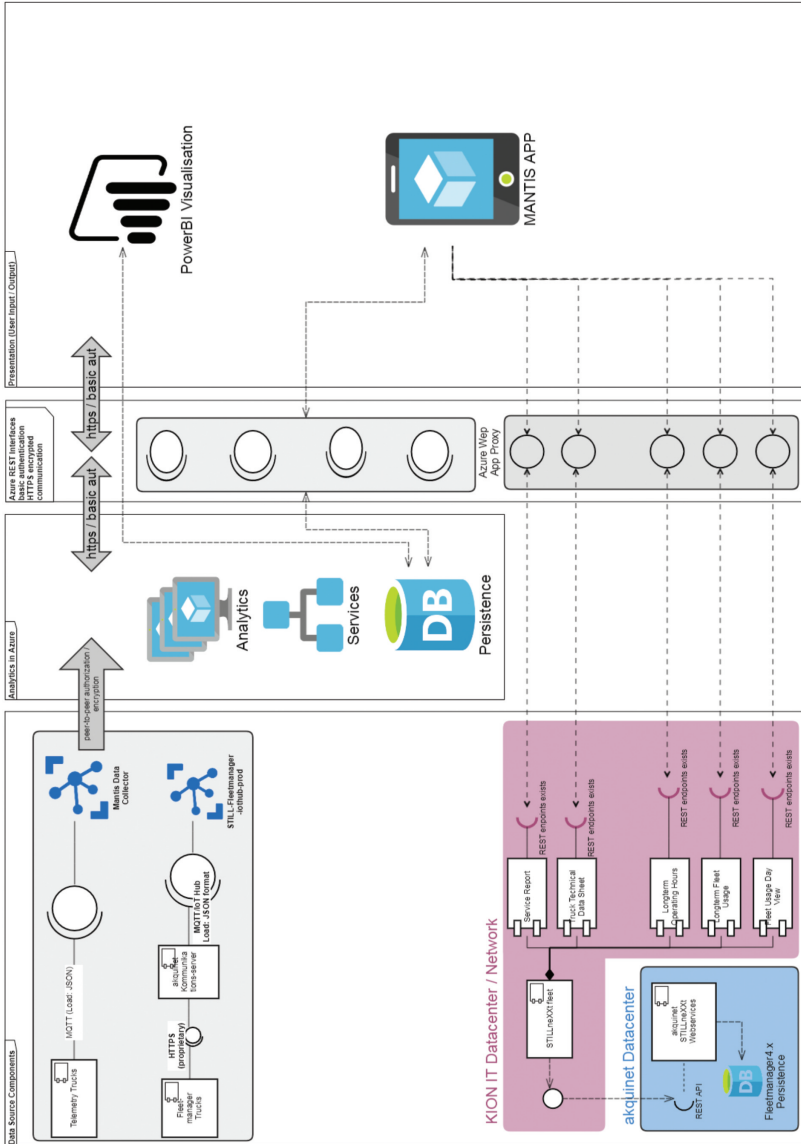**Figure 7.92**   Data sources for downstream analytics.

Figure 7.93   MS Azure demonstrator architecture approach.

of sensors that are used to control its movements and processes to the operator. These sensors are primarily intended for internal control operations, so their measurement data are not transferred to the outside world for further processing as a standard. Therefore, there is a need for an additional external sensor that records on one hand all relevant external measurement data autonomously. So this sensor should provide the following external information to draw a precise picture of the environment conditions:

- Acceleration in x, y and z direction;
- Ambient temperature;
- Humidity;
- Pollution degree.

On the other hand he should also have access to internal control values, with the aim to draw a precise picture of the detailed usage profile and transfer this data to the Azure cloud.

This sensor is shown in Figure 7.94 and it was built prototypically, but due to the complex mechanical interfaces, e.g., for dirt detection, the sensor could not be implemented in this project for a real validation on the forklift truck. Nevertheless, the functional principle is promising and will be used in further developments. Without this sensor, the environmental data are initially determined via questionnaires and weather information from the internet. The vehicle information is provided by so-called soft sensors. Soft sensors are software solutions that convert existing system variables into algorithmic data.

First, all available and relevant data for the creation of possible soft sensors were recorded via data logging and visualized for the forklift. For
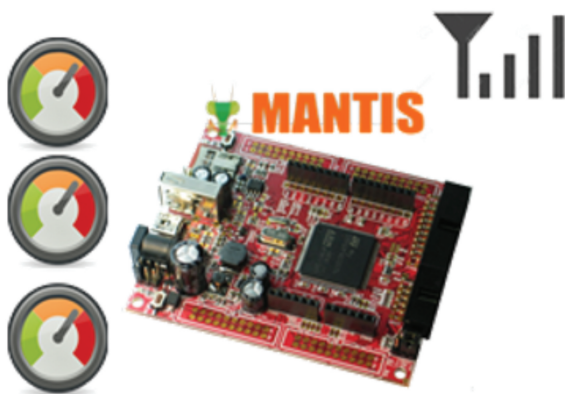

Figure 7.94 Raspberry Pi based sensor based.

logging, powerful multi-channel industrial data loggers are used, which record the data traffic on the can bus in real time (see Figure 7.95). It is important to ensure that the process does not interfere, as the vehicles have a high risk potential during operation and are subject to the Machinery Directive. Especially since most measurements have to be carried out on real customers in order to obtain data sets that are as variable as possible. These several gigabytes of data were then processed via Matlab[2] to determine the most promising constellation.

7.5.4 Data Analytics and Maintenance Optimization

The core problem for the subsequent analytical processes lies in both data preparation and the transfer of expert knowledge as described in the chapters before. The data is therefore first processed after the ETL[3] process to extract the core information. The data situation in mobile systems is fundamentally problematic for statistical processes, since the high transfer costs and the low storage depth of these devices mean that less data is available about these systems than needed for statistical analytics. The high influence of the environment also reduces the possibility to extract algorithms from the data, because each additional element need more validation data. The last burdening influence on analytics is the lack of precision in describing the problem. Many data sources (e.g., service reports) are not designed for later

Figure 7.95　Data Logging on the internal CAN bus.

---

[2]Matlab (spelling: MATLAB) is a commercial software from the US company MathWorks for solving mathematical problems and displaying the results graphically.

[3]ETL is a process in which data from several possibly differently structured data sources is combined in a target database.

analysis processes. Several preliminary studies are conducted to find out which data sources and constellations are suitable for the expansion of its existing processes and business models. Among other things, tools such as Spotfire[4] and programming languages such as R[5] are used for faster analysis visualization.

The example in Figure 7.96 illustrates the analysis of vehicle usage data with regard to fault events for RCA. In particular, pattern analysis plays an important role. Are there recurring patterns to which special errors can be assigned? Once such patterns are identified, appropriate processes can be stored to perform automatic service actions. For this reason, STILL initially selected errors for its basic analysis that have little scope for interpretation. For meaningful results, the ambient boundary conditions are also of great importance (see also Chapter 5 Section 2). For the final results in real-time operating systems, it is necessary to choose learning systems that allow an adequate response also to each new situation.
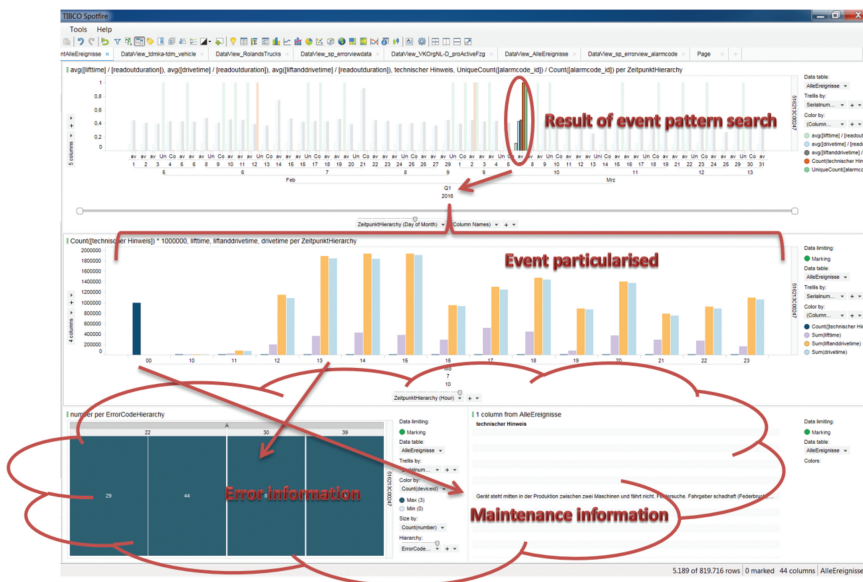

Figure 7.96   Example for graphical pattern recognition by Spotfire.

---

[4]TIBCO Spotfire Analytics is a commercial software platform for business intelligence solutions for the systematic analysis of internal and external data.

[5]R is a free programming language for statistical calculations and graphics.

In the area of wear, as mentioned above, STILL has concentrated on tires, assuming that the tires in particular give a good picture of how the truck is used by the driver and of the interaction with the environmental situation. STILL has analyzed the data sets of approximately 70 forklift trucks in the topic RUL in terms of use and tire wear.

However, the first approaches basing on the service reports were problematic, since the reason for replacing a tire can be quite different:

Not every tire is changed because it is worn. A lot of tires are also changed, due to damages (see Figure 7.97), so the existing database is faulty with respect to precise information about the real wear. Due to this fact we start with a basic analytic, which will be optimized during operation.

In order to find a pragmatic solution, STILL approached the problem by analyzing the behavior of the different forklift trucks. This means that the forklift trucks are examined over a long period of time and their behavior is classified. The basis for this assumption is the fact that there must

Figure 7.97   Tire changed due to wear and damage.

be a statistical correlation between physical usage and the resulting wear. Therefore, various operating conditions of the forklift truck are considered with regard to their possible influence. It turned out that agility seems to be the one of the most promising values, since its influence on wear is disproportionately high. For the first analytics the agility was divided into 100 elements, where 100 is representing an extreme dynamic driving (Figure 7.98). As a rule, the forklift trucks are observed over a period of one year, so that it can be assumed that we have an average of the typical usage of the trucks. The individual dynamic elements are summed up per vehicle and per class. Due to a simple weighting of the dynamic classes, a ranking of the dynamic use could be created, which corrodes in parts with the wear. Since the ground condition, load and environmental conditions were not considered, certain deviations can be explained. For the complete model further analyses are necessary, but these are not carried out in the course of this project.

The graphic in Figure 7.98 also shows that the weighting does not always allow an explicit conclusion. Vehicles with low dynamics can be clearly identified and the result fits to their wear. But there are also vehicles in the midfield in particular that have entries in the very dynamic classes. However, due to the weighting selected so far, they are classified as medium in terms of wear. Since the associated tire wear does not match the placement, it can be assumed that a linear weighting between the classes does not seem to finally apply. There are currently too few data sets available for a clear statement, so that the results still need to be sharpened with appropriate self-learning mechanisms.
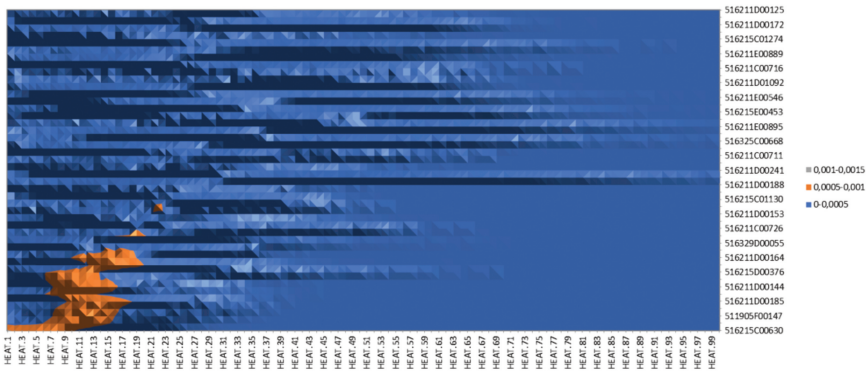

Figure 7.98   Agility heat map of analysed forklift trucks.

The extent to which the knowledge gained can be used to improve maintenance processes must be demonstrated by subsequent field studies. An installed example architecture (shown in green in Figure 7.99) will produce over a period of time to be defined, results of the algorithm. This results are then rated and used to refine the analytics.

The high level of service product requirements is due to the fact that a large part of these products are subject to a fee for the customer. For this reason, the highest diligence is required when developing solutions so that these do not have a negative cost effect on the customer and burden the business relationship. However, there is always a need to improve processes, since all services provided are subject to considerable cost pressure from the market. For this reason, the provision of individualized solutions and the increase in effectiveness is of great importance to the company.

The developed solution modules from the wear area are used in particular for the optimization of demand-based billing of services like rental or full service, while the RCA solutions are to be used for increasing the effectiveness in the processing of defects in the field.
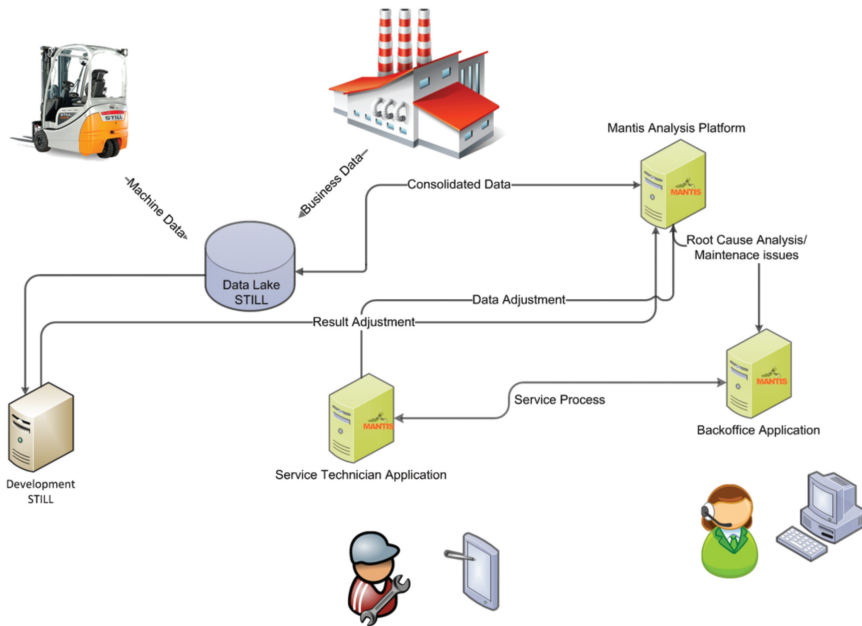

Figure 7.99   Second step in architecture to improve process results.

7.5.5 Conclusions

Considerations in the project have shown that mobile systems with small data volumes are particularly difficult to handle. For this reason, technical aspects are required in such systems in a much higher than expected dimension. Purely statistical observation cannot lead to success with such conditions.

It became clear that especially the knowledge of the required data is of great importance and an important key to success. It also became clear that the quality of the data has to be checked at an early stage and ensured by appropriate measures.

For STILL, these considerations provide clear starting points that can be used in future applications. Due to the very complex relationships between physical vehicle use and the interaction with environmental influences, an expansion of the data structures will be necessary in the future. The results achieved so far are refined in further validation loops and enriched by artificial intelligence. This is necessary in order to realize the high potential in the general area of optimization of maintenance processes. However, in addition to the classical support in problem solving tips and analyses results, there are also elements that will result from visual and/or HMI technology. These will even open up opportunities for further business areas.

A chosen system architecture must therefore always offer the possibility of expansion. The platform of this system architecture plays a rather subordinate role, as the solutions present themselves as a kind of modular system rather than an integrated solution.

7.6 Proactive Maintenance of Railway Switches

Contributors: Csaba Hegedűs, Paolo Ciancarini, Attila Frankó,
Aleš Kancilija, István Moldován, Gregor Papa, Špela Poklukar,
Mario Riccardi, Alberto Sillitti, Pal Varga, Paolo Sannino,
Salvatore Esposito, and Antonio Ruggieri

A larger pressure on the railway infrastructure has been created by a strong necessity for faster mass transport with high capacities and frequent runs. This makes it fundamental to continuously monitor the technical equipment of the railroad tracks in the most efficient way possible. By detecting fatigue wear of the track system in an early stage – due to issues such as broken rails or increased rail wear, caused by natural hazards or by excess loading on the track system – it is possible to avoid serious damage, also by means

of correct interpretation of collected data, which allows for rapid intervention on the track system.

The railway use-case within the MANTIS project [The MANTIS consortium] is dealing with these issues through a proactive maintenance approach of the railway system [Hegedus et al., 2018]. This concerns the interlocking system and the study of possible complications that affect railway signalling – i.e., non-functional and out of control situations – with its main focus on describing the development of a set of approaches and support tools which allow to continuously analyse the status of specific components within the infrastructure. The use case aims at determining whether and within which limits it is possible to make reliable predictions for improving the maintenance process. In particular, it targets identifying anomalies and reducing emergency maintenance, since it is very costly and cases major train delays.

7.6.1 Introduction to Railway Monitoring

In the railway infrastructure, the prevailing maintenance approach is still following the preventive model where most of the maintenance operations are based on periodical check-ups and substitutions of parts when a failure is detected. These tasks are carried out at given periodical intervals designed to mitigate risk with a considerable safety margin involving having to send maintenance staff to the asset site on a regular basis, exposing them to the usual safety risks of a running railway [Cocciaglia, 2012].

Modern railways have very low level of signalling installed. For switches, this comprises only of the detection if a switch is in the correct end position and locked. The development of new maintenance systems, including the integration of heterogeneous monitoring and diagnostic technologies, plays a key role in the improvement of railway safety operations. Existing monitoring solutions show some limitations due to their non-standardized, proprietary nature and very low integration level. Consequently, they are not able to monitor properly the degradation of complex asset, and to detect correlations between the condition of assets [Cocciaglia, 2012].

7.6.2 Scope and Logic

A railroad switch allows trains to change tracks (Figure 7.100). When a train is destined to run on another track, the switch-man on the train or another employee in the railroad yard will turn the switch to direct the train toward the chosen direction. The railroad switch is activated by moving a long arm
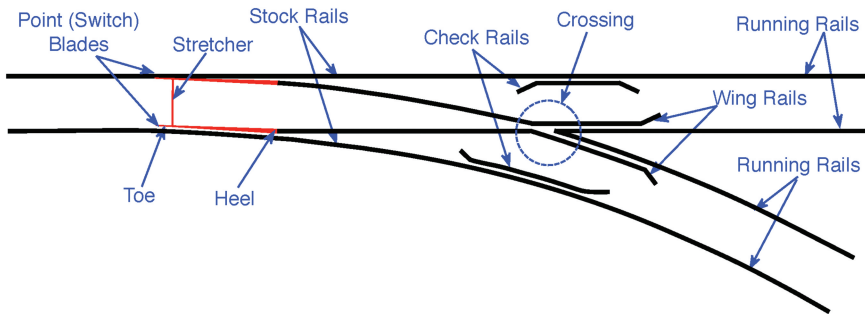
Diagram of simple turnout showing the names of the principal parts.
Moving switch blades shown in red.

Figure 7.100 Simple turnout schematic. Moving switch blades are shown in red.

from side to side and moving the train tracks to the desired position. While many railroad switch activations are accomplished by hand, nowadays some are electronic and can be changed by an employee in an elevated office at the railroad yard.

A realistic proactive maintenance solution for railway switches is based on the concept of Cyber-Physical Systems (CPSs), where a cyber-twin of the physical system is modelled, and its status is kept up-to-date through data collection from physical sensors deployed on-site.

The rest of this section provides some details on data processing, presents the proactive measurement system, and describes the developed data visualization subsystem.

7.6.3 Data Processing

The data processing requires the consideration of all the available datasets connected to each switch. In particular, the data available are in the form of time series and can be grouped as follows:

- **Control:** data generated by the switch control unit. They include commands sent to the switch (i.e., start, stop, etc.), and some feedback data provided by already existing sensors in the switch. The collected information is coarse grained, with a log sequence structure;
- **Physical:** data generated by sensors temporary added to the switches to measure some specific parameters. The most interesting data used in this analysis are the electric current consumed during the movement [Ampere], the duration of the movement [second], and the environmental temperature [°C].

The goal is to identify anomalies that could require a maintenance activity. We have focused on the following behaviours:

- **Drifts of the profiles:** could be caused by accumulation of dust on the switch resulting in an increased amount of current leading also to a failure of the switch;
- **Unexpected behaviour:** could be caused by physical obstacles in the switch that may cause damages to the device.

During data exploration, we have identified the following behaviours:

- **Behaviour 1** (Figure 7.101)**:** This is a very noisy profile that makes the identification of the behaviour difficult and may highlight problems in the data collection. In particular, in the correct positioning of the sensor and/or the presence of sources of noise that may alter the collection;
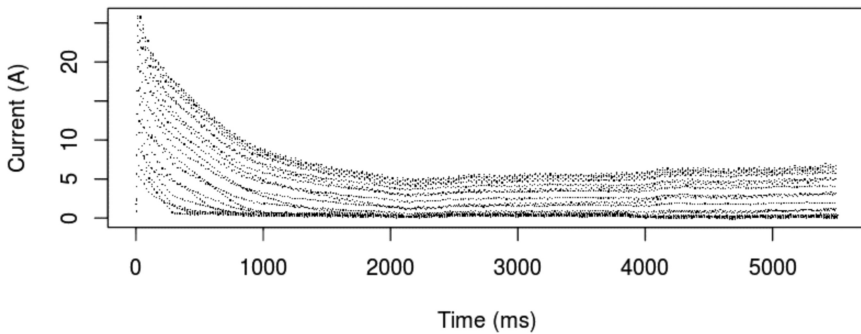- **Behaviour 2** (Figure 7.102)**:** Similar to Profile 1 but with a limited amount of noise;

Figure 7.101   Switch Data – Behaviour 1.
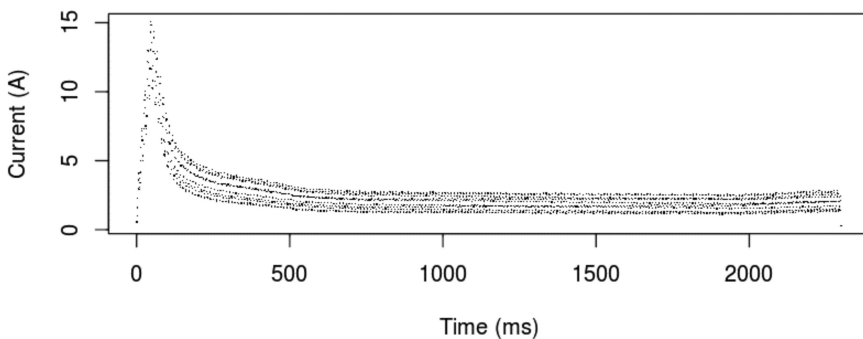
Figure 7.102   Switch Data – Behaviour 2.

- **Behaviour 3** (Figure 7.103)**:** Expected profile of a double switch;
- **Behaviour 4** (Figure 7.104)**:** Expected profile of a switch;
- **Behaviour 5** (Figure 7.105)**:** Profile of a switch with an abnormal behaviour.

The profiles of the current depend on several physical variables linked to the mechanical and electrical components that compose switches. These profiles are linked to the specific model of the switch from which the data are collected. The current is influenced also by environmental factors: temperature, humidity, and dust.

Due to the large variability of the profiles, our main problem is the identification of an approach to define the default correct behaviour. This can be achieved in different ways:

- **Physics:** this approach is able to define the physical model of each switch, and it is able to predict the correct behaviour in many different
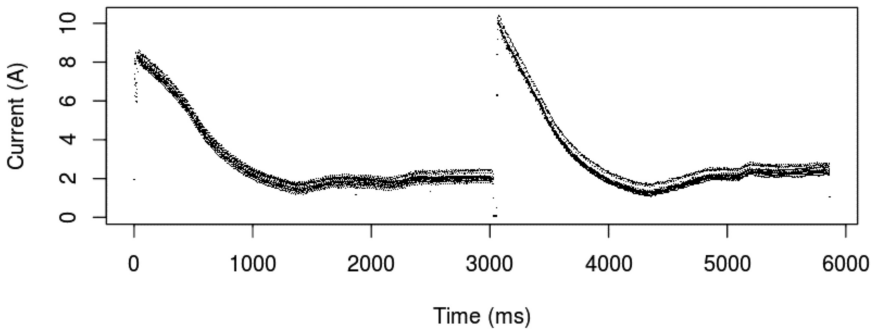
Figure 7.103   Switch Data – Behaviour 3.
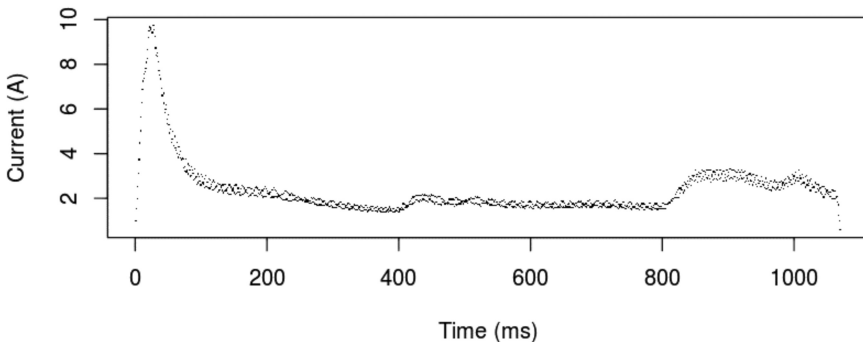
Figure 7.104   Switch Data – Behaviour 4.
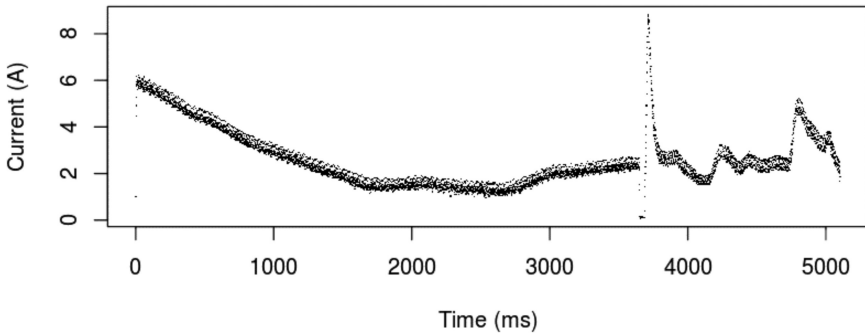
Figure 7.105 Switch Data – Behaviour 5.

environmental conditions. However, it requires building a model for each kind of switch and tuning the parameters for each installation;

- **Statistics:** this approach requires the collection of data from a wide set of devices in different operating conditions to define the default behaviours that are known with some level of uncertainty, but it does not require the manual development of a physical model for each switch. The model can be derived from the data and can be adapted to different switches collecting additional data.

Figure 7.106 shows the statistical features of a specific switch. The black line is generated calculating the median of the different time series representing the current profiles of hundreds of movements that happened correctly in the past.

Threshold detection was used. If the current is outside the bounds, a warning is risen. The definition of proper bounds is of a great importance
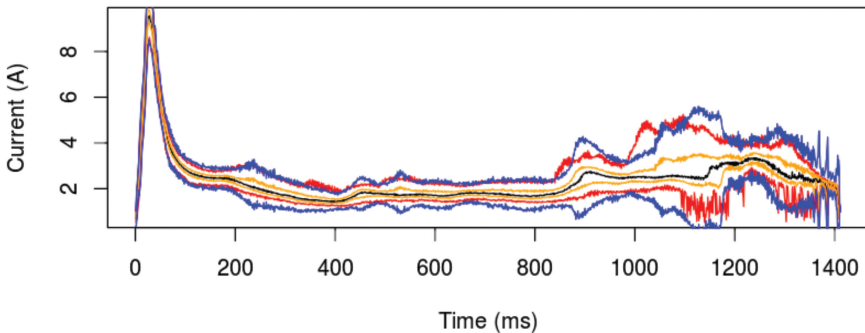
Figure 7.106 Identification of bands for quartiles and outliers.

for the detection of an abnormal behaviour. As the distribution of the samples at each time instance is not normal, the outliers' bounds are defined using the 1st and 3rd quartiles, according to the Tukey's range test for outliers [q1-1.5*IRQ; q3+1.5*IRQ].

There is quite a large range, especially at the end of the movement. After a deeper investigation, it was found that the behaviour was caused by the fact that the analysed single set of data was hiding two different data sets. Actually, the behaviour of the switch in the summer and in the winter is different due to the temperature sensitiveness. There are several aspects that depend on the temperature, such as the duration of the movements (longer in winter) and the current peaks (higher in winter). For these reasons, the statistical model had to consider the current season (the temperature of the environment). Therefore, the same analysis was repeated, but the dataset was divided into two sets based on the time of the year of the data.

The model validation was done by a bootstrap approach building the model using a random subset of correct movements in the same season and verifying it with the rest of the data. This analysis was helpful for defining the statistically correct behaviour of a switch using data coming from the field and tuning a model without any specific knowledge of the internal structure of the switch. Such a model can be easily adapted to different switches working in different conditions, and it was tuned using data in the different seasons.

The data can be analysed by log analysis approaches. However, the coarse grain and the lack of a sufficient information from the field (including tagged data describing anomalies) resulted in an analysis that was not able to build a relevant model that can actually be used.

So, the aim for analysing such data by applying those different statistical approaches is to determine a model of the default behaviour of the switch, and to identify anomalies in the behaviour. Among various purposes of diagnosis and prognosis [Jantunen et al., 2016], this can be used for failure prediction [Fronza et al., 2013], and other proactive maintenance purposes [Lenarduzzi et al., 2017], including root cause analysis and the calculation of remaining useful life.

7.6.4 Measurement System for Proactive Maintenance of Railway Switches

For failure prediction and diagnostics, a new maintenance system was needed. We built a new, low cost non-invasive measurement system that can be attached in retrofit to operational switches. The measurement

system measures the factors that affect the life expectancy of the railway infrastructure. The choice of the appropriate attributes is based on expert knowledge since the different types of switches are not equally affected by these impacts.

The data acquisition system is part of an architecture which is based on the MANTIS platform [Hegedus et al., 2018] and complies with the architecture of the platform in full extent. The architecture of the system is shown on Figure 7.107 and consists on the following modules: (1) Standalone data gathering edge device; (2) Edge broker implementing MQTT; (3) MIMOSA database on a Microsoft SQL Server; (4) Data analytic modules; and (5) MANTIS Human-Machine Interface (HMI).

The edge device – which is the embedded subsystem deployed with the railway switch – is not only responsible for gathering new data but pre-processing and forwarding it to the cloud in an appropriate, MANTIS-enabled message format. The heart of this device is an STM32F4 series MCU (Microcontroller Unit) which employs a single ARM-Cortex-M4 core is capable of collecting, storing and pre-processing the information, while also handing the communication tasks as well. It offers numerous interfaces – including UART, SPI, I2C –, and 12-bit analogue-to-digital converters; thus both analogue and digital sensors can be used.

In this use case, the edge device contains one digital integrated humidity and ambient temperature sensor, a digital temperature sensor and four analogue displacement sensors.
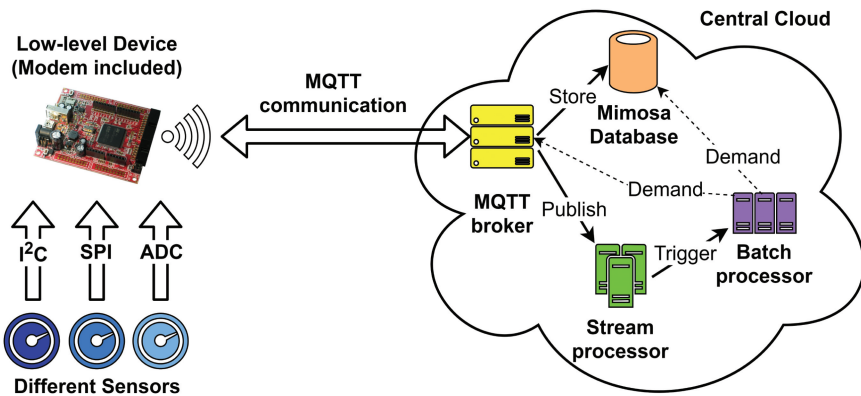

Figure 7.107   Measurement system setup.

7.6.4.1 New factors collected

The system measures several factors that can affect the wear of the railway switch over time. These expert-identified factors can be divided into two groups:

- **Operational factors:** These parameters are directly related to the operation of switches – they have direct impact on condition deterioration. In our implementation, we measure lateral and longitudinal displacement of point blades. These point blades direct trains to one of the possible paths, i.e., they are the moving parts of a switch. Here the excepted resolution is high, and we are interested in gathering data only during switching sequences;

- **Environmental factors:** These parameters are well-known to affect almost every cyber-physical system. The most significant one is temperature. Both the ambient temperature and the temperature of the rails are measured. The rail temperature can cause dilation of rails thus it affects the operation of switches indirectly. Another environmental factor is humidity, which plays a lead role in corrosion. Since the ambient parameters are changing slowly, reading the values periodically, every half an hour provides appropriate accuracy and resolution for this use-case.

The blade movement measurement must be event driven, data is collected when a switching of blades occurs. Therefore, the switching itself must be detected. Detecting the start of a switchover is tricky, since measurement noise and passing trains may interfere. Therefore, a threshold based triggering is used together with a pre-fetch measurement phase, as Figure 7.108 shows. The thresholds are set high enough to avoid false positive triggers, and the pre-fetch phase ensures that the acquired data contains the full movement of the blades. Moreover, if the device starts a measurement and the actual position does not reach the end position – just nearly approaches it –, the measurement cycle will not stop. In this case all information about the movement between the real end positions and the threshold levels would be lost.

The state-transitions of the measurement are presented on Figure 7.109. In the case when the measurement takes longer than a predefined (expected) interval, the measurement stops and triggers the device to send a warning message to the central cloud. This function indicates an error, which means that the point blades cannot reach their end position – so the switching operation failed.
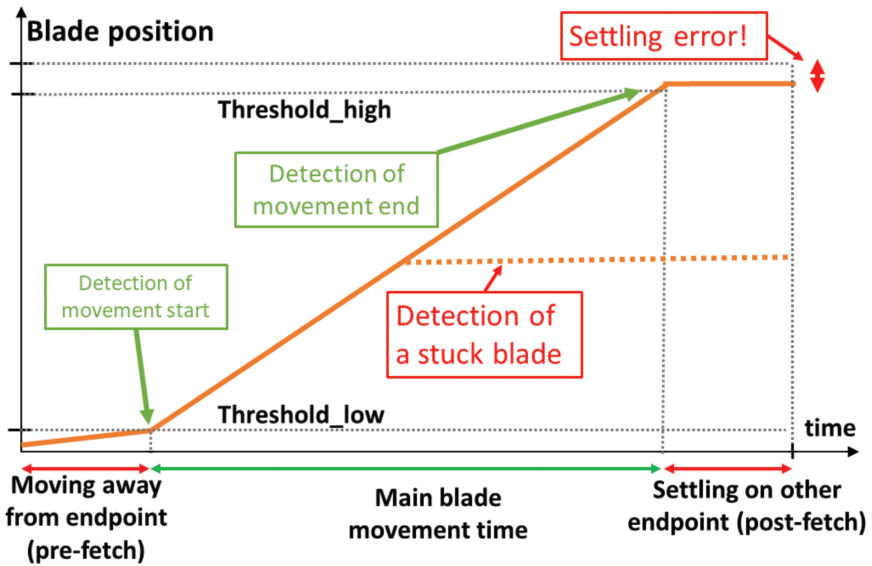
Figure 7.108 Switch's point blades displacements.

7.6.4.1.1 *Platform level*

The gathered information is encoded in an interoperable JSON-based message format developed within the MANTIS project, based on the MIMOSA [MIMOSA Consortium] domain ontology. The messages contain not only the results of measurements, but additional information: (i) exact timestamp, (ii) duration of the measurement, (iii) identifier of the edge device instance and (iv) additional values that help the re-assembly of the message at broker side.

The messages are transmitted via the MQTT protocol over TCP/IP. The wireless connection between the edge device and the central cloud is provided by a SIMcom SIM800 based GPRS modem which is attached to the MCU via serial line. The central cloud contains an MQTT Edge Broker, which handles the messaging, while both the Low-level Device and the cloud have an MQTT implementation each.

In the central cloud, the message is received by a Mosquitto MQTT broker [Mosquitto^{TM}, 2010] with a parser client. The information is then stored into a MIMOSA OSA-CBM database, which is a standard architecture for condition-based maintenance systems. The parsed datasets will be processed offline by data mining and analysing tools. Future work includes that the incoming message can be analysed online, automatically by a stream processor. This will enable an automated alerting and forecasting system.
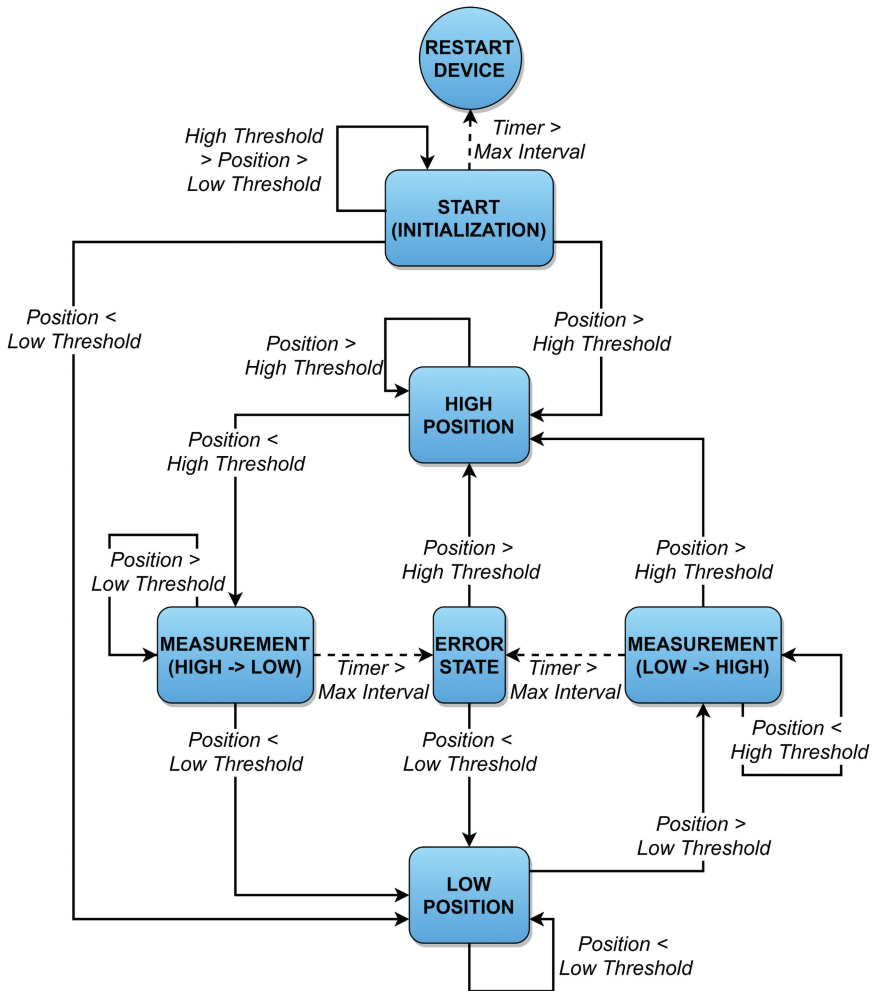
Figure 7.109 Measurement state diagram.

The processed and analysed information is stored in the database, thus the central cloud can provide relevant information to different parts of the MANTIS architecture, for example for the Human-Machine Interfaces.

7.6.5 Data Visualization

To increase the efficiency of the maintenance personnel and to evaluate the results of the data analysis [Korošec et al., 2013], an intelligent HMI had

to be developed. Following the scenario-based design approach, user needs and the context of use were described in the human-machine interaction scenarios. In the iterative process of scenarios refinement, five main human roles have been identified, ranging from the maintenance technician to the business manager. Further refinement of the scenarios led to identification of three main functionalities of the user interface:

- Monitoring the parameters given by the measurement box;
- Displaying the alarms that indicate the abnormal movement of the railway switch;
- Displaying the task schedule for the maintenance service.

The interface was developed on top of the generic MANTIS HMI, described in Chapter 6. It supports multiple users with different roles, where each user or role can be presented with one or more dashboards covering their intended interaction. Dashboard is customizable and does not require any web development skills.

As it can be seen in Figure 7.110, the HMI allows the user to quickly see the position of the railway switch through the corresponding graphics. An additional graphics with the IoT image indicates the connection to the measurement box to ensure the reliability of the data. When the connection is established, the image turns green. Otherwise, the user should not assume
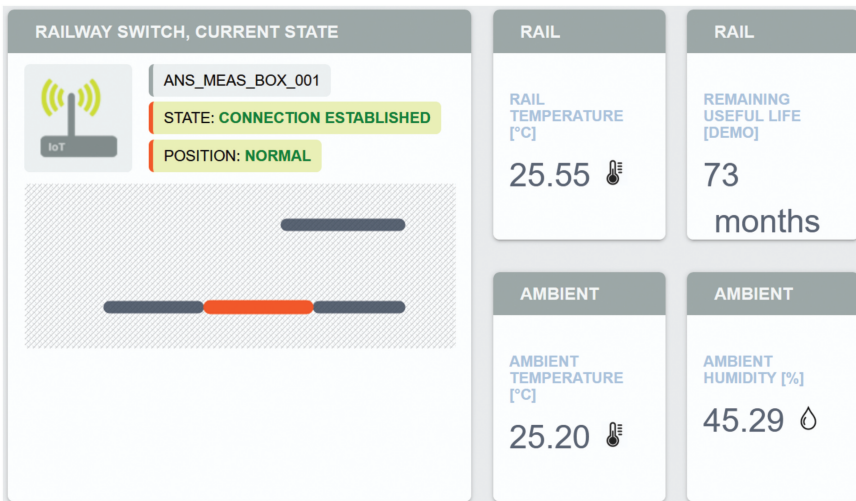

Figure 7.110   Graphics, displaying the state and the position of the switch (left) and the instant values of the environmental parameters (right).

that the data presented on the interface is accurate or up to date. In addition to the switch position and state, instant parameter values related to the railway switch, such as the rail temperature, switch status, ambient temperature and humidity, are displayed. If the alert thresholds for a measurement are set, values out of range will also be shown as alarms. Historic values of the raw sensor measurements and environmental parameters are displayed as a graph (Figure 7.111). Visualization of the data analysis and prediction results is done through the same monitoring widgets, which can also show predictions, remaining useful life estimations. For more in-depth analysis, Kibana visualizations are integrated.

Scheduled maintenance tasks are currently displayed in the alarms table, but they can also be displayed separately. The table is editable, filterable and sortable and allows the user to acknowledge the task/alarm as well as to enter textual feedback. To assist the maintenance personnel working on the field, a map with the location of the railway switch is displayed on a separate widget.

Several context-awareness features, mainly based on location and the user role, have been proposed to assist the maintenance personnel in performing their tasks. Such features proved to be most useful in performing the maintenance actions on the field, where the visualization of the information varies depending on the location of maintenance team. Another such example is a personalised suggestion of the user's next step according to their past interaction with the interface. In this way, the users are provided with the right information in the right moment and context.
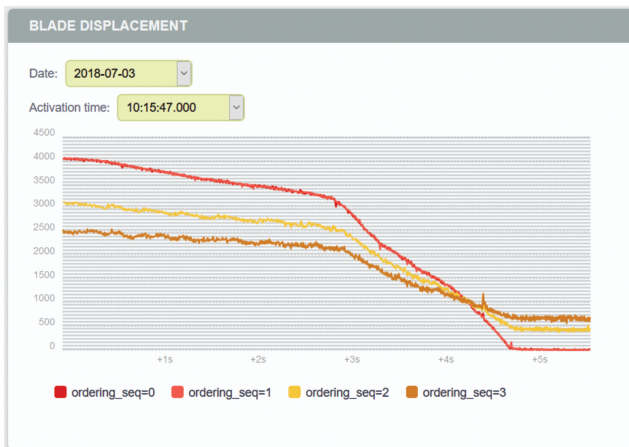

Figure 7.111   Graph displaying blade displacement sensor measurements.

7.6.6 Conclusion

The Industrial Internet of Things, the concept of CPS and the industrial initiatives force the evolution of proactive maintenance solutions for industrial systems. This section presented some results, where the MANTIS concepts were applied to the use-case of railway switches. These include the measurement method, the analysis of the measured data, and the result visualization HMI tailored for various types of contexts and users – ranging from the maintenance technician to the business manager.

7.7 Fault Detection for Photovoltaic Plants

Contributors: Achim Woyte, Babacar Sarr, Karel De Brabandere, Tom Tourwé

The high-level objectives of the Photovoltaic plants use case are mainly concerning the reduction of efforts for operation and maintenance (O&M) by more cost-efficient monitoring. This is pursued through smart sensing and data acquisition as well as through analysis and decision-making functions applied in real time to the operational data. These developed analytical methods are meant to reduce downtime and subsequent losses in electricity production due to component failures. Also one of the main objective is to improve the energetic performance of the plants by detecting design flaws, bad installation and maintenance practices, performance degradation of components over time, and sudden changes in the performance of components. Such possibility of early detections requires improvements of O&M scheduling departing from root cause analysis, alerting and prediction functions, and maintenance optimisation.

7.7.1 Introduction to PV Plants

Established in 1999, 3E is an independent technology and consultancy company. 3E provides solutions as well as guidance to improve renewable energy system performance, to optimise energy consumption and facilitate grid and power market interaction. 3E pursues innovation to provide leading energy intelligence and practical solutions to its customers and it disposes of long-term monitoring data sets recorded with high time resolution for more than 3000 PV installations distributed over the world with a total installed capacity of more than 2GW via its monitoring service SynaptiQ. 3E has worked on projects in more than 40 countries and operates with an

international team of around 100 experts from its headquarters in Brussels and offices in Toulouse, Paris, Beijing, Istanbul, Cape Town, and London. 3E is certified ISO 9001:2008 since early 2010.

A huge potential for more effective pro-active maintenance actions passing through automated fault detection and identification lies in the large amounts of PV monitoring data that are recorded but currently used in a very limited way. It is in this perspective of exploiting this potential that 3E has been developing analysis and decision-making functions for proactive maintenance of Photovoltaic (PV) plants.

In view of the objectives of this use case, mentioned at the beginning of this section, the exploration and validation of analysis and decision-making functions for proactive maintenance of PV plants has therefore been the principal focus. 3E and its partners have been developing intelligent functions for pyranometers sensors, and overall automatic PV plant analysis to assess the "health" of the plant.

The following section provides an overview of a practical application of RCA developed for the Photovoltaic Plants use case lead by 3E. It focuses on the illustration of one of the techniques used for fault detection: Limit checking applied on the PV use case.

7.7.2 Practical Application of Root Cause Analysis in Photovoltaic Plants

Photovoltaic plants are energy conversion systems. They convert the power of light, i.e., photon beams or electromagnetic waves, into electricity that can be used in an off-grid system and/or fed into the public utility grid in terms of frequency and voltage. The efficiency of this energy conversion step is influenced primarily by ambient temperature and secondarily by wind speed (Sw). Wind speed is often neglected.

Consequently, the primary input variables for this energy conversion process are the solar irradiance in the plane of the PV array (GPOA) and the ambient temperature (Tamb). The output variable is the electric AC power to the grid (PAC) as indicated in Figure 7.112. The PV module temperature (Tmod), the DC voltage, current and power at the output of the PV array (VDC, IDC, PDC, respectively), and the AC voltage and current to the grid (VAC, IAC, respectively) may be considered measurable state variables of this conversion process. The so-called yields (Y) and losses (L) describe the energy balance throughout the system in operation and are represented in Figure 7.112 at the different stages of the plant.
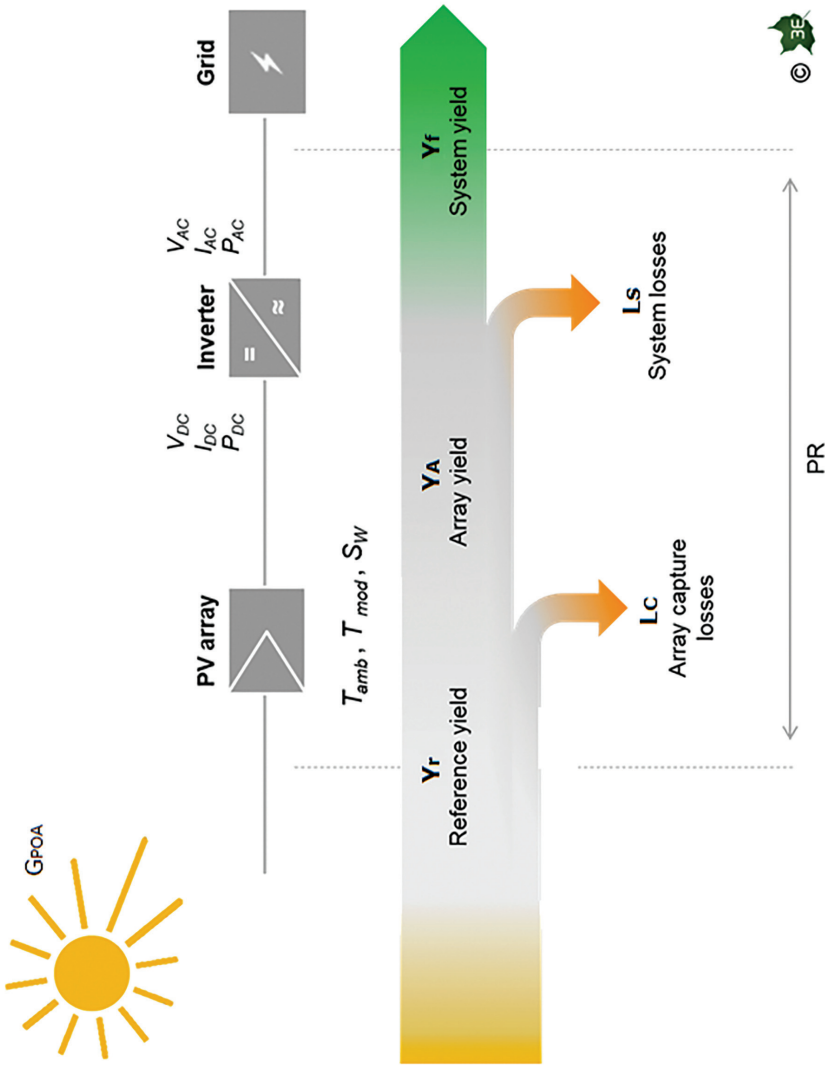
Figure 7.112 Energy flow in a grid-connected photovoltaic system.

Data received from the PV array comes in the form of time-series data. After standard data cleaning procedures, based on pre-mentioned variables listed above, normalized performance parameters are derived. They allow to quantify the energy flow and losses through the PV array per loss type.

They are:

- Availability loss: due to unavailability of grid, inverter, or DC input. This is encountered in the situation when power monitored from the plants is equal to zero while there is still light coming from the sun;
- Array-current loss: due to deviations of the measured DC current from proportionality with irradiance through STC (Standard Test Conditions), for times when the plant is available;
- Array-voltage loss: due to deviations of measured DC voltage from 'STC voltage';
- Inverter loss: due to deviations between measured AC and DC power.

The yields and losses are typically hourly average values but can be integrated over time.

The main variables used for limit checking are solar irradiance in the plane of the PV array (GPOA), ambient temperature (Tamb), PV module temperature (Tmod), DC voltage and current at the output of the PV array (VDC, IDC) and electric AC power injected to the grid (PAC). The AC voltage (VAC) and power factor (PF) are not used for limit checking.

For checking the operational performance over different energy conversion steps, a performance loss ratio per step is defined. This performance loss ratio is computed for a given time span, e.g., a day up to several months. It is the useful energy lost over the energy conversion step divided by the energy available, i.e., the incoming solar energy on the PV array as represented by the solar irradiance in the plane of the PV array (GPOA); all normalized to standard rating conditions of the PV array. Accordingly, the overall performance of a PV plant is described by the performance ratio (PR), i.e., 100% minus the sum of all performance losses.

In practice, we compare the performance loss ratios from measurements to model-based performance loss ratios and thresholds. The model is fed with measured values of GPOA and Tamb. The model parameters can be set from data sheet parameters of the devices in the PV plant or identified from measurements from the plant in a healthy state. Accordingly, adequate limits can be derived either from tolerances on the data sheet parameters or from choosing percentiles from the healthy plant. Both the model-based performance loss ratios and their limit values vary depending on the PV plant and the weather during the evaluation period.

Figure 7.113 illustrates this application of limit checking for a PV plant located in Belgium. The current-related array losses ('Array (current)') in the upper half of Figure 7.113 by far exceed the threshold. During a thorough
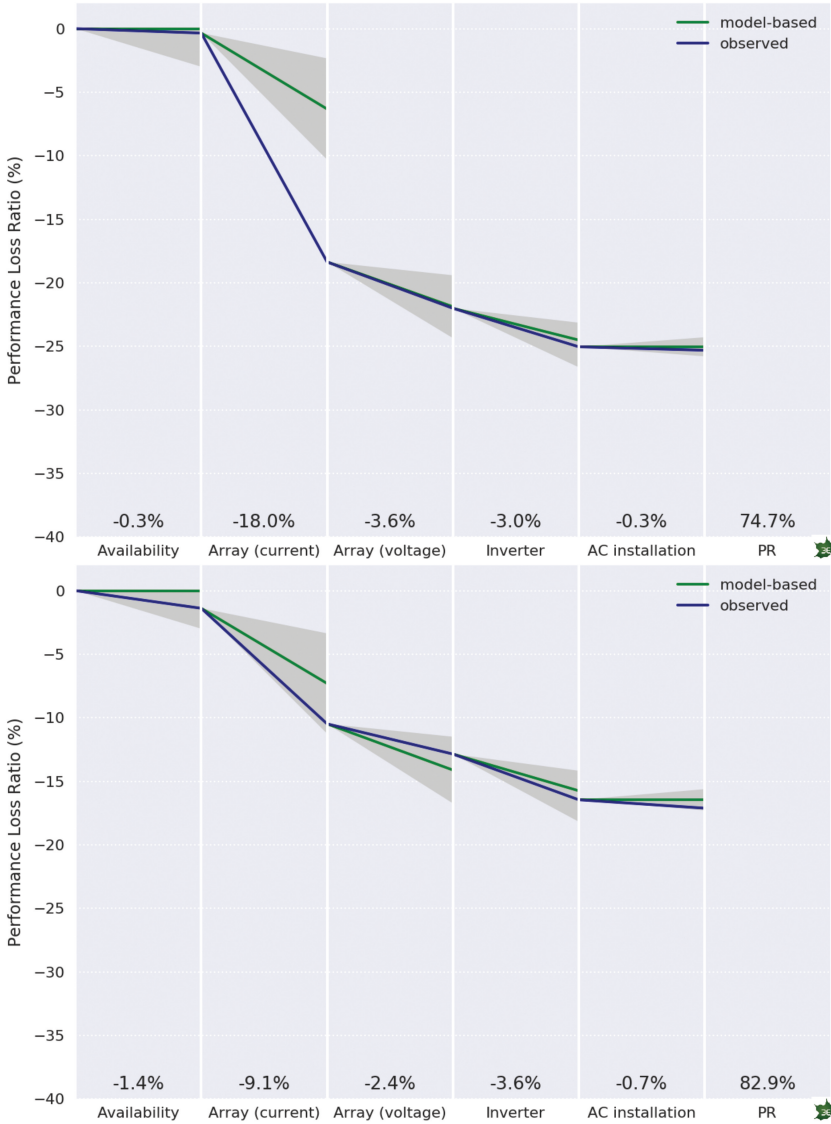

Figure 7.113 Example of limit checking results for the energy conversion process in a PV plant; performance loss ratios per conversion step are compared to the model for each conversion step.

maintenance action after this problem was detected, several smaller PV module failures were fixed. After maintenance action, all performance loss ratios were back within their expected ranges, yielding a much higher PR of 82.9% (lower half of Figure 7.113).

7.8 Conventional Energy Production

Contributors: Matti Kaija, Antti Niemelä, Juha Valtonen, Ville Rauhala, Veli-Pekka Salo, Erkki Jantunen

The most important goal of a power plant operator is to maximize the availability and performance of their power plant and to minimize the generating costs of energy. To reach this goal, the maintenance of the power plant is crucial and maintenance can account up to 30 % of the cost of the produced energy.

Most modern power plants use a combination of scheduled maintenance and corrective maintenance. In this strategy, critical components, such as turbines and pumps, are serviced based on statistical trend data and less critical components are left to a run-to-failure strategy, where maintenance is not done until the machinery fails. These maintenance strategies are not cost optimized since scheduled maintenance usually results in either unnecessary repairs if servicing is done too frequently or potentially catastrophic failures if service is neglected. Ineffective maintenance not only wastes materials and resources, but lost operating time resulting from equipment failure is a significant expenditure to a power plant operator.

Effective maintenance planning and scheduling is essential for shortening revision and downtime periods. Most of the revision work is done by external contractors and work delays can be expensive. By maintaining components based on their condition, revision work and cost can be optimized.

Current condition based maintenance strategies usually rely on periodical measurements done by experts, because the power plant staff do not have the necessary skills to analyse the measurement data. Most power plants collect information that could be utilized in condition based maintenance, but the data is not used because it is difficult to interpret or because it lacks parameters for fault prediction. Implementing new sensors in power plant equipment can be difficult.

7.8.1 Introduction to the Plant Under Study

The Järvenpää power plant is combined heat and power (CHP) power plant owned by Fortum Power and Heat Oy and operated by Maintpartner Oy. The power plant is located in the city of Järvenpää in the Tuusula municipality in southern Finland. The power plant main boiler K1 is a fluidized bed boiler (BFB) with a fuel capacity of 76 MW. The K1 boiler was commissioned in 2013 and has a wide range of utilizable fuels, mainly wood biomass, peat and waste based fuels. The plant also has three natural gas boilers for peak- and reserve situations. Yearly the power plant generates 250-330 GWh heat for the district heating network and 100 GWh electricity.

The flue gas recirculation blower was selected as the monitored machine due to its easy accessibility and constant run status. The blower is used to recirculate scrubbed flue gas back to the boiler to lift the fluidized bed in the boiler in order to cool the bottom of the boiler furnace. The recirculation fan is not an immediately process critical component, but a failure can affect the combustion process and prolonged operation can reduce the boiler lifecycle by overheating the boiler grate.

The blower consists of three parts, the engine, bearing and impeller inside the impeller housing. Most common faults for any rotating machines are bearing failures, imbalance and misalignment. All of these causes have distinct vibration patterns that can be identified with proper vibration instrumentation. The pilot instrumentation consisted of vibration sensors installed to the blower bearing and a tachometer for measuring the rotation frequency of the blower. The monitored blower is shown in Figure 7.114.

Figure 7.114   Flue gas recirculation blower monitored in the pilot project.

7.8.2 Scope and Logic

The pilot project consists of instrumenting a flue gas recirculation blower at Järvenpää power plant with several different types of vibration sensors and creating a data collection and data storage system for the data. Also implementing data collection from the plant process data to the same data storage as the vibration data is studied. The goal of the project is generate a pilot condition monitoring system including installed sensors, data connection and collection and data analysis and display possibilities. The current planned pilot structure is presented in Figure 7.115.

The pilot has been divided into three phases for management purposes. The first phase consists of installation of the sensors and local data collectors at the site. The installation takes into account the current Finnish standards regarding vibration measurements and monitoring [PSK Standards Association, 2006, 2007]. In the second phase, the local data collectors are connected to the MANTIS data storage that is created following the MIMOSA data structure presented in the OSA-CBM and OSA-EAI standards [MIMOSA, 2010, 2014]. The data storage also utilizes the reference architecture [Mantis Consortium, 2016]. From the MANTIS data storage, the data is distributed to the individual systems. Some of the sensors also provide direct access via internet that is used in the first phases of the pilot to collect preliminary measurement data. The third phase focuses on the analysis of the data and failure prediction.

The pilot structure allows for comparison of several different sensor types to find the most applicable sensor for this type of condition monitoring if such a preference can be made. The standardized data structures for data communication and storage provide a basis for collaborative use and development of the MANTIS platform.

The analysis and failure prediction study of the pilot is focused on rotating machines. The applied techniques will be specified later in the project after a preliminary technical study is complete and the data collection system provides preliminary results for analysis.

As several partners already have commercial sensor and data collection solutions available, these are used for maximum benefit in order to focus the research work to areas that have not yet been studied. The main focus of the pilot is the collaboration of different standardized solutions and protocols and the utilization of the collected data to create new value. The data analysis and failure prediction is a relatively new area that is not implemented in the commercial systems and will require
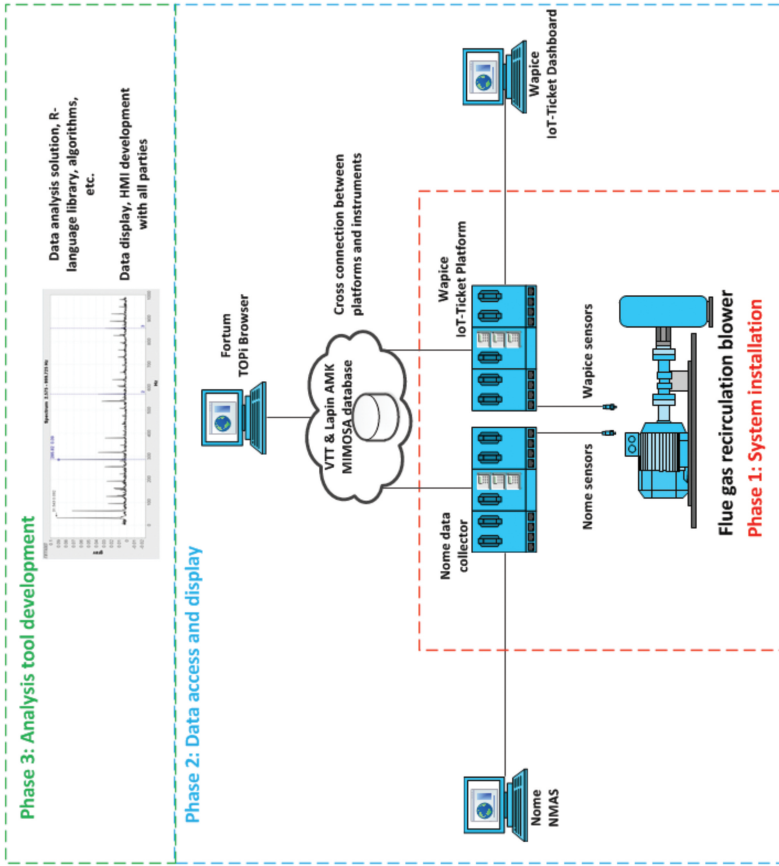
Figure 7.115   Conventional energy production use case platform.

research on the analysis methods, not only implementation and collaboration research.

7.8.3 Monitoring Rolling Element Bearings

Rolling element bearings are vital components of rotating machinery and they can be found in almost all rotating machines. In theory they are usually designed so that they should last the whole life time of the machine assuming that proper lubrication is provided and that no over-loading takes place. Unfortunately in real life it is rather typical that something goes wrong either with the lubrication or the loading and that initiates the wear of the bearing. Assuming that the outer or inner rolling surface has suffered, the wear takes place with increasing speed because the wear particles tend to cause further wear and worn surfaces cannot withstand the loading as well as new intact surfaces. Consequently bearing wear develops in exponential way. From financial point of view, it's important to know if the worn bearing will last until the next planned stoppage since in many cases it is very costly to stop the production in order to change one bearing. Naturally, one option could be to have redundancy i.e., a spare machine that could be used while the one suffering from bearing wear is repaired. However, it is easy to understand how costly this kind redundancy would be as in such a case two factories would be needed to do the work of one.

Traditionally the condition monitoring of bearings is carried out manually so that a trained professional is manually doing measurements once a month or every two weeks and possibly so that if something strange has been noticed in the measured signals the time period between consecutive measurements has been reduced to much shorter level e.g., once a day measurements. Clearly, this kind of manual monitoring is rather costly and takes a lot of effort in industry because of the high number of potential bearing failure objects. The aim is to automate the above described measuring process and also to be able to carry out the diagnosis automatically i.e., define whether a bearing fault is initiated with signal analysis and diagnosis based on artificial intelligence. In addition, the capability of predicting the remaining useful life is one of the objectives. In practice this means that we can predict when the latest date would be when the bearing has to be changed.

During recent years, the price of sensors and processors has reduced dramatically and this is the reason why three different type of vibration monitoring solutions for the detection of bearing wear have been tested. One of the tested solutions is the Nome nmas system, which is developed for this

type of purpose i.e., condition monitoring of machinery. The second tested solution WRM can be seen as a more generic platform that supports not only vibration measurement but also a wide range other kind of techniques. The third option is based on low cost components Mems sensor and RaspberryPi [Junnola, 2017] for processing the data, (see Figure 7.116). The idea is to find out how well this kind of solution performs when compared to more sophisticated equipment. The results of that comparison are discussed in chapter 4 of this book.

The most common technique today to diagnose whether a bearing fault is present or not is so called envelope detection, which is based on the detection of the vibration impulses that are caused when a rolling element hits the worn surface. The impulses vibrate at the first natural frequency of the structure in question i.e., at relatively high frequency 0.5–5 kHz. The frequency at which the impulses take place reveal which kind of fault is present. Is it outer or inner race or possible cage fault? Naturally the reliable detection of bearing faults at an early stage is the key action in condition monitoring. Quite a lot of effort is dedicated in being able to predict the development of the wear process. This is a very challenging task as there are so many factors that influence this phenomenon. For example, the loading and lubrication conditions together with the bearing geometry and material have an influence.
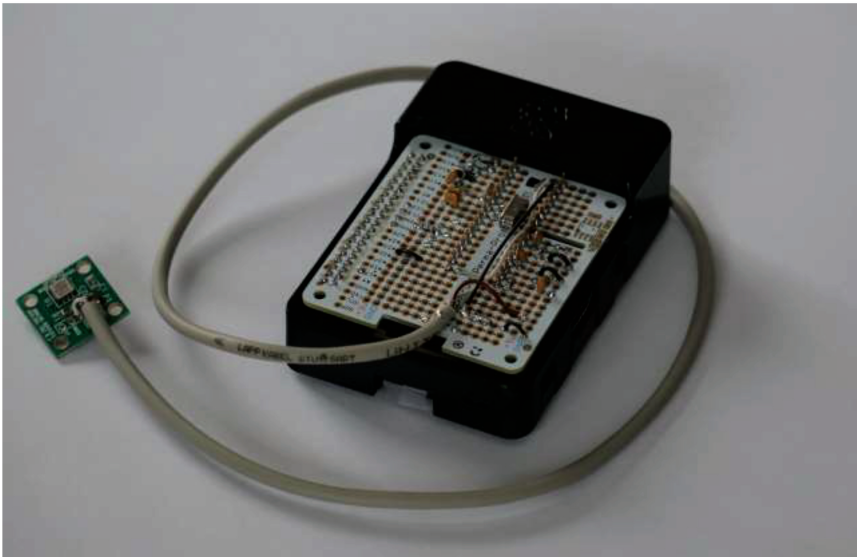

Figure 7.116   RaspberryPi based measuring system with a mems accelerometer.

The fact that the signal analysis indicators that reveal the existence of a bearing fault do not increase linearly but instead both increase and decrease at certain phases. The development in the end of life is so rapid that it makes the prognosis process challenging [38]. These techniques have been more thoroughly discussed in chapter 5 of this book.

With the introduction of new sensor types and processing power of current CPSs, the amount of data that needs to be managed increases dramatically and this in turn emphasises the role of the platform that is for this purpose. In this use case the idea is to use hierarchical data structure so that most of the data is processed locally and only meaningful information of exceptions is passed to higher levels. For example, the RaspberryPi processor is capable of carrying out the necessary signal analysis tasks together with the diagnosis of possible faults. The maths are programmed with Python which is an open programming language dedicated to mathematical programming. It should be noted that due the openness of the programming language a lot of useful material is available free of charge. The data is at all levels (locally e.g., RaspberryPI, plant, cloud -Azure in this use case-, service centre) managed with MIMOSA [Gorostegui, 2017]. This is again an open solution for maintenance related data. MIMOSA can hold data of the bearing type and geometry, its maintenance history (who, when, what . . .), measurement data, data to support the making of diagnosis and prognosis, results of diagnosis and prognosis. Basically, this is all data that is needed for a CMMS that is handled with MIMOSA. MIMOSA has also served the purpose of integration (installation in Kemi) between the various systems that the individual partners have been using. Mimosa is the platform for supporting the development of OSA-CBM Web Services that are developed for diagnostic and prognostic purposes.

7.8.4 IoT-Ticket Platform

One of the measurement platforms used in the pilot was industrial IoT platform IoT-Ticket and reference edge computing device WRM247+ both developed and owned by Wapice. For data connectivity IoT-Ticket offers several possibilities to connect into data sources. These are e.g., OPC, OPC UA, MQTT and other industrial standard communication methods. Custom connectivity is possible through ready-made developer libraries and REST API. Using the WRM247+ multi-purpose data collection and edge computing device it is also possible to execute vibration measurements. As an off-the-shelf solution for the vibration measurements it is possible to connect the

device into IFM VSE sensor gateway solutions. Connectivity to any industrial standard IEPE 4-20mA vibration sensor is built during the pilot phase in addition to existing methods. In pilot setup the WRM 247+ device was directly connected to vibration sensors using a signal condition amplifier. The benefit of this approach is that it gives a full control of all measurements that are done in the device. IMI 603C01 sensors are selected and connected to the system under test using magnetic connectors.

A support for computing RMS (Root Mean Square), Peak (Maximum Peak) and Crest (Peak/RMS) time domain analyses (KPIs) and FFT analysis for sampled signal is implemented and configured appropriately. The pre-processed data is then uploaded to IoT-Ticket server in regular intervals for further analysis and condition dashboards and reports (see Figure 7.117).

Using the Interface Designer and graphical flow programming tools the connection to MIMOSA database is implemented by creating the necessary flows to connect, read and analyze required data. From IoT-Ticket connectivity to MIMOSA was done using a standard flow-component that allows connectivity to external REST sources. As a parameter this component takes a combination of username and password, source URL and REST method (contains the XML/JSON payload). Virtual data tags allow forwarding REST response into IoT-Ticket's system. In order to post-process the data further several diagnostics flows are created to automatically monitor
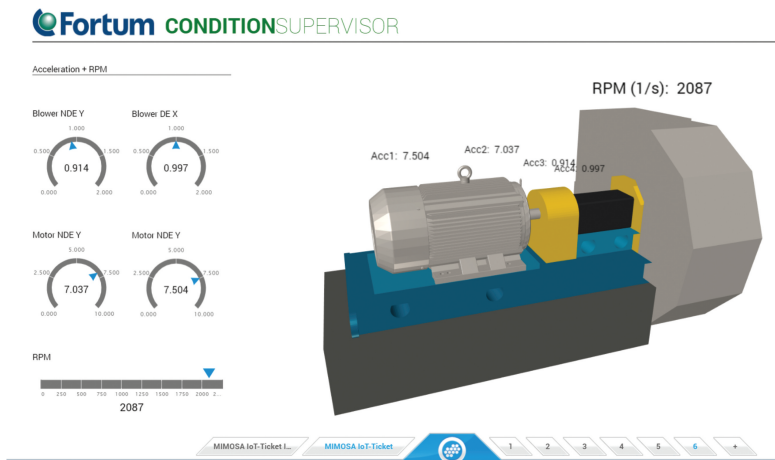

Figure 7.117   IoT-Ticket dashboard utilizing the 3D model built by LapinAMK. Live values are fused into 3D model.

the vibration levels. Data driven events are enabled by utilizing the IoT-Ticket alarms and reports feature. Automatic reporting was setup to trigger if something exceptional would happen in the diagnosed data. This could be for example an exceptional signal level in vibration measurements or a decision based on statistical information computed inside IoT-Ticket or one of the external prognostics service providers available through MIMOSA.

7.8.5 nmas Measuring System

Nome used and further developed the nmas monitoring system in the use case. The nmas monitoring system is developed and owned by Nome and is made for condition monitoring of rotating machinery. Connectivity to MIMOSA database is made through a REST API. The pilot case monitoring system included a local measurement unit capable of measuring, calculating, and storing data. A remote connection is build using 4G connection and locally the measurement unit is connected through WLAN.

Industrial standard 100mV/g IEPE acceleration sensors and optical tachometer are used for measurements. All measurement channels are measured simultaneously. Local device calculated velocity RMS (Root Mean Square) values and acceleration peak value continuously and stored time signals in defined intervals. Pre-Alarm and alarm limits are set according to ISO10816 standards.

Analysing of signals is done using nmas Analysator or View Java based analysing tools. A browser based viewing interface is made to view results with cellular phone or tablet. Viewing software and browser interface displays measurement trends, raw time signal, velocity spectrums, and alarm statuses. More sophisticated analysis including different mathematical functions, band-pass filters, window functions, different spectrums are available with analysis software. Basic view of nmas View and Analysator software is presented in Figure 7.118.

During the project Nome developed nmas Simple measuring device that is highly adaptive and easy to install condition monitoring system for local and remote monitoring of critical machinery. Nmas Simple measurement device is presented in Figure 7.119.

7.8.6 Mantis Cloud Platform

A Microsoft Azure based MIMOSA deployment is used as an information exchange platform. A REST interface into MIMOSA was developed by

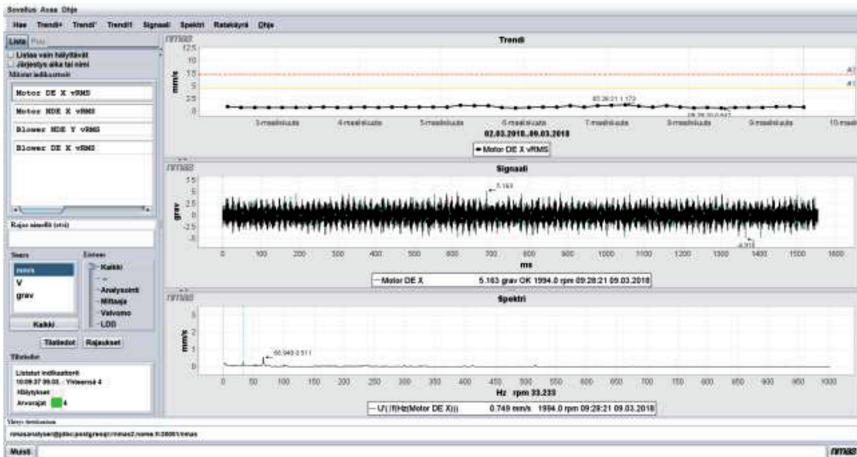**Figure 7.118**   nmas Analysator UI basic view.

Figure 7.119   nmas Easy condition monitoring device.

Lapland University of Applied Sciences (LUAS) in order to integrate various different measurement platforms together. It's been developed in JAVA using existing libraries such as Jersey2 and Jackson to provide basic REST and JSON functionalities. The REST interface is named MIREI short for

MIMOSA REST Interface. RESTful approach is chosen due to its simplicity, ease of use and bi-directionality. Any possible performance related downsides in the way REST operates is outweighed by the overall ease of integration and flexibility of the approach.

MIREI was initially released as two separate variants with slightly different HTTP command structures; the standard MIREI and the experimental MIREI. The standard MIREI provides helpers to make CRUD operations more streamlined and concise. It also can contain vendor specific data mappers that allow systems to store and retrieve native data from MIMOSA while still retaining MIMOSA compliancy. Such a mapper is developed for Nome's measurement system. The standard MIREI release however is only available for specific tables and does not allow access to other MIMOSA tables without further expansion. The MIREI experimental enabled access to all MIMOSA tables. This, however, requires better understanding of the MIMOSA data model to be of use. It does not contain any helper functions and will require users to fill in all the fields marked NOT NULL and adhere to the constraints existing within the data model. Though the new commands automatically fill the information related to row updates.

Later the two were merged into a single release enabling both functionalities, however the REST URLs are still kept separate. This is done in order to make it easier to access the REST commands in scenarios where it is necessary to use both the helpers and have a more complete access to the underlying MIMOSA database. Figure 7.120 shows the REST interface and its role in this use case.

The type of data inserted into the MIMOSA database is mostly focused on measurement data and generated data provided by the analytics, prognostics and simulation tools developed by VTT. Envisioned CMMS and ERP data integration is not completed, however the REST interface would make this possible. The MIMOSA database is used to store different types of location information used in augmented reality and virtual reality applications. This is accomplished by creating new data types to the MIMOSA and using the existing tables related to assets, segments and measurement locations.

HTTP Basic authentication is used to restrict access to MIREI. Partners received their own username and password for the system. There is also possibility to enable more secure token-based authentication for the MIREI. In the token-based authentication method the client needs to request authentication token from the server using the login information provided and then include this token in the REST headers.
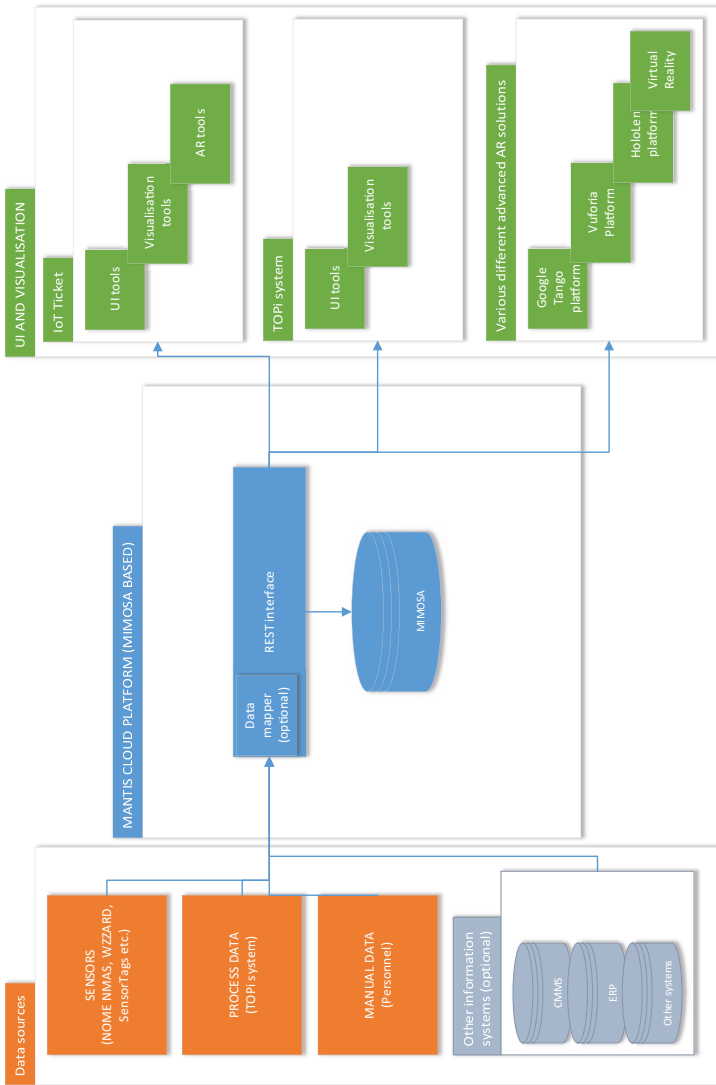
Figure 7.120 REST interface for this use case.

7.8.7 Data Analytics and Maintenance Optimization

The main optimization problem in power plant maintenance is to minimize the total lifecycle maintenance costs and to maximize the total lifecycle availability. This means that it could sometimes be beneficial to select "sub-optimal" operational modes to reduce maintenance costs or vice-versa. However these decions need to be made based on actual data that is carefully analyzed.

The success/failure of lifecycle optimization is measured by total plant availability and maintenance costs. However power plants are designed to operate for 25–40 years making lifecycle optimization difficult. In order to collect better data and thus make more informed decisions, better data collection and analysis tools need to be developed. Modern data collection, remote monitoring and analysis tools (neural networks, statistical analytics, physical models) also allow cost effective implementation of more advanced maintenance methods on less critical components.

7.8.8 Conclusions

This use case represents a quite normal situation that occurs in power plant environments. Power plants are long-term investments and house IT systems from several different vendors and technologies that need to be integrated with each other and need to be able to communicate. Each of the different technologies operate in their individual fields of expertise, but can have common elements such as data collection and databases.

This use case represents such an integrated system and the research done in the use case provides a reference architecture of how such a system can be built as well as benchmarking some of the open source technologies such as MIMOSA database and REST API needed for the integration.

Each of the partners also continued developing the individual components of the integrated system to provide improved analysis, connectivity and prognosis capabilities for the users. HMI and AR/VR is developed by the partners and is presented and discussed in Chapter 6.

7.9 Health Equipment Maintenance

Contributors: Jeroen Gijsbers, Mauro Barbieri, Verus Pronk, Hans Sprong, Jaap van der Voet, Godfried Webers, Karel Eerland, Marcel Boosten, Kees Wouters, Mike Holenderski, and Alp Akçay

This chapter provides an overview of practical application of several elements developed for the Health equipment use-case lead by Philips Healthcare. The chapter focusses on the most essential activities carried out by Philips Healthcare and its research partners, as the whole use-case description would go into too much detail for a chapter.

7.9.1 Introduction to Health Imaging Systems

Healthcare Imaging Systems are essential for the diagnosis and treatment of patients in hospital and private clinics (Figure 7.121). Due to the complexity of these systems and the large costs involved, it is not economically feasible to implement backup systems. Therefore, system uptime has to be maximized, planned downtime has to be minimized and unplanned shutdown has to be prevented. To cope with the exploding cost of healthcare, the cost of ownership has to be reduced, which also implies that maintenance budgets are under pressure. In response, Philips Healthcare has developed maintenance services for hospitals based on remote monitoring of their systems.

The biggest challenge there is to retrieve, store and analyze large amounts of data from globally distributed systems such that predictive maintenance

Figure 7.121   Unplanned system shutdown has a large impact on patients and hospital staff.

can be offered instead of maintenance at fixed time intervals. Furthermore, an alerting system is necessary when the online data analysis detects a threat of shutdown.

Due to the large purchase cost and the cost of housing, unplanned shutdown has a large impact on the hospitals and on the patients who may not get the care they need. Philips Healthcare made use of MANTIS Reference Architecture for equipment asset optimization, thereby aiming to move from a reactive to proactive and predictive maintenance.

The objective is to accurately predict upcoming failures by mining large amounts of data from heterogeneous systems distributed globally, such that maintenance can be timely scheduled or in urgent cases, the responsible person can be alerted. The main challenge is understanding how to get from large amount of data to accurate and precise failure detection and prediction.

Next to that, the availability of data and the analytical outcomes can give additional opportunities to exploit this information.

Every Healthcare Imaging Systems contains many sensors and generates large log files daily. Since these systems are heterogeneous by nature, the first challenge is to optimize logging such that data mining success can be optimized (anamnesis in Figure 7.122). The next challenge is to make all data available worldwide in the cloud (transport in Figure 7.122). Once the data is centrally available, it has to be translated to behavioral models and consolidated in a limited set of relevant parameters (translation in Figure 7.122). This translation requires significant computing power and storage space (infrastructure in Figure 7.122). Next, the obtained parameters have to be analyzed with respect to the maintenance challenges (analytics in Figure 7.122) and the results have to be visualized by end-users (visualization in Figure 7.122).

From the Healthcare Imaging Systems division of Philips two modalities participate in the MANTIS project: the *Magnetic Resonance* modality and the Interventional X-ray modality recently renamed to Image Guided Therapy Systems. They are introduced separately in the following sections.

7.9.1.1 Introduction to magnetic resonance

In the Business Unit Magnetic Resonance, medical *Magnetic Resonance* systems are developed (see Figure 7.123). These systems are mainly used to diagnose diseases. There is a variety of Magnetic Resonance system configurations to cover different magnetic field strengths, different gradient power strengths and different clinical application areas.

Anamnesis
Log reliability, service and utilization data. Collect all data at time of issue occurrence.

Transport
Reliable transport of logdata.

Translation
Build device behavior model from logfile data, consolidate in Nested Parameters

Infrastructure
Storage of parameters and running analytics scripts require storage capacity and CPU power

Analytics
Translate parameters to information required by Stakeholders.
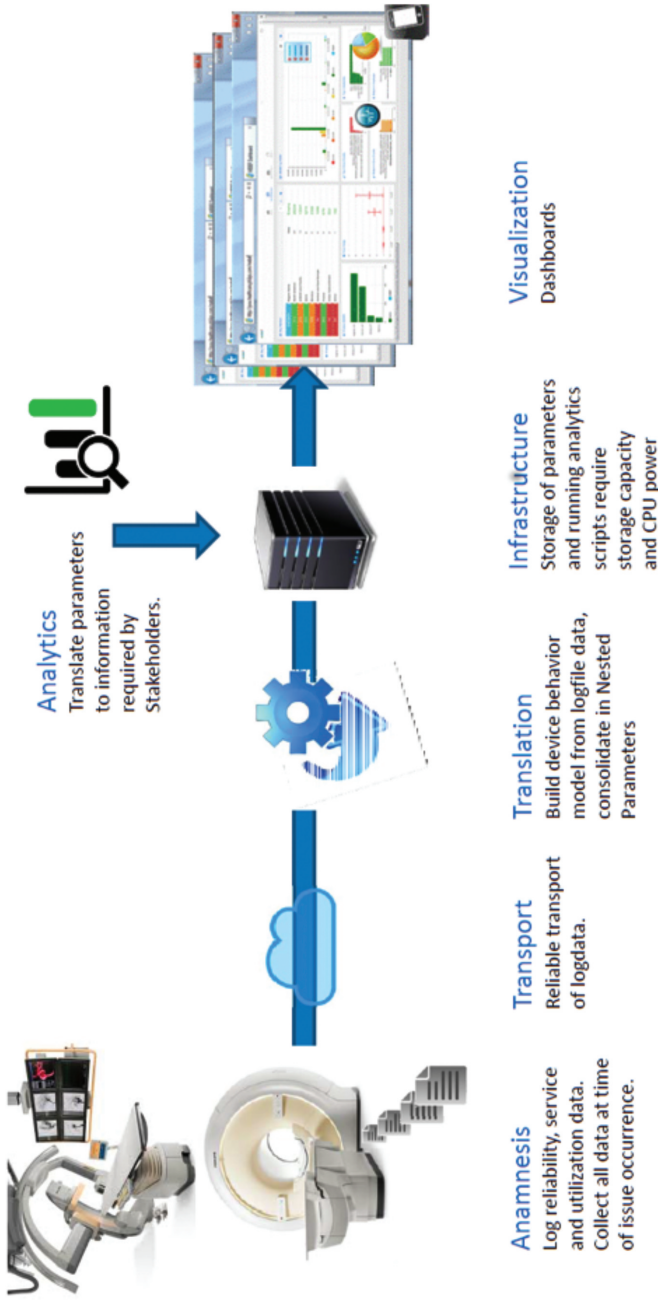
Visualization
Dashboards

Figure 7.122 Graphic depiction of the healthcare use-case.

Figure 7.123   Magnetic Resonance Unit.

A Magnetic Resonance system contains a cryogenically cooled superconductive coil that generates a static magnetic field. When the cryogenic cooling shuts down, the liquid helium level reduces and, within a few days, the superconducting coil loses its superconducting properties and results in a quench. The net effect is that the static magnetic field is lost and operation of the Magnetic Resonance system is no longer possible. It requires days to refill and ramp up the magnetic field back. Early detection of the loss or reduction of the cryogenic cooling function may stop the cascade of events that lead to a quench.

7.9.1.2 Introduction to IGT systems

In the Business Unit Image Guided Therapy Systems, medical X-ray systems in a C-arm configuration are developed (see Figure 7.124). The focus will be on the larger motorized 'fixed' systems (also called Cath labs), though some of the ideas may also be viable for the smaller "mobile" C-arms.

These larger X-ray systems can be used for diagnostics, but are most useful in minimal invasive interventional X-ray procedures like the treatment of coronary disease (Dotter treatment or stenting), structural heart disease (valve placement or repair), stroke treatment (aneurisms or stenosis), vascular disease (aortic aneurism or revascularization of limbs) and many other less known treatments. Because vital organs are often targeted, it is essential that the doctor can follow accurately what he is doing inside the patient's body. A wrong movement with a lead wire or catheter may cause serious harm to the patient or even death. Hence, IGT systems philosophy of avoiding interruptions of the image chain while there is a patient being treated. In the equipment of interventional X-ray systems, there obviously is a serious need for reliability.

Figure 7.124   Image Guided Therapy Unit.

7.9.2 Data Platform

The existing data storage platform is outdated, not scalable, unstable, and requires too much maintenance. The focus therefore is to transform the existing platform towards a platform that adheres to the MANTIS Reference Architecture. This is the foundation for all other activities in the project as well for future activities. Figure 7.125 shows a high-level overview of the implemented data architecture.

    Healthcare Imaging Systems can upload their data via the Philips Remote Services VPN. A Data Lake provides the high-volume storage required to
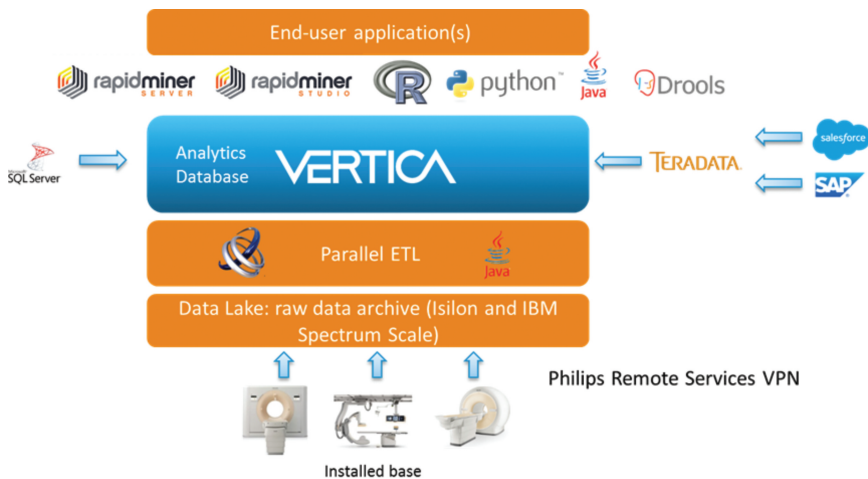

Figure 7.125   High-level overview of the data architecture.

store large amounts of historical device files. The device files in the Data Lake are processed by parallel ETL scripts to extract more useful information that is then stored in a Data Warehouse. The Data Warehouse is also the destination for structured data coming from administrative databases such as MS SQL Server, SAP and Teradata. Data analytics can be performed on the data stored in the Data Warehouse by using RapidMiner or programming languages such as R, Python or Java.

The next sections provide more details on the individual elements of the architecture.

7.9.3 Data Lake

Data from Healthcare Imaging Systems (*Installed base*) has been centrally collected by Philips via the Philips Remote Services VPN and stored in a Data Lake. The Data Lake is a component allowing low-cost secure storage of large amounts of data. The current capacity is of a few hundred terabytes but the system can be scaled out to multiple petabytes. It is realized using a Spectrum Scale storage cluster with GPFS file system and disaster recovery that is mounted as network shares on Linux and Windows machines.

Approximately 3 years of data have been stored in the Data Lake. Logically, the data lake space is divided in an archive and a live data-landing zone. The archive contains all the historical data available while the live data-landing zone contains the data received from medical devices that has still to be processed.

Both the archive and the live data-landing zone are further divided per imaging modality (e.g., *Magnetic Resonance*, IGT) and, within each imaging modality, each medical device has its own space. Within the reserved space for a medical device, the archive and the live data-landing zone are further divided per year, month and day.

7.9.4 ETL Scripts

Although it is possible to run analysis scripts on the Data Lake, this is not done very often due to the costs of developing parallel scripts and the lack of interactivity. The preferred way to analyze the data is by interactively querying a data warehouse that contains a pre-processed version of the information contained in the log files.

Log files come in many different file formats depending on the type of equipment, type of log file, type of device and release. For example, IGT equipment stores log events in textual format, parameters in a Microsoft Windows registry format, and configuration information in XML. Magnetic Resonance equipment produces log event files and parameters in textual format, XML, and proprietary binary formats.

For each log file format, a parser has been developed or adapted from existing tools to process the files and extract relevant data that can support root cause analysis and predictive maintenance. Given the diverse nature, formats and the large size of the log files from Healthcare Imaging Systems, ETL scripts have been developed to extract known, potentially useful, information from the log files. This information is stored in a Data Warehouse along with structured data to enable the scenarios. The ETL scripts are written in Java, Ruby, and Python and run on a cluster of computers that uses TORQUE, a distributed resource manager for cluster management, to parallelize the import operations. Each script is therefore written to support parallel processing and be resilient to failures.

To this aim, two aspects are very important: *idempotence* and ensuring correct *data provenance*. With *idempotence* we mean that an ETL script can be applied multiple times to the same input file without resulting in the data being imported more than once. This facilitates re-running ETL jobs that have failed or partially failed. With *data provenance,* we mean that every record in the Data Warehouse should have associated at least three key pieces of information:

- A reference to the Healthcare Imaging System, preferably a direct reference (e.g., serial number) that does not require joins with other tables;
- The file it originates from with full path and last modification timestamp;
- The version of the ETL script that created the record;
- The date and time at which the record was created.

7.9.5 Data Warehouse

For the architecture, we choose a distributed column-based storage solution (Vertica) that allows to store large amounts of data and to perform SQL queries as if it were a "standard" relational database. The main characteristics of the chosen data warehouse are the following:

- Distributed: this allows scaling-out when more storage or speed is required, increases robustness by replicating data on multiple nodes, increases speed of access by storing multiple copies of the data and by distributing load across nodes;
- Column-oriented: optimized for data access. Data is logically organized in tables as in traditional relational databases, though on disk, the data is stored "per column" instead of "per row". This allows speeding up queries by only reading the files of the columns involved in the queries;
- Advanced compression: aggressive compression of the data is used to replace slow disk I/O for fast CPU cycles. Because the data is stored in columns, different compression schemes can be used depending on the property of a column (e.g., type, cardinality, order) achieving extremely high compression rates;
- SQL-compatible: data can be retrieved using standard SQL queries and via ODBC/JDBC connections. This makes it easy for the applications and ETL scripts to upload data and for the users to retrieve it and to perform data analytics. Additionally, structured data from existing relational databases can be imported directly 1-1 without having to change the data models.

Before designing and implementing ETL scripts, domain knowledge experts and data scientists decide which information from device log files they consider useful for further analysis. This is typically done in an interactive way that involves interviews with R&D experts, gathering of documentation and specification documents of the Healthcare Imaging System, gathering and manual exploration of sample data as well as automatic analysis to determine: structure, data types, data value boundaries, etc.

A "data model document" provides the specification of the ETL task as well as the final format in which the data will be stored in the Data Warehouse. Identifying the dataset to import is also part of the specification (e.g., which files should be skipped or declared invalid).

After the data model document is approved by domain knowledge experts and data scientists, the ETL is designed, implemented (this may include porting or adapting an existing parser), tested and applied to the set of historical device files. The resulting data is then verified using basic analytics (e.g., checking data boundaries, number of records, etc.) and validated by data scientists and domain knowledge experts. The process is repeated until the desired quality level is achieved.

Note that the data in the Data Warehouse is typically written once and read multiple times. Unless the data needs to be corrected, updates and deletes will be done rarely. Furthermore, due to the column-oriented nature of the chosen Data Warehouse technology, the data models are de-normalized to achieve fast access and simplify data analysis.

The data model document has also the function of *data dictionary* documenting the definitions and the logic behind the records stored in the Data Warehouse.

The ingestion in the Data Warehouse of structured data from existing databases follows a similar process with the difference that instead of ETL scripts, data loading scripts are written that use JDBC/ODBC connections or simple CSV files to move sets of relational data from the sources to the Data Warehouse.

7.9.6 Sensors

To monitor the performance of critical components in a Healthcare Imaging System, intelligent components are developed. These intelligent components sense and record their state in real-time. For components that have a wear-out mechanism that would make them fail within the estimated lifetime of a system, the known wear can be removed from computation by means of calibration during planned maintenance activities. For the IGT equipment one of these calibrations is automated in real-time with the use of an intelligent function. There are more opportunities to automate calibrations in the future, but therefore the required feedback loops need to be in place, which is not always the case.

In the installed base, there are still many Healthcare Imaging Systems that lack the required sensing for critical components. For those devices an intelligent sensor, e-Alert sensor, has been developed. This stand-alone sensor has embedded sensors to measure physical properties of the equipment or environment, to process the signals and to send alerts. Such a sensor is not available in the Healthcare Imaging System itself. The sensor is connected to the healthcare facility network and can communicate via E-Mail and/or SMS. Figure 7.126 shows the context diagram of the e-Alert sensor. In case of unfavorable conditions that require a corrective action to resolve the issue, messages can be sent directly to customers as well as to service engineers. Next to that, the e-Alert sensor can be connected to the Philips Remote Service Network. The e-Alert sensor uploads sensor logs and alert logs to Philips Remote Service Network, where the data is stored and pre-processed.
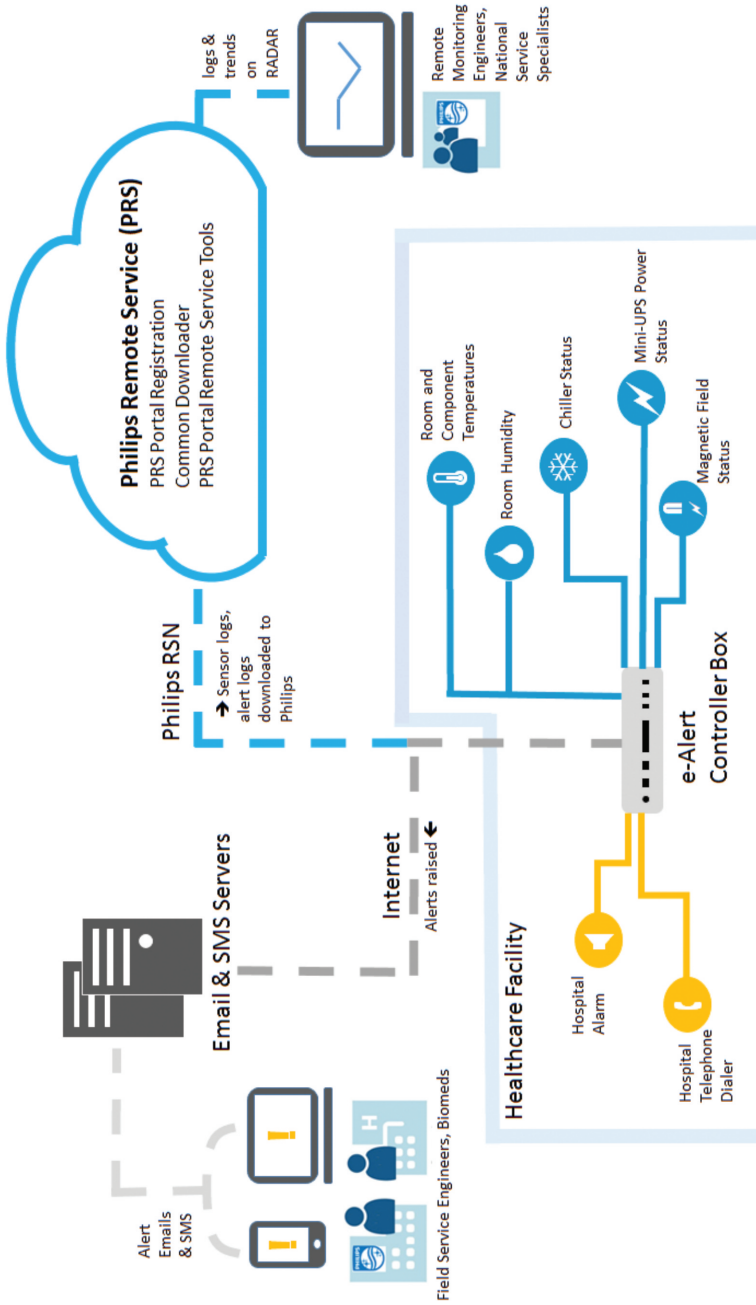
Figure 7.126 e-Alert sensor context diagram.

This data is accessible via a Philips Remote Service portal to enable Philips to determine operational profiles (aggregated or on e-Alert sensor level). This information is used to define control limits to keep the Healthcare Imaging System in optimal operational condition.

7.9.7 Analysis and Decision Making Functionalities

The next sections explain on high level some topics with respect to analysis and decision making functionalities. First, the scoring of predictive models using live data, data sources and data preparation steps for the Log Pattern Finder is explained. Then physical modeling of a unit that dominantly fails due to wear-out, and finally a mathematical model that optimizes the decision making process of remote service engineers in the presence of imperfect failure alerts.

7.9.7.1 Predictive model deployment and live scoring

The data ingested in the data warehouse is used to develop predictive models for particular failure modes. Predictive models are software programs that take as input "live" and "historical" data from a Healthcare Imaging System and calculate a probability of failure of a certain component or group of components within a given period. Predictive models can be based on simple and static rules and thresholds when the failure modes are well understood and can be easily modelled with the data provided by the Healthcare Imaging System. In the case of static rules and thresholds, historical data is typically used to choose the best threshold values that minimize the false positive rate while providing a high number of true positives.

Very often, simple and static rules and thresholds are not sufficient to predict failures with the desired level of accuracy. In these cases, statistical learning is applied. A machine-learning algorithm, such as a neural network or a support vector machine, is trained on a historical dataset until it reaches sufficient predictive performances on historical data.

Once a predictive model has been developed (and trained or tuned with historical data), the model is deployed in a *Quality Assurance* environment where it is scored daily with new data coming from the Healthcare Imaging Systems in the field. The results of the models, called "alerts", are stored in the data warehouse for being consumed by a web application called the "remote monitoring dashboard" where a team of remote monitoring engineers evaluates them for their accuracy and predictive power. During this evaluation, the remote monitoring engineers check whether an alert actually

corresponds to a situation of imminent failure using all the knowledge at their disposal. Furthermore, the remote monitoring engineers provide feedback to the model development team on the text and plots to be used in the alerts in order to make them more actionable and easier to understand. The model development team uses the feedback of the remote monitoring team to improve the predictive models until the alerts they generate in *Quality Assurance* are deemed good enough for being promoted to production. At this stage, a predictive model is deployed into production and its alerts are displayed in the production remote monitoring dashboard. These alerts are used to create the actual proactive cases for Healthcare Imaging Systems in the field.

7.9.7.2 Log pattern finder data

In the previous section, we have seen how predictive models can be developed and used within the data processing architecture. A particular set of models that is useful in this context are the so-called log patterns. A log pattern is a logical sequence of log events that correlates with a particular failure mode. Ideally, we would like to be able to discover log patterns automatically using data. In this section, an approach for automatically finding log patterns is described.

7.9.7.3 Data sources

For the log pattern finder for IGT equipment, we make use of various data sources. The primary data source is the *calls* data source. As the objective is to find reactive patterns, we make use of a calls table to identify (*i*) the time of the call, (*ii*) the system to which the call applies, (*iii*) the parts that were replaced, if any, and the log events that were generated by the identified system during a time window prior to the call. These four data sources are linked as shown in Figure 7.127, where an arrow from data source A to B indicates that one or more fields from an entry in A are used to identify the proper entry or entries in B.

For any call, there is always a single associated Healthcare Imaging System, but there may have been various parts replaced, depending on the outcome of a root-cause analysis by the service engineers. There are also calls where no part has been replaced, but other actions have been undertaken to resolve the issue. We do not consider these and concentrate on those calls where at least one part has been replaced.

For collecting the appropriate set of log events, we use various methods, ranging from taking a fixed *observation* interval of n days prior to the call in
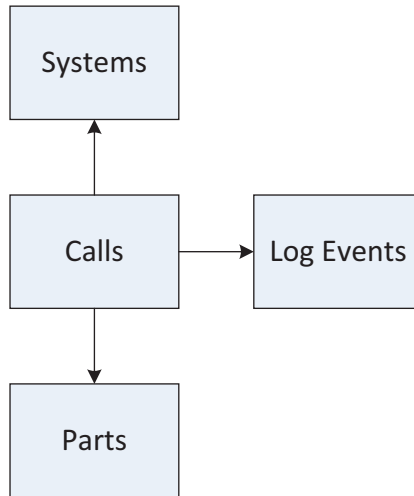
Figure 7.127　Overview of the main data sources used.

order to retrieve a set of log events to more carefully selecting an observation interval containing a sufficient number of log events.

7.9.7.4 Inspect and normalize the data

It was necessary to invest quite some time into getting to "know" the data. Blindly applying an algorithm to the data is usually not a good idea. For example, for log events only a portion of the available fields is used to obtain a concise representation of an event. This approach, however, results in millions of different log events, a situation that is not appropriate in the current context. We discovered that event ids with multiple descriptions are present. The main reason for this is that these descriptions contain numbers, dates, times, ip-addresses and such. As these numbers are most likely not interesting, a *normalization step* is applied to reduce the number of different descriptions. A specific filter has to be created for each specific issue. These filters are handcrafted, based on the inspection of numerous descriptions.

　　As another example, there is the issue of dates and times. As the Healthcare Imaging Systems are located in many different countries, one inevitably has to investigate how time is represented in the data coming from the various countries. This encompasses time zones, daylight savings time, and the use of local time without any time zone. Especially for identifying the proper log events, the call-open date is used, so the time information for log events and calls should be encoded in the same way.

7.9.7.5 Data pre-processing

As retrieving a complex set of data from various large databases can be quite time consuming, we decided to implement a pre-processing step to generate intermediate data, stored in easily accessible files. In particular, for a certain configuration (e.g., a time window, a selected set of system codes and system releases and the (maximal) length of the observation interval) we determine for each part replaced, during a call that opened in the given time window, details of all part replacements during this time window. These details include the system for which the replacement was done, the exact system release, various parameters and, finally, all the log events that occurred.

Although we do not consider calls wherein no part has been replaced, we do generate data for these calls, as their contribution in all calls is significant. We do this by creating a virtual part "NO_PART" and treating all calls without any part replacements as if this "NO_PART" has been replaced. By consistently doing this during the pre-processing step, we already prepare ourselves to look in more detail into these calls.

As this pre-processing is only done once for each configuration, much time is saved and the time required to run the experiments is greatly reduced.

7.9.7.6 Data representation

Finding log patterns entails combining the occurrence of combinations of log events during an observation interval and the replacement of a specific part during the associated call. To enable an efficient implementation, we decided to encode calls, together with their observation intervals, as integers, meaning that one integer is used to encode both a call and the associated observation interval. In this way, a part p can be represented by a set of integers, i.e., those that encode calls in which p has been replaced. A log event e can analogously be represented as a set of integers, i.e., those that encode observation intervals in which e has occurred. Note that log patterns in general can be represented in exactly the same way.

This representation of parts and log patterns allows efficient computation of all kinds of logical operations. For example, all calls wherein part p has been replaced and before which log event e has occured is represented by the intersection of their respective sets. A log pattern $AND(E_1, E_2)$ is also represented as the intersection of the representation of its two arguments.

Another advantage of this representation is that, by precomputing a *fingerprint* for each part and log pattern, checking equality can be done very efficiently. Concatenating the integer elements of a set in a string in increasing

order, separated by an appropriate character, allows string comparison to be used. This will be further elaborated upon in the next subsection.

In addition, log events are encoded to allow efficient computation. The encoding is by using integers, prepended with the letter E. This allows quick comparison of log events, as well as a concise notation of log patterns. It is noted that especially the additional info field can be as large as a few KBs, so that a concise representation not only aids in readability, it can potentially save a considerable amount of memory during computations.

7.9.7.7 Equivalent log patterns

When two log patterns have the same representation in terms of their sets or fingerprints, they are called equivalent. This is handy when looking for log patterns: of the equivalence classes that can be created by this equivalence relation, only one representative needs to be used, as the others will give the same results. This can significantly reduce the number of log patterns to be searched through when looking for good log patterns, and even more when, testing for a limited number of hypotheses, the total number of log patterns to be taken into account gets small.

Although two equivalent log patterns give the same results, this does not mean that they are equivalent in every sense. A service engineer may make a distinction between two log patterns based on the available information. It is also important to mention that this equivalence relation depends on the dataset at hand. On a different dataset, e.g., the test set, they may not be equivalent. Therefore, they may have differing performance. A service engineer may be able to identify the most suitable candidate from an equivalence class to act as representative. Of course, multiple log patterns from an equivalence class can be selected.

7.9.7.8 Log pattern selection problem

Once candidate log patterns have been identified, they will have a certain performance in terms of the number of true positives (TP) and false positives (FP) found in the training data. During the search for log patterns, these numbers have been subject to a number of constraints in order to generate log patterns of sufficient quality.

Part of the functionality of the log pattern finder is the *false-positive analysis*. Once, for a given part p, a log pattern with sufficient quality in terms of the number of true positives (TPs) and a sufficiently low false discovery rate (FDR) has been identified, an important and useful exercise is to investigate why a false positive (FP) ended up as such. This gives us a

better insight into the issues that caused these false positives and allows us to improve the log pattern finder.

7.9.7.9 Design decisions

In order to limit the wide spectrum of possibilities, we have made a number of simplifying assumptions and design decisions. The most important one is choosing a fixed observation interval length of one day. This has disadvantages, but we have experimented with longer intervals, and, although the results are different, they were not significantly better. Currently, investigations into choosing a proper observation interval length are ongoing. Note that the observation interval length could be chosen differently for different parts.

Another important decision is that we look at binary occurrences rather than frequencies of occurrences. In other words, either a log pattern occurs or it does not occur during an observation interval. We also do not consider the ordering of individual log events in a log pattern containing more than one log event. Yet another decision is not to apply processing on the description and additional info other than normalization.

Further, we restrict ourselves to log events that only occur a limited number of times, i.e., at most 400 times per year. This is a heuristic that we introduce to deal with the issue of significance of individual log events in a log pattern. We could have been stricter by also here considering p-values, but this is for further investigation.

Finally, we adopt a file naming approach in order to facilitate the management of these files. Although seemingly unimportant, we prepend all generated files with a timestamp, so that the order in which they are generated can be reflected in the directory and files are never overwritten. We use one timestamp for each individual run of the software, so that multiple runs can be performed in parallel.

7.9.7.10 Output

As output, we create files for individual parts and list all patterns found, their performance, their equivalents, as well as the results of the FP-analysis. We also report on the number of possible, allowed and actually generated hypotheses. In the end, the individual log patterns are combined into an overall log pattern, consisting of ORs of ANDs of individual log events. Several dozen of these log patterns are now actively monitoring thousands of Healthcare Imaging Systems daily and a few dozen are still under development.

7.9.7.11 Failure prediction

To come to a prediction, research is done on neural networks and their capability to predict failure of a specific component, a high power amplifier in equipment of the Business Unit Magnetic Resonance.

A neural network algorithm is used to predict failures in the high power amplifiers, based on system utilization data and power amplifier demands. For this purpose, utilization and demand data prior to an amplifier replacement was compared to utilization and demand data after an amplifier replacement. Data prior to amplifier replacement is considered as Failure data whereas data after an amplifier replacement is considered as Good data. The Failure data is gathered from a period of 17 days prior to device failure to 4 days prior to device failure. The Good data for period of 13 days is collected after 10 days of part replacement. Please refer to Figure 7.128 for a summary of the timeline for data collection. The cooling period masks the uncertainty in dates where the amplifier was actually installed and masks potential "burn-in" related failures.

The problem of fault prediction is here onwards addressed as a classification problem where, the model reads 13 days of historical data and predicts if it is Failure data or Good data. If it is predicted as Failure data, then it is more likely that the device is going to become faulty at least after 4 days and if it is predicted as Good data, the device is less likely to become faulty in the next few days. The dataset covers 219 amplifier replacements in a period of 1.5 years (July 2015 until the end of 2016). However, due to connectivity issues, we have got data for 134 systems only prior to amplifier replacement, and data of 154 machines only after amplifier replacement. Summarizing, we have 134 failure data points and 154 good data points. For each replacement, 16 features (F16) have been defined. Hence, the total data set consists of (134 [replacements] + 154 [Good]) * 16 [features] * 13 [days data/features] = 59.9k feature points. The data points are ordered as a single dimensional array (1D) (see Figure 7.129). The 1D arrays are formed by lexicographical

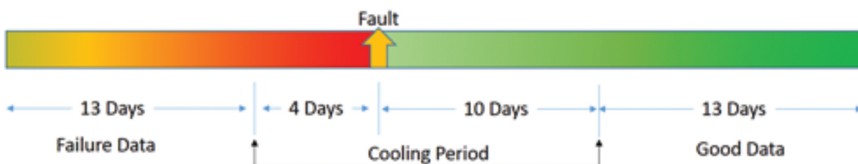**Figure 7.128**   Timeline of gathered categories of data from a single system.

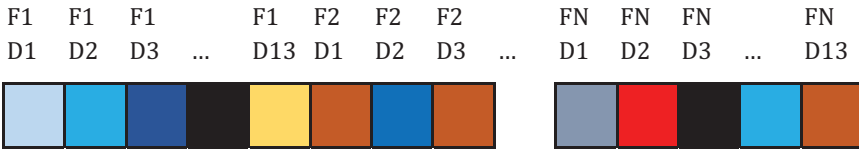| F1 | F1 | F1 | | F1 | F2 | F2 | F2 | | FN | FN | FN | | FN |
| D1 | D2 | D3 | ... | D13 | D1 | D2 | D3 | ... | D1 | D2 | D3 | ... | D13 |

Figure 7.129 Conversion of features into arrays.

ordering of the above features where F1 D1 denotes first feature recorded at Day 1.

The neural network used in this research is an Artificial Neural Networks (ANN), where the input is fed as a single dimensional array. The ANN used consists of 7 fully connected layers. The initial 6 layers consist of 50 neurons and a relu activation function and the last layer consists of two neurons and a softmax activation function. The network consists of two dropouts. The initial dropout in forward direction drops 25% of the features and the next dropout eliminates 50% of the features. This helps the network to develop a generalization that prevents overfitting. Once trained, the output layer of the model returns a probability of how it is likely that the device corresponding to the input data fails.

The data is split into training and testing sets at a proportion of 70% of samples and 30% of samples respectively. The ANN architecture, represented in Figure 7.130 was built and the training data was used to train it for 20,000 epochs. Figures 7.131 shows the learning curves (accuracy, loss, validation accuracy, validation loss) for increasing epochs of the ANN architecture.

It was necessary to train for 1500 epochs to attain the target accuracy of 95% over 1000 epochs. It can also be observed that the learning late curves are oscillatory in nature. This is because of the dropout which helps to reach a higher level of generalizability for the model. The resulting validation
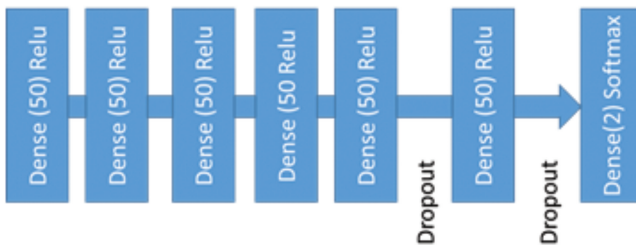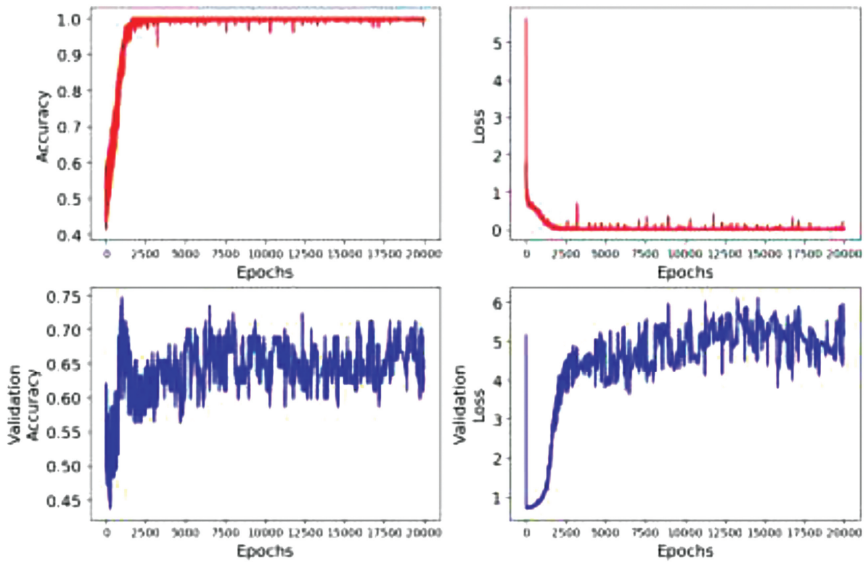
Figure 7.130 ANN Architecture.

Figure 7.131   Learning curves.

accuracy is 67%. Subsequently, the algorithm was fed with data from the years 2016+2017, covering more power amplifier replacements. The resulting validation accuracy is 54%; we could not achieve better validation accuracy. This is understandable since the number of amplifier replacements is still limited and deep learning architectures generally require thousands to obtain a generalised model. However, the ability of the deep learning architectures to perform satisfactorily on the limited data is promising. Additional work will be carried out to collect more data in order to achieve a validation accuracy of more than 90%.

7.9.7.12 Physical modeling

This section will reflect on the 10-step method used to come to a RUL prediction for the X-ray tubes. X-ray tubes are the most expensive parts to replace of IGT equipment and they are as such a major concern for the service organization. It is known that X-ray tubes are subject to wear-out, so it can be expected that they fail after some usage. Because of the major impact (in terms of downtime and cost) of an X-ray tube replacement, it is important to understand the failure of X-ray tubes better in order to improve the service to customers.

Step 1: Weibull analysis

As a starting point, do a Weibull analysis of the failures of the unit you want to make a RUL prediction for. The presence of a significant wear-out means that a RUL prediction is feasible. Next to working on the prediction of the wear-out, a second activity should be started to eliminate other failure modes as much as possible.

Step 2: Identify the responsible failure mode

The Weibull plot for the unit only tells whether there are failures due to wear-out, but not which is the related failure mode. After a candidate failure mode is found, a Weibull plot for just that failure mode should confirm that this is indeed the failure mode responsible for wear-out. As a rule of thumb, at least 80-85% of the failures should be related to wear-out in order for the prediction to be useful.

Step 3: Gather knowledge about the physics of the failure mode

X-ray tube cathode filaments wear, because they are heated to a high temperature during an X-ray run to emit sufficient electrons to produce the desired X-ray dose, but the heat causes the Tungsten, the material they are made of, to evaporate. Over time, a hot spot forms where heat and evaporation are exponentially increased until the material melts at the hot spot resulting in the opening of the filament. Understanding the stress and having a damage indicator at hand are two prerequisites for finalizing this step and continue with the next step in the method.

Step 4: Establish the relationship between stress and damage indicators and an end criterion in controlled experiments

Controlled experiments can be performed in the lab or in the field, but they usually require extra instrumentation. The controlled experiments will allow modeling a first order relationship between damage and stress for the unit, without the interference of secondary influences. These will appear later in the field, but can be recognized by comparing them with the first order model.

Step 5: Measure variables with a strong relationship to the stress and damage

Sometimes stress and damage can be measured directly, but often only indirect methods are available in the field. Make sure that the relation between the field data and the variables in the wear model are well established and understood.

Step 6: Collect, plot, and monitor the data

Once the field data becomes available, they can be collected and plotted. To be able to predict the amount of time that is left before a failure, you need to know how much wear is accumulated per time unit. A convenient way to represent the data is a (Damage/Load/Time) DLT plot. An example of a DLT-plot is given in Figure 7.132.

In Figure 7.132, the blue curve represents the ln(c-factor) against the linear wear (damage vs. load). This curve is called the *wear curve*. The orange curve represents the linear wear against calendar dates (load vs. time). It is remarkably straight for this particular DLT plot, indicating this user on average keeps on using the system in the same way every day over a long period of time. The blue curve is called the *usage curve*. At this stage, it is important to look for anomalies in the data, be able to explain these anomalies, and see if the failure is actually close when the wear curve starts to bend.

Step 7: Segment the usage curve

The usage curve in Figure 7.132 shows a relatively straight line for the linear wear against time plot. There are examples, where the usage curve changes significantly as can be seen in Figure 7.133.

When the usage changes, the best predictor for the future will also change. A change in usage depends on human decisions. In our case, it is related to the number of patients being treated on the system per unit time, the type of treatments, the availability of staff and all kinds of causes that would make the use of the system fluctuate. Therefore it is necessary to look for significant changes in the usage and adjust the prediction when such a significant change is detected. This can be done and tuned just based on the shape of the usage curves themselves.

In Figure 7.133, the dashed line represents the prediction of load at a particular time considering the usage change. In this particular case, the usage curve was segmented in two segments and the dashed line is based only on the second segment and obtained by means of linear regression.

Step 8: Collect sufficient wear curves to failure and establish an end criterion

When sufficient units have actually failed with reasonable certainty that they failed with the failure mode related to wear-out, the ends of their wear curves can be used to establish values for end criteria. In step 4, the variable for the end criterion is selected. In our case, this is the slope of the wear curve. When
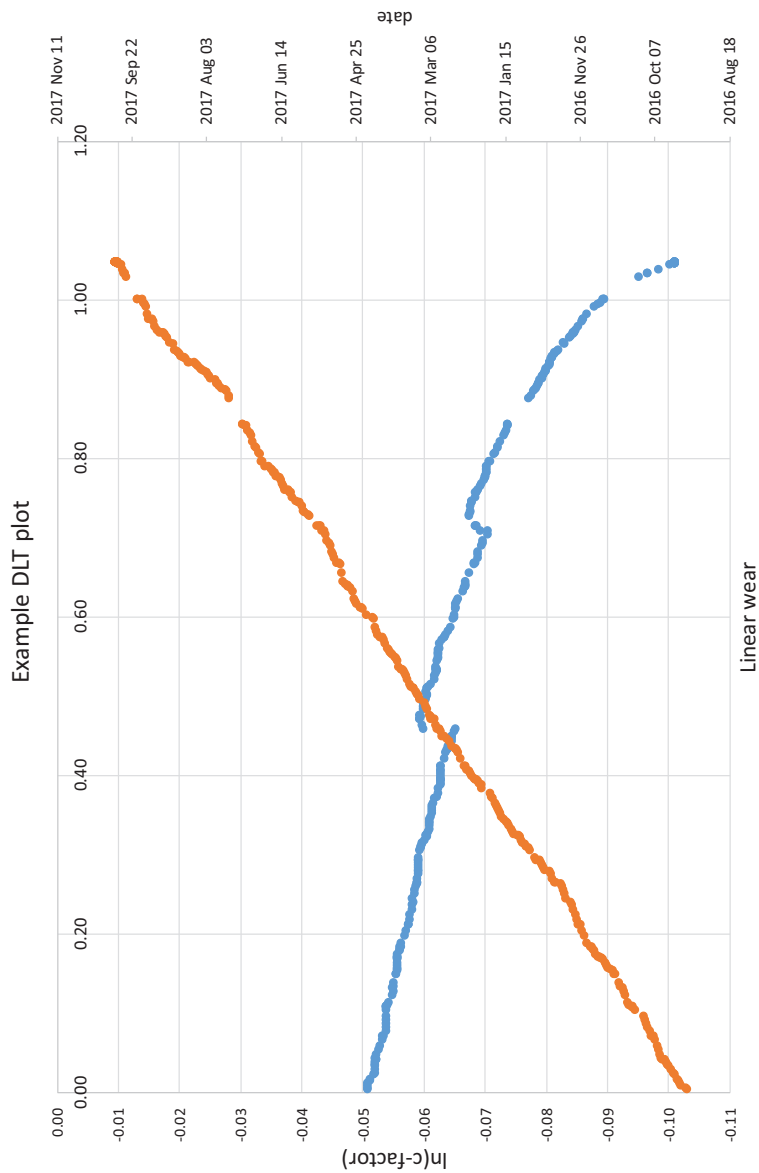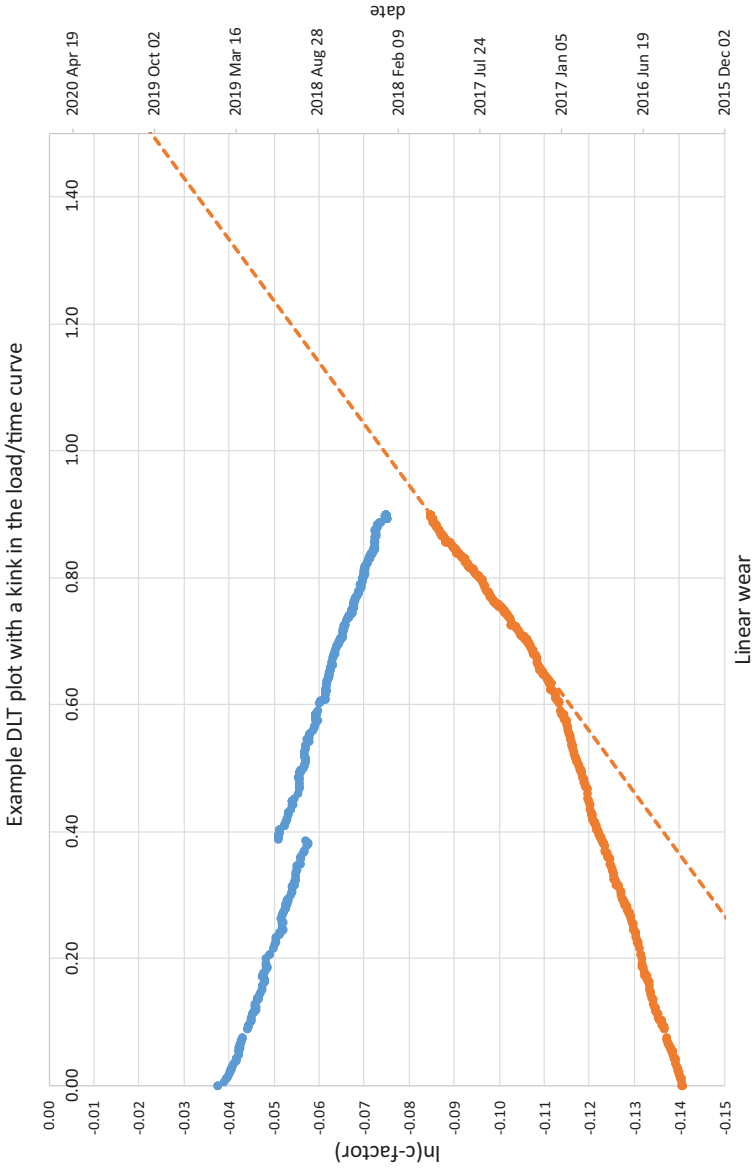
Figure 7.132 Example DLT plot.

Figure 7.133   Segmenting the usage curve.

sufficient units have worn out, an average value at which failure happens can be established. This may not be necessary in all cases, for instance there may be a limit value already defined (like a minimum profile depth of a car tire).

Step 9: Predict the wear curve progression for curves approaching failure

Once the end criteria are found, predictions can be made for units approaching failure.

Step 10: Assess the prediction stability

As time goes by, more data will be available and this will have influence on the prediction. At a certain moment, the prediction should however become more or less stable.

7.9.7.13 Maintenance and inventory optimization

Philips Healthcare is interested in developing a cost effective proactive maintenance strategy, relying on mathematical tools for statistical life-cycle and reliability analysis. The goal is to come to the optimal planning of maintenance and related resource planning. This consists of two different elements:

- Develop maintenance optimization models that can be used to make a balanced tradeoff between the cost of a failure and the cost of proactive maintenance, taking the uncertainty in the prediction models into account;
- Create a decision support system for remote monitoring engineers to come to the business optimal decision, and to guide them on what to do in case of imperfect predictions.

In addition to providing a recipe for the remote monitoring engineers to follow, we expect that the maintenance optimization model will also shed light on how much can be invested in improving the predictive/proactive models, i.e., the value of improved predictive/proactive models will be revealed by comparing the optimal expected costs under different levels of imperfectness in alert predictions.

7.9.7.14 Model and analysis

This section describes the creation of a mathematical model that supports remote monitoring engineers in their proactive maintenance decision making.

The model should support in the decisions that follow on an alert raised by predictive models. Whenever an alert is seen, a remote monitoring engineer needs to decide whether to create a case or to reject the alert.

Decision variables

As reported in Table 7.3, there are two decision variables in the model: a represents the decision to create a maintenance case and y represents the decision to combine the case with an already scheduled maintenance activity.

Table 7.3   Decision variables of the model

Decision variable	Notation
Initiate maintenance actions	$a \in \{0,\ 1\}$
Combine with next maintenance activity	$y \in \{0,\ 1\}$

Model parameters

Table 7.4 summarizes all the parameters used in the model to support the remote monitoring engineers.

Table 7.4   Parameters used in the model

| Parameter	Notation
No Facility Systems Engineer on site	$m \in \{0,\ 1\}$
No open case	$o \in \{0,\ 1\}$
Customer contract	$v \in \{0,\ 1,\ \dots,\ 6\}$
Outside working hours coverage	owh
Costs of downtime per unit time	c_d
Downtime compensation	$cc_{dt} \in \{0,\ 1\}$
Customer's region monitored by RME	$r \in \{0,\ 1\}$
Expected costs of PdM	c_{PdM}
Expected costs of PdM combined with PM	c'_{PdM}
Expected costs of CM	c_{cm}
Expected costs of diagnostics	$c_{diagnostics}$
SLA Response time for contract **V**	Tr_v
Estimated time for diagnostics	$t_{diagnostics}$
Estimated repair time	t_r
Time to next Planned Maintenance	t_{sm}
RUL	X
Time to on-site maintenance	T_{os}
Probability that the alert is true	P
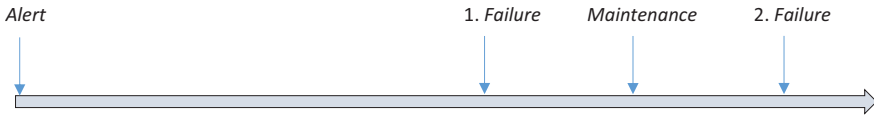
Figure 7.134 Timeline of Events (1: failure occurs before maintenance, 2: maintenance occurs before failure).

Timeline of events and probabilistic scenarios in the model

The model includes two random variables, X and T_{os}. These variables measure the time from the alert arrival to the failure and on-site maintenance, respectively (see Figure 7.134). When the remote monitoring engineer decides to create a case, the local service organization schedules the maintenance activities to resolve the case. Since X is random, we have to distinguish several scenarios when a case is created (see Table 7.5).

A proactive case is scheduled on T_{os} when $a = 1$, $y = 0$, and $T_{os} < t_{sm}$. We assume that if the realization of T_{os} is greater than t_{sm}, the LSO makes the decision to combine the case with the already scheduled maintenance case on t_{sm}. It makes no sense to execute the case later than this moment because it will lead to an additional visit and costs. Therefore, a proactive case is scheduled on t_{sm} when $a = 1$ and $y = 1$ or when $a = 1$, $y = 0$, and $T_{os} > t_{sm}$.

The equipment can fail before the case is solved. This happens if $X < T_{os}$ when the case is scheduled on T_{os} or if $X < t_{sm}$ when the case is scheduled on t_{sm}. The downtime is equal to the response time plus the repair time when the equipment fails before the case is solved; corrective maintenance (CM) costs are incurred.

The proactive maintenance case can also prevent a failure. This happens when $X > T_{os}$ or $X > t_{sm}$. The downtime is equal to the repair time and proactive maintenance costs are incurred. Costs of c'_{PdM} are incurred if the proactive case is combined with another case, and costs of c_{PdM} are incurred when the proactive case is not combined with another case.

When no case is created and the equipment fails, CM and diagnostic costs are incurred because the local service organization did not receive a case. The problem needs to be diagnosed first because the problem is unknown, when the customer calls. In this situation, the downtime consists of response time, time for diagnostics and repair time.

When the remote monitoring engineer decides to create a case, it is always possible that the alert was false. The local service organization discovers that the alert was false and costs of c_{FP} are incurred in such a scenario.

Table 7.5 Possible scenarios for the maintenance cases that are created

Action	Alert Realization	Probability on alert realization	Scenario realization	Probability on scenario realization	Costs incurred	Downtime incurred
	True Positive	P	$X < T_{os}, T_{os} < t_{sm}$	$P \cdot \Pr(X < T_{os}, T_{os} < t_{sm})$	c_{cm}	$Tr_v + t_r$
			$X > T_{os}, T_{os} < t_{sm}$	$P \cdot \Pr(X > T_{os}, T_{os} < t_{sm})$	c_{PdM}	t_r
			$X < t_{sm}, T_{os} > t_{sm}$	$P \cdot \Pr(X < t_{sm}, T_{os} > t_{sm})$	c_{cm}	$Tr_v + t_r$
Case			$X > t_{sm}, T_{os} > t_{sm}$	$P \cdot \Pr(X > t_{sm}, T_{os} > t_{sm})$	c'_{PdM}	t_r
	False Positive	$1 - P$		$1 - P$	c_{FP}	0
	True Positive	P	$X < t_{sm}$	$P \cdot \Pr(X < t_{sm})$	c_{cm}	$Tr_v + t_r$
Combine Case			$X > t_{sm}$	$P \cdot \Pr(X > t_{sm})$	c'_{PdM}	t_r
	False Positive	$1 - P$		$1 - P$	c_{FP}	0
SNAR (Seen No Action Required)	True Positive	P		P	$c_{cm}+ c_{diagnostics}$	$Tr_v + t_r + t_{diagnostics}$
	False Positive	$1 - P$		$1 - P$	0	0

Mathematical formulation of the model

After an alert arrival, the remote monitoring engineer has to make the decision such that the expected costs and downtime are minimized. We define two objective functions. The first one aims to minimize the expected costs, and the second one is used to minimize the expected downtime:

$$\text{Min } E[C] = a \cdot (1-y) \cdot E[C_{case}] + y \cdot E[C_{combine}]$$
$$+ (1-a) \cdot E[C_{SNAR}] + cc_{dt} \cdot E[D] \cdot c_d \qquad (7.11)$$

$$\text{Min } E[D] = a \cdot (1-y) \cdot E[D_{case}] + y \cdot E[D_{combine}]$$
$$+ (1-a) \cdot E[D_{SNAR}] \qquad (7.12)$$

s.t.

$$a \leq r, \ o, \ m, \ v \qquad (7.13)$$

$$y \leq a \qquad (7.14)$$

$$a, \ y \in \{0, \ 1\} \qquad (7.15)$$

Equation (7.4) represents the minimization objective of the expected costs because of the decisions made by the remote monitoring engineer. $E[C_{case}]$ represents the expected costs of creating and sending a case to the local service organization ($a = 1$, $y = 0$). $E[C_{combine}]$ represents the expected costs of creating a case and suggesting to combine it with an already scheduled case ($a = 1$, $y = 1$). $E[C_{SNAR}]$ represents the expected costs of SNAR ($a = 0$). The expected downtime costs are represented by $cc_{dt} \cdot E[D] \cdot c_d$ and are only incurred when the customer is entitled to downtime compensation ($cc_{dt} = 1$).

We can calculate the expected costs of each action by summing the multiplications of scenario probabilities for that action with the associated costs. These expected costs expressions for the different actions are given below:

$$E[C_{case}] = P \cdot (c_{cm} \cdot (\Pr(X < T_{os}, T_{os} < t_{sm}) + \Pr(X < t_{sm}, T_{os} > t_{sm}))$$
$$+ c_{PdM} \cdot \Pr(X > T_{os}, T_{os} < t_{sm}) + c'_{PdM}$$
$$\cdot \Pr(X > t_{sm}, T_{os} > t_{sm})) + (1-P) \cdot c_{FP} \qquad (7.16)$$

$$E[C_{combine}] = P \cdot \left((Tr_v + t_r) \cdot \Pr(X < t_{sm}) + c'_{PdM} \cdot \Pr(X > t_{sm})\right.$$
$$+ (1-P) \cdot c_{FP})$$
$$E[C_{SNAR}] = P \cdot (c_{cm} + c_{diagnostics}) \qquad (7.17)$$

Equation (7.5) represents the minimization objective of the expected downtime. $E[D_{case}]$ represents the expected downtime of creating and sending a case to the local service organization ($a = 1$, $y = 0$). $E[D_{combine}]$ represents the expected downtime of creating a case and suggesting to combine it with an already scheduled case ($a = 1$, $y = 1$). $E[D_{SNAR}]$ represents the expected downtime of SNAR ($a = 0$). We can calculate the expected downtime of each action in the same way as the expected costs. The expected downtime expression for each action are given below:

$$[E[D_{case}] = P \cdot ((Tr_v + t_r) \cdot (\Pr(X < T_{os}, T_{os} < t_{sm}) + \Pr(X < t_{sm}, T_{os} > t_{sm})))$$
$$+ t_r \cdot (\Pr(X > T_{os}, T_{os} < t_{sm}) + \Pr(X > t_{sm}, T_{os} > t_{sm})) \qquad (7.18)$$

$$E[D_{combine}] = P \cdot (c_{cm} \cdot \Pr(X < t_{sm})) + t_r \cdot \Pr(X > t_{sm})$$
$$E[D_{SNAR}] = P \cdot (Tr_v + t_r + t_{diagnostics}) \qquad (7.19)$$

Equation (7.6) makes sure that an alert is SNARed if, the customer's region is not monitored ($r = 0$), the customer has no contract ($v = 0$), the Facility Systems Engineer is already on-site ($m = 0$), or there is already a maintenance case opened for the system ($o = 0$). Alerts with such characteristics should be SNARed automatically.

Equation (7.7) enforces that a case can only be combined with an existing case when the remote monitoring engineer decides to create a case for the alert.

Equation (7.8) ensures that a and y can only take binary values.

7.9.7.15   Results and insights

The model is implemented in a case study for flat detectors, which converts X-ray into electrical signals. Since we have two objective functions, there is not always a single solution for the optimization problem. Lower costs can result in higher downtimes. The remote monitoring engineer can take three different decisions: (*i*) SNAR the alert ($a = 0$), (*ii*) create a case ($a = 1$, $y = 0$), or (*iii*) create a case and combine it with an already scheduled case ($a = 1, y = 1$). The model aims to evaluate all three options in terms of expected costs and expected downtimes. The output of Model 1 consists of a summary of each option with the expected costs and downtime of each option. This gives the remote monitoring engineer support in their decision making because they can account for the possible consequences of their decisions.

With the Flat-Detector-specific default values, creating a case is the optimal decision to make by the remote monitoring engineers. In Figure 7.135, we see that the optimal action outperforms the other actions in both expected costs and expected downtime at a specific value of the model parameter *P*.

Notice that Figure 7.135 is made for a credible alert with $P = 0.8$. If we use the same input values but set P to 0.15, we receive the plot in Figure 7.136. It can be
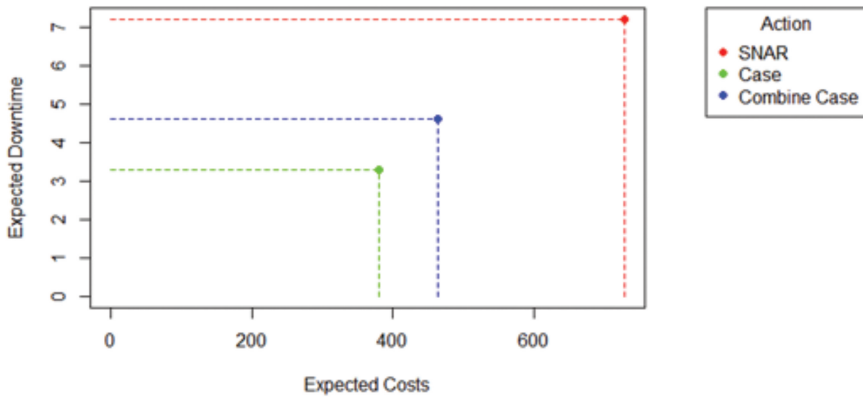
Figure 7.135 Model output default values for flat detector with P=0.80.

seen that there is no optimal decision to make. No action outperforms all others in terms of both expected downtime and expected costs. Create a case is the best option in terms of costs while combining the case is the best option in terms of downtime.

If we vary P from 0 to 1 with steps of 0.01, we receive the plots in Figure 7.137 and Figure 7.138. We observe the existence of a probability threshold for creating a case.

We next evaluate the influence of the service contracts on the optimal decisions according to the model. We create three fictitious customers, which we refer to as Customer A, Customer B and Customer C. Customer A has a contract with the most extensive entitlements. Customer B has a contract with no coverage options. Customer C has the most basic service contract.
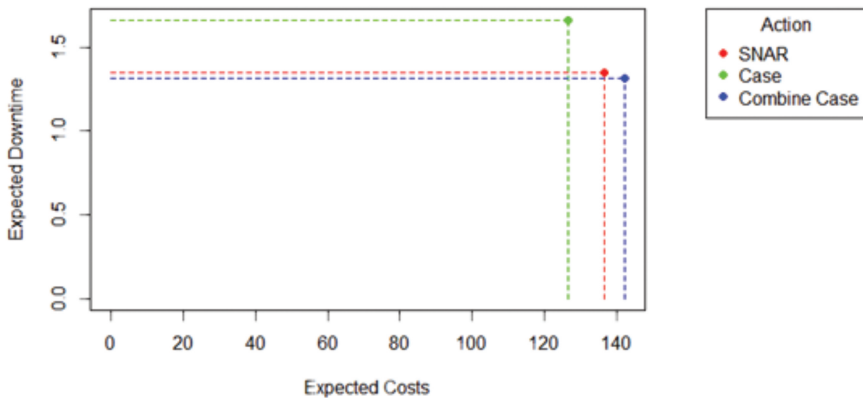

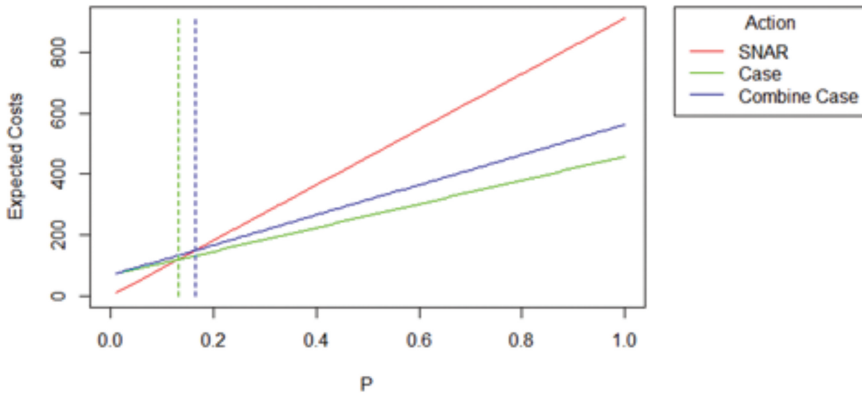**Figure 7.136** Model output default values for flat detector with P=0.15.

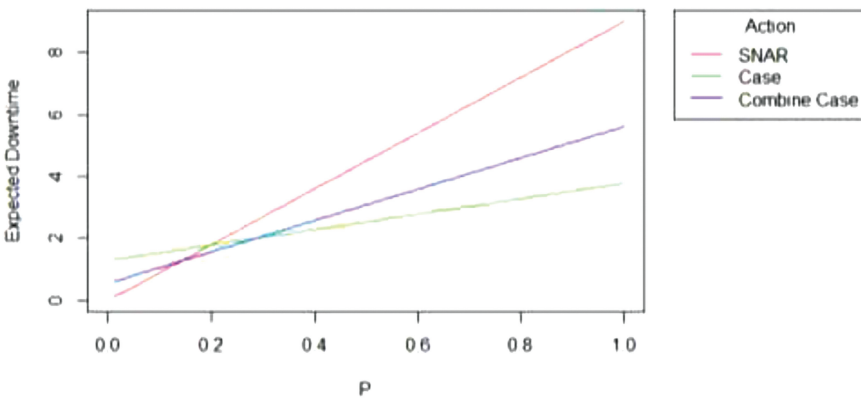**Figure 7.137**   Influence of P on expected costs in flat detector case.

Figure 7.138   Influence of P on expected downtime in flat detector case.

We vary P to find out if different probability thresholds exist for different types for customer. Figure 7.139 shows the influence of P on the expected costs and downtime for the different types of customers.

We observe that the expected costs incurred for Customer A are the highest. This is due to the compensation of downtime received by this type of customer. The expected downtime of all actions is the lowest for Customer A. The reason behind this is that shorter on-site response times are offered to customers with higher contracts. After a customer call, the field service engineer is faster on-site to conduct maintenance on the failed equipment. In addition, the customer is entitled
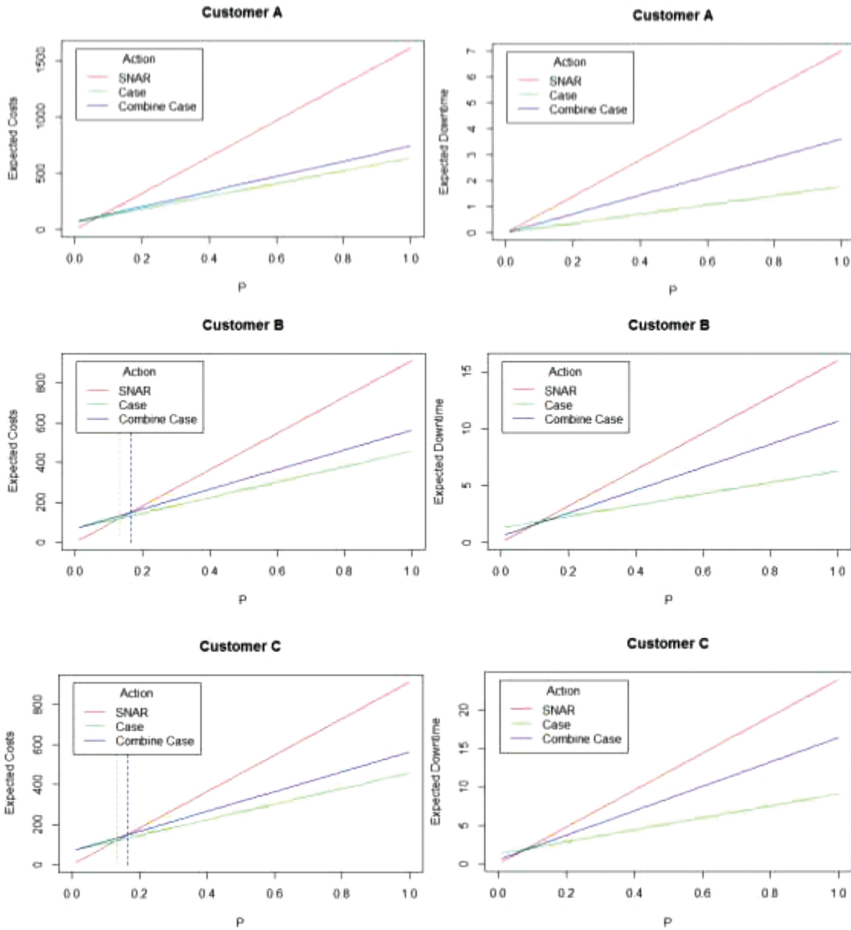
Figure 7.139 Influence of P on expected costs and downtime for different customers.

for maintenance outside operating hours. For Customer B and C, the expected costs of each action are equal. The value cc_{dt} is the only customer-specific parameter that influences the expected costs.

7.9.7.16 Visualization and HMI
Two HMIs were created related to the Philips Healthcare use case. One related to the development of the e-Alert sensor and one to assist the remote monitoring team. Both are explained in the sections below.

7.9.7.17 E-Alert portal

The e-Alert sensor provides a web-based user interface to configure sensors and to configure the control limits. When the healthcare facility allows it, the staff of the healthcare facility can access the user interface of the e-Alert sensor. This user interface provides capabilities to view the history of sensor values to support root cause analysis. Service engineers can also access the user interface of the e-Alert sensor. The user interface provides capabilities to view the history of sensor values, to configure the control limits and to update the embedded software.

Each e-Alert sensor is able to upload its sensor logs and alerts to the Philips Remote Service Network. A portal (see Figure 7.140) has been developed to gain access to these logs and alerts for offline data analysis. This enables Philips to determine operational profiles, specific to the Healthcare Imaging System where the e-Alert sensor is connected. This information can be used to fine-tune the configured control limits for that specific Healthcare Imaging System to keep it in optimal operational conditions.

7.9.7.18 Remote monitoring dashboard

A dashboard to created, to provide an overview of all Health Imaging Systems that generated an alert for the remote monitoring engineers. On the highest level of the dashboard an overview of all alerts is presented, as shown in Figure 7.141.

When one of the alerts is selected, a detailed report is presented (see Figure 7.142). This report contains all details for the related Health Imaging System

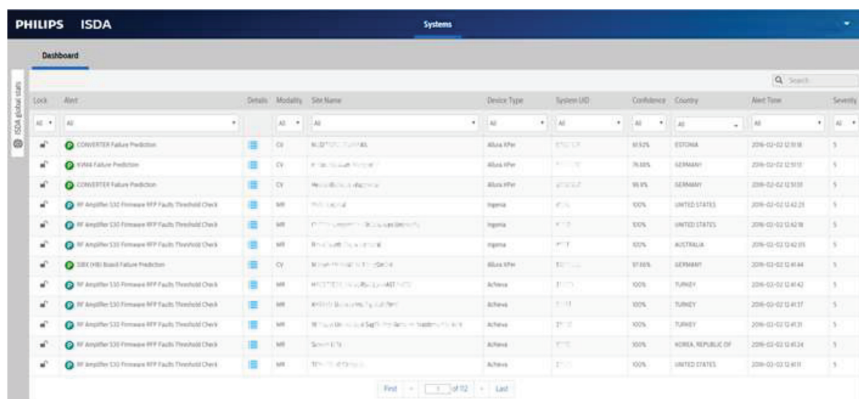**Figure 7.140**   Portal to gain access to e-Alert sensor data.

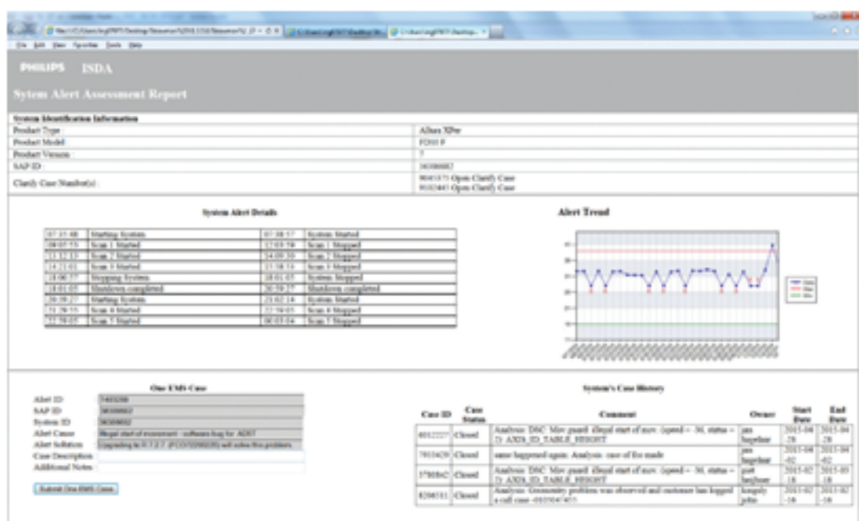**Figure 7.141** Remote monitoring dashboard high level overview.

Figure 7.142 Remote monitoring dashboard detailed report.

like system type, model version, software release, alert history, parameter trends, and maintenance history. This will support the remote monitoring engineer to be able to take further action. When further action is required, the remote monitoring engineer can directly initiate a service work order. Depending on the urgency of the service action, the engineer can combine it with an already scheduled appointment.

7.9.7.19 Conclusions

For the Philips Healthcare use-case, quite a lot of effort has been spent on the data storage platform and the analysis and decision making functionalities of the Mantis project. Having a large set of historical data already available, made it possible to make a lot of progress from the start. It also made the cooperation much smoother with both our main research partners, Philips Research and the Technical University of Eindhoven, because from an early stage we were able to share data. In our case, the misconception that only providing data eventually will lead to results is once again proven true. Without decent domain knowledge, it is virtually impossible to get usable results. Besides explaining the research partners what all the available data means, domain knowledge is also needed to review and filter possible outputs. This means that in the beginning intensive collaboration is required to get the research partners up-to-speed, followed by frequent updates. This is something that needs to be taken into consideration.

The Philips Healthcare use case featured a few challenging topics and goals set at the beginning of the project. Most of them were realized. The data storage platform is mature and ready to be extended, sensors and intelligent functions have been designed, and a considerable amount of predictive and proactive models has been created. All the different aspects combined, resulted in a remote monitoring capability. We envision that, from 2018 onwards, one in every five system service events worldwide will be triggered by careful analysis of system data – and will therefor take place before any major issues arise. This maintenance can also be planned so there is no disruption to the workflow.

References

Aaron Parecki, C. M. 'OAuth 2.0,' [Online]. Available: https://oauth.net/2/. [Accessed 01/04/2018].

Albano, M., Ferreira, L. L., Pinho, L. M., and Alkhawaja, A. R. (2015) Message-oriented middleware for smart grids, *Computer Standards & Interfaces*.

Apache, 'Spark Apache,' [Online]. Available: https://spark.apache.org/. [Accessed April 2018].

The Apache Software Foundation. (2018) 'Apache NiFi,' [Online]. Available: https://nifi.apache.org/. [Accessed 01/04/2018].

Carroll, J. M. (2002) *Making Use: Scenario-Based Design of Human-Computer Interactions*, The MIT Press.

Cheung, D. W., Fu, A. W.-C., and Han, J. (1994) *Knowledge Discovery in Databases: A Rule-based Attribute-oriented Approach*, Berlin: Springer.

Cocciaglia, D. (2012) 'Case Study: Switch & Crossing Diagnostics,' in *Track Maintenance & Renewals Congress*, London, United Kingdom.

CRUD, Managing the Data-base Environment, p. 381, at Google Books.

De Barr, A. E. and Oliver, D. A. (1975) Electrochemical Machining, Surrey: Unwin Brothers Ltd., 1975.

Ferreira L. L. et al. (2017) 'A Pilot for Proactive Maintenance in Industry 4.0,' in *13th IEEE International Workshop on Factory Communication Systems (WFCS 2017)*, Trondheim, Norway.

Fronza, I. et al. (2013) 'Failure prediction based on log files using random indexing and support vector machines,' *Journal of Systems and Software*, 86(1), pp. 2–11.

Gorostegui, U. (2017) *Automatic Integration of Cyber Physical Condition Monitoring Systems as Part of Computerized Maintenance Management Systems*, Mondragon University.

Han, J., Cai, Y. and Cercone, N. (1992) 'Knowledge discovery in databases: An attribute-oriented approach,' in *18th Int. Conf. Very Large Data Bases*, Vancuver, Canada.

Hegedus, C., Varga, P. and Moldován, I. (2018) 'The MANTIS architecture for proactive maintenance,' in *5th International Conference on Control, Decision and Information Technologies CODIT*, Thessaloniki.

Hegedus, C., Ciancarini, P., Frankó, A., Kancilija, A., Moldován, I., Papa, G., Poklukar, Š., Riccardi, M., Sillitti, A. and Varga, P. (2018) 'Proactive Maintenance of Railway Switches,' in *5th International Conference on Control, Decision and Information Technologies CODIT*, Thessaloniki, Greece.

https://kafka.apache.org/

http://projects.spring.io/spring-boot

IEEE Standard. (1998) 'IEEE 830-1998 – Recommended Practice for Software Requirements Specifications,'.

'Internet of Things Architecture,' [Online]. Available: http://www.iot-a.eu.

ISO, 'ISO 13374-1: Condition monitoring and diagnostics of machines – Part 1: General Guidelines.'

Jantunen, E. et al. (2016) 'Optimising maintenance: What are the expectations for cyber physical systems,' in *Emerging Ideas and Trends in Engineering of Cyber-Physical Systems (EITEC)*, Vienna, Austria.

Junnola, J. (2017) *The Suitability of Low-Cost Measurement Systems For Rolling Element Bearing Vibration Monitoring*, University of Oulu.

Korošec, P., Bole, U., and Papa, G. (2013) 'A multi-objective approach to the application of real-world production scheduling,' *Expert systems with applications*, 40(15), pp. 5839–5853.

Larrinaga, F. et al. (2018) 'Implementation of a reference architecture for cyber physical systems to support condition based maintenance,' in *5th International Conference on Control, Decision and Information Technologies CODIT*, Thessaloniki.

Lenarduzzi, V. et al. (2017) 'Analyzing forty years of software maintenance models,' in *39th International Conference on Software Engineering (ICSE)*, Buenos Aires, Argentina.

Levina, E. and Bickel, P. (2001) 'The EarthMover's Distance is the Mallows Distance: Some insights from statistics,' in *Proceedings of ICCV*, Vancouver, Canada.

The MANTIS consortium, 'The MANTIS project website,' [Online]. Available: http://www.mantis-project.eu/.

Mantis Consortium. (2016) 'D2.2 Reference architecture and design specification,'.

McGeough, J. A. (1988) *Advanced Methods of Machining*, London & New York: Chapman and Hall Ltd., pp. 55–87.

McKay, M. (2018 to be published) *Bearing Failure Prediction Based on a Wear Model*, Aalto University.

MIMOSA. 'OSA-CBM 3.3.1 standard,' 29/6/2010. [Online]. Available: http://www.mimosa.org/mimosa-osa-cbm. [Accessed 22/10/2015].

MIMOSA. 'OSA-EAI 3.2.3a standard,' 12/5/2014. [Online]. Available: http://www.mimosa.org/mimosa-osa-eai. [Accessed 22/10/2015].

MIMOSA Consortium. 'An operations and maintenance information open system alliance,' [Online]. Available: http://www.mimosa.org/mimosa/.

MIMOSA Consortium. 'The MIMOSA project site, [Online]. Available: http://www.mimosa.org/.

Mosquitto^{TM}, (2010) 'An Open Source MQTT v3.1/v3.1.1 Broker,' [Online].

National Instruments Corporation, 'LabVIEW,' 2018. [Online]. Available: http://www.ni.com/en-gb/shop/select/labview/. [Accessed 01/04/2018].

PSK Standards Association. (2007) *Vibration Measurement in Condition Monitoring. Choice and Identification of Measurement Point*, 3nd edition.

PSK Standards Association. (2006) *Vibration Measurement in Condition Monitoring. Selection and Mounting of Sensor, Connector and Cable*, 4th edition.

'RabbitMQ,' [Online]. Available: http://www.rabbitmq.com. [Accessed April 2018].

Stuart, A., Ord, K., and Arnold, S. (1999) *Kendall's Advanced Theory of Statistics: Volume 2A—Classical Inference & the Linear Mode*, Wiley.

Vinoski, S. (2006) 'Advanced message queuing protocol,' *IEEE Internet Computing*, 10, pp. 87–89.

Wold, H. (1975) Path Models with Latent Variables: The NIPALS Approach, vol. Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling, A. A. F. M. B. &. V. C. H. M. Blalock, Ed., New York: New York Academic, pp. 307–357.

'WSO,' [Online]. Available: https://wso2.com/. [Accessed April 2018].