Advanced Methods of Analyzing Operational Data to Provide Valuable Feedback to Operators and Resource Scheduling

(HQ-KPI, BigData /Anomaly Detection, Predictive Maintenance)

Dennis Braun, Urs Steinmetz

SES Energy Services GmbH (SES)

Essen, Germany

1. Introduction

Whilst the already high stress from markets and political influences on utilities remains on a high level, the demands for intelligent support for operators and strategy are increasing.
The paper will cover the experiences STEAG Energy Services (SES) has made in recent internal and external projects, discussing the practical consequences of "worlds colliding" when powerful IT approaches meet high expectations from the power plant managers, operators, and trading departments.
SES is by its business model exactly at this interface of IT and real power plant world since STEAG - the parent company - owns and operates a variety of power plants worldwide (conventional + renewable, 7 GW+ owned and ~ 200 GW operating experience). In-house engineering experts develop and maintain modules to answer issues in power production that cover the whole scope.
For early and reliable warnings to tackle developing problems, advanced approaches for Key Performance Indicators (KPI) with high information content are available and well established. Recently, STEAGs projects, as well as customers/partners from the wind power industry, refineries, cement plants and coal/gas-fired power plants have been increasingly demanding a fast and extensive approach that helps to efficiently prioritize and focus on the developing problems.
This paper will discuss the differences between two alternative approaches:
1. "traditional" (supervised learning) approaches to monitor plant operation and distill early warnings for predictive maintenance from the operational data
2. "big data/anomaly detection" approaches, based on unsupervised machine learning
Experiences and conclusions from recent projects with both approaches will be shared from the perspective of an owner & operator of power plants.

2. Different plants and requirements

As described in the introduction, this paper will focus on the perspective and experience of owners and operators of power plants of all types.
A preliminary remark: The expectations to such an IT system differ tremendously over the industries. From our experience, the willingness to pay for an IT solution to support operations as well as the correlated expected level of quality of such a solution somewhat scales with the size of the plant. One could argue that this is a commonplace to state, but the

result is that often that the smaller plants content themselves with a cheap and simple solution whilst forgetting to factor in the possible savings and findings that would be possible with a more advanced solution.

This being said, the situation of STEAG Energy Services GmbH turns the tables on that issue: Since the experience comes from the large scale plants, these tools that have proven themselves in practice to provide actual benefits are demanded also for the projects of the recent past, for example the development and operation of wind parks etc.

For a reliable operation of the plants, it is fundamentally significant to have reliable and proper knowledge about the current condition of the plant to be able to plan maintenance as well as scheduling the resources accordingly.

By transferring for example the High Quality KPI, which are based on pre-engineered KPI with neural networks and lots of specific engineering know-how to the new projects outside of conventional power plants, valuable findings were made. One example will be discussed in chapter 4.

3. Definitions and System Structure

As described above, this paper discusses two different approaches to predictive maintenance: On the one hand, pre-engineered Key Performance Indicators (KPI) are used. They can be implemented by using supervised machine learning algorithms, set up by experts. These KPI combine high sensitivity with a very low rate of false alarms.

On the other hand, unstructured and unlabeled data can be analyzed fully automatically using unsupervised machine learning algorithms.

Since this paper will share practical experiences from recent projects, it is helpful to understand where these generic approaches fit into the IT-systems from SES that are used worldwide:

High-Quality KPI / supervised learning algorithms are traditionally part of the **SR::SPC** system. Developed in the large scale conventional coal fired power plant environment, this system has reached a high level of sophistication in terms of ease-of-use and precision / efficiency. SR::SPC has been adapted to other industries since, so good experiences in the wind and cement industry etc. are available now.

Unsupervised machine learning algorithms, or the “big data approach” are implemented in the same software environment in the module **SR::SPC ML**.

Both tools (SR::SPC and SR::SPC ML) have the capability to detect anomalies.

The following table shows the advantages and disadvantages as experienced in our application scenarios:

	Advantage	Limitation
“Expert HQ KPI“	• Systematic configuration per KPI and individual training period allow for high sensitivity • Easy interpretation of the results • Also good detection of cyclical changes incl. projection	• Initial engineering effort higher than for Big Data approach • Faults for which no KPI has been defined are not detected.
“Big Data“	• Lower efforts when creating the system • Anomalies no longer go unnoticed because no KPI has been defined.	• Attribution of detected anomalies to possible errors/faults requires analysis by engineers. • Selection of the learning period influences the result; more data / longer periods are required.

Figure 1: Advantages and disadvantages of HQ-KPI and Big Data approaches

In connection with Figure 1 please make sure to also read the summary of this paper for an important update.

4. HQ-KPI and Big-Data approaches in a practical example

When applying the two approaches to a practical example, the pros and cons become clear:

At the end of the day, Predictive Maintenance means that when maintenance work is done on time and at the right time, the maintenance cost can be reduced and the availability of the unit can be improved.
Figure 2 gives an overview:

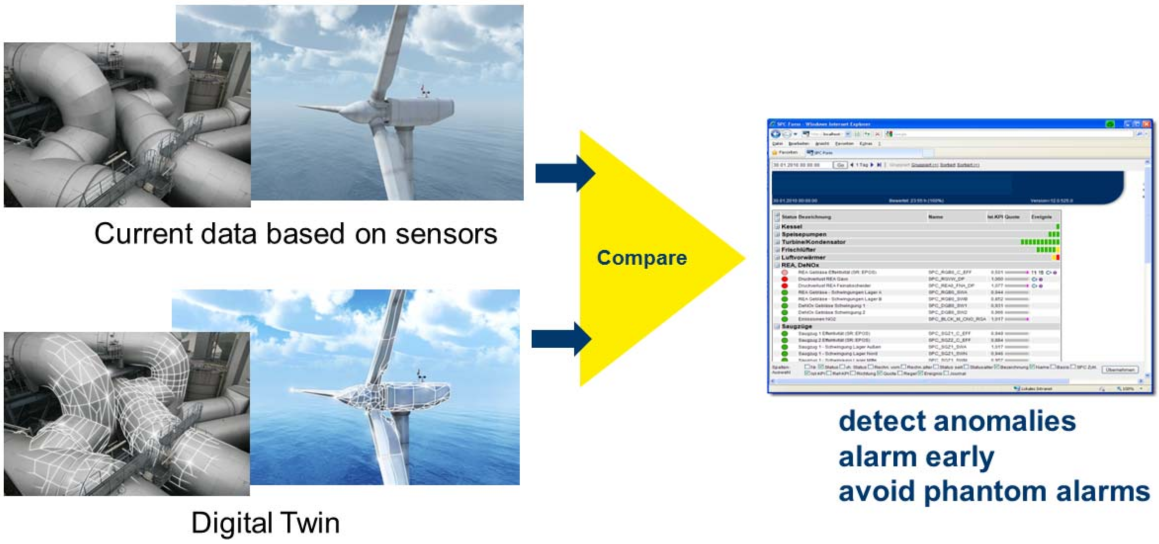

Figure 2: Comparison of current sensor data with a Digital Twin

As Figure 2 shows, having a reliable Digital Twin (a model of the real facility that provides the necessary data about the behavior of the real facility with adequate precision) can help to spot irregularities and issue alarms early on. However, it is crucial to make sure not to flood the users with too many and phantom alarms (false alarms).

In some cases, a Digital Twin is used to spot not only anomalies in terms of faults and defects but in terms of performance losses. In this case, the Digital Twin can alternatively to a statistical model be based on a thermodynamical model of the plant, mainly the water-steam cycle and the air-fluegas path. The SR-systems as introduced in chapter 3 can be extended to SR::EPOS, which is a performance monitoring system based on an EBSILON®*Professional* model of the unit. Since a model based on physical equations can be used for further analyses (such as evaluating the monetary impact of process deviations) or be accurate even in modes of operation that never occurred before, for some customers' problems a solution as in Figure 3 can be the right solution.

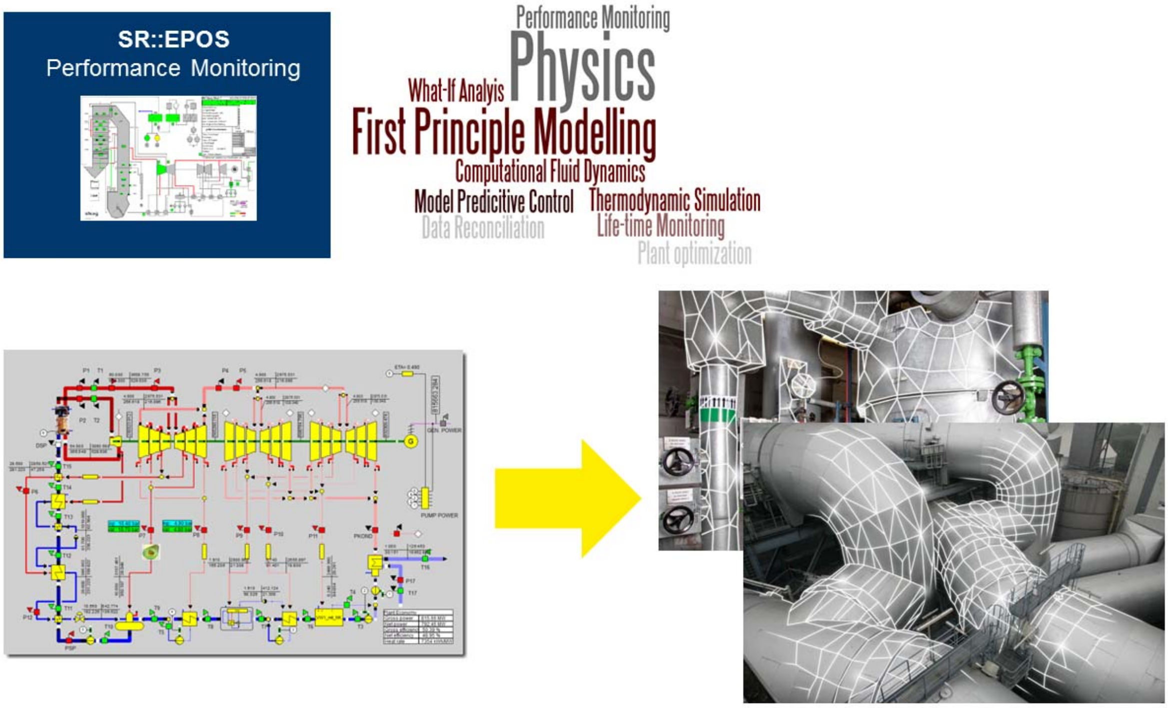

Figure 3: Physical model for a Digital Twin

For problems that are not based on thermodynamics, such as vibrations for example, the Digital Twin can be based on statistical models.
As shown in Figure 4, the reference value can be generated either via a High-Quality-KPI with a supervised learning approach or with Big Data methods based on unsupervised learning.

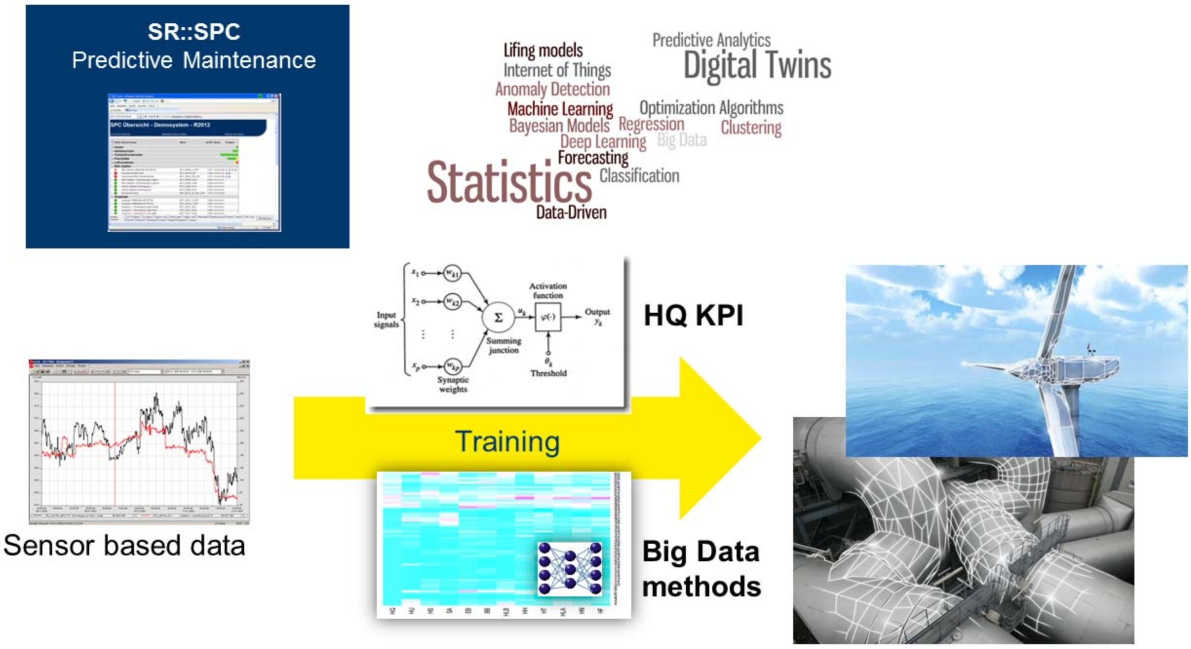

Figure 4: Digital Twins based on High-Quality-KPI or Big Data methods

Going into more detail, the following Figure 5 shows the temperature main bearing of a wind turbine (y-axis) over time (x-axis). **The time scale for the following graphs (x) is 01.11.2010 to 01.09.2013, almost three years.**

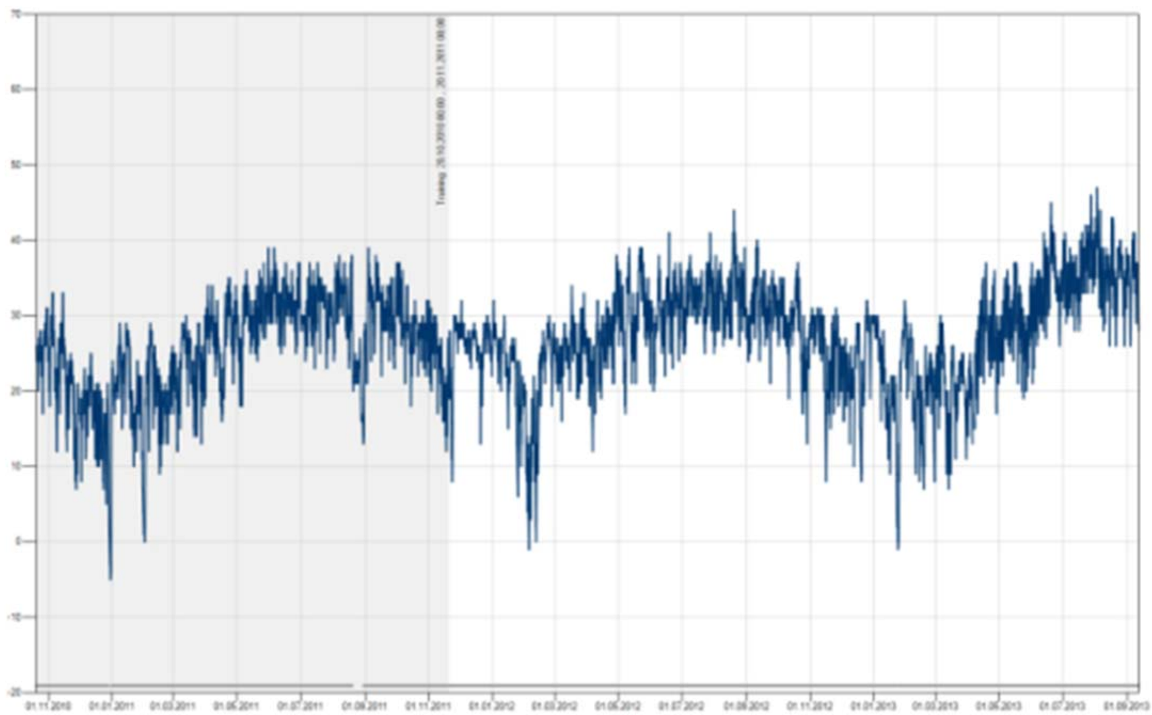

Figure 5: Temperature main bearing of a wind turbine (y) over time (x)

The annual fluctuation can be well observed in Figure 5, as expected, the temperature is lower in winter than in summer operation. The area marked in grey on the left hand side is the time period that has been used for the training of the model. With the bare eye, no anomalies are easily to be detected so far.

Using more engineering knowledge about the behavior of wind turbines, the following Figure 6 shows three dimensions:

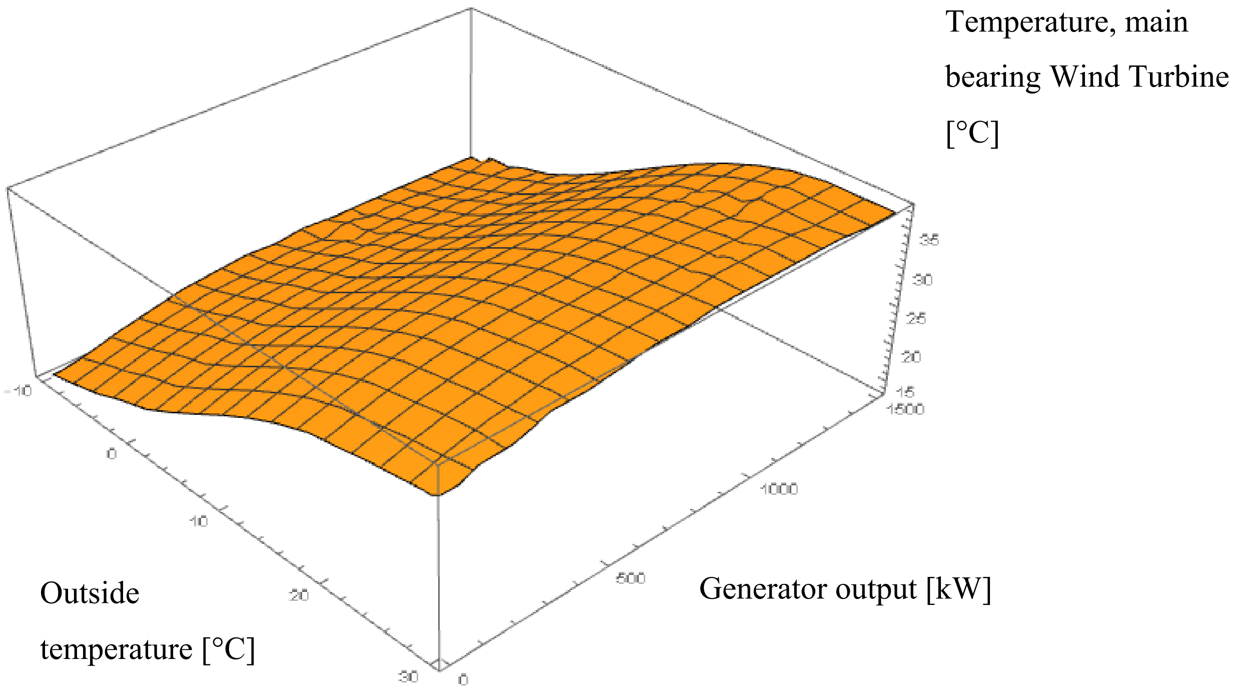

Figure 6: Relevant data to assess wind turbine performance

For a person with technical experience, the view above makes sense: Low outside temperatures lead to lower bearing temperatures, even if the generator output is constant (e.g. generator output 1.300 kW, temperatures approximately -10 to +30 °C and bearing temperatures approximately 15 to 40 °C).

This model as shown in Figure 6 represents the digital twin for this wind turbine, specifically for these three values. For other values there might be additional models.

Therefore, this diagram can be used to generate a reference value that can point out in comparison with the measured values if any measured value is OK or an anomaly. This is shown in Figure 7 below:

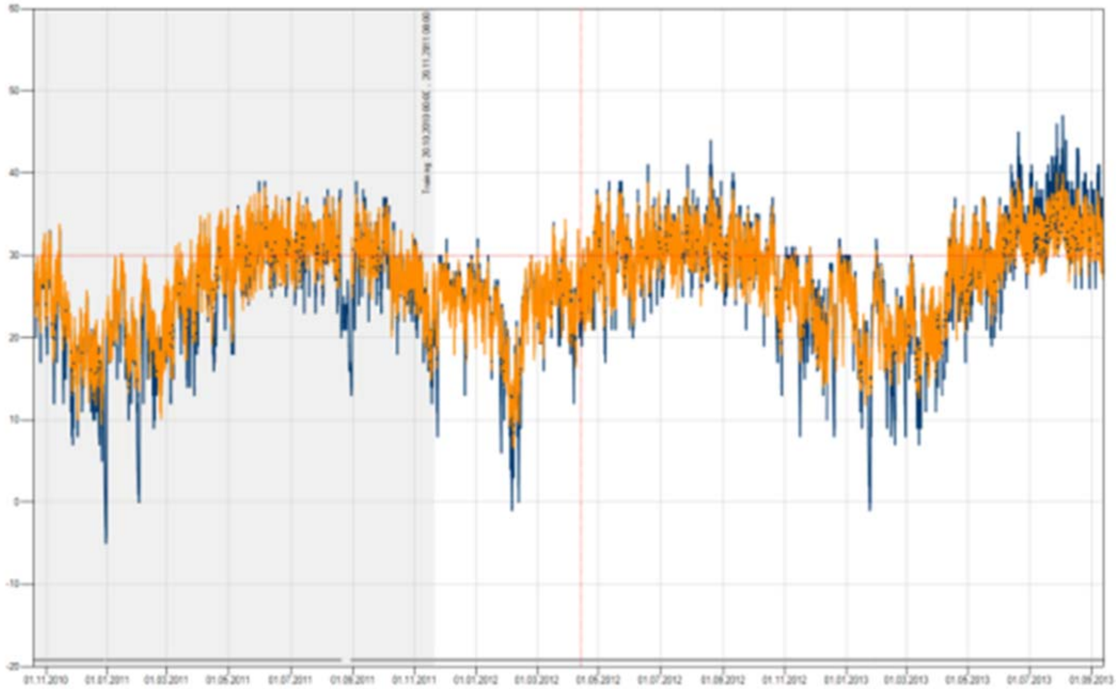

Figure 7: Temperature main bearing (blue) and reference value (orange) of a wind turbine (y) over time (x)

The person with technical experience can now analyze the graph easily. More or less, the two values correspond, most of the time the measured temperature (blue) is nearly equal to the reference value (orange.) On the other hand, on the right hand side the blue line exceeds the orange one. That could mean nothing or it could mean something is going on.
A more precise assessment requires more information. It is important to point out that changes of this relatively small magnitude often go unnoticed.
It is even more important to be aware that these are only two values the reader of this paper has to keep an eye on in this example in order to detect the deviation. In the real world, per wind turbine there are dozens to hundreds of values to analyze and there might be many wind turbines within the company or wind park to evaluate. Why would someone have exactly these two values on the screen that are shown in Figure 7? And would the reference value be readily available? In many cases that won't be the case. Of course, if for any reason the owner or operator has no access to the data on the wind turbines bus system, for example due to contractual or technical reasons, monitoring and predictive maintenance is notably harder to implement.
For deeper evaluation, the data from above is applied to a HQ-KPI made with SR::SPC. Amongst other features, SR::SPC can run different statistical tests on the KPI. The result of

one of these tests - the so called CUSUM-chart - is shown in figure 8.

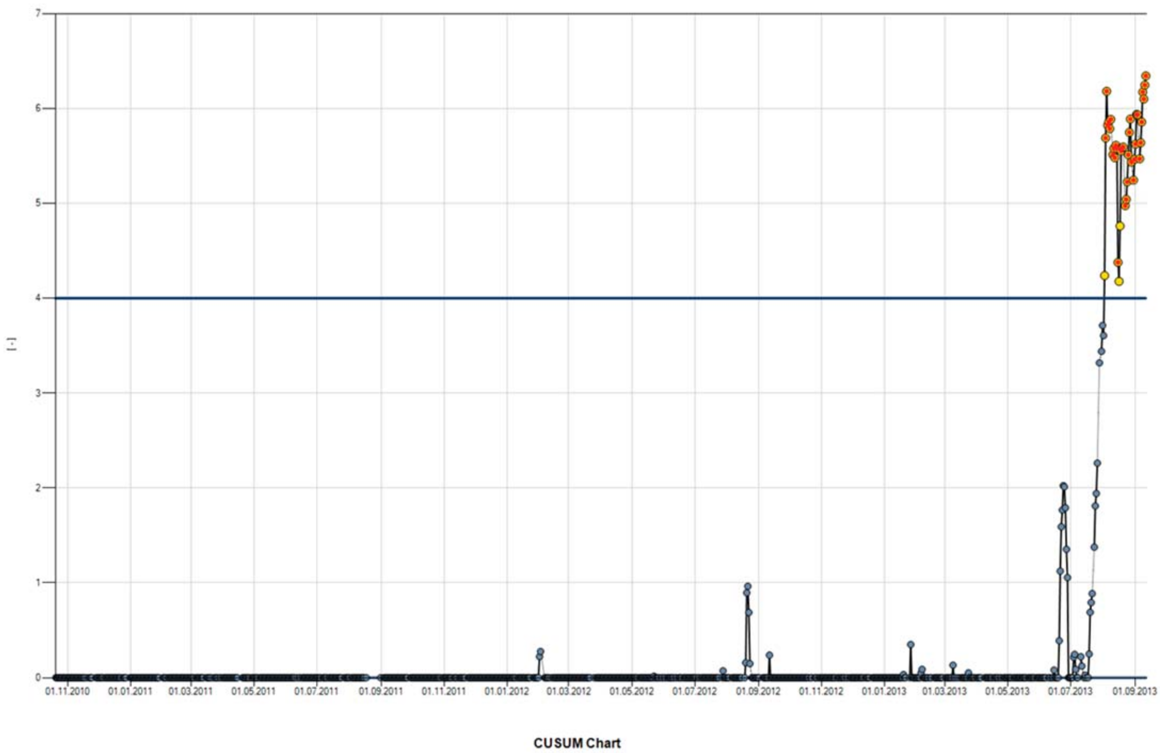

Figure 8: CUSUM (cumulative sum) chart of the KPI

A system such as SR::SPC will process the measured data e.g. by filtering the raw data as well as managing the reference model. A KPI based on this is normalized, meaning it doesn't have dependencies otherwise than from condition or process quality. In CUSUM-charts the KPI shown is typically close to zero and below a predefined limit or above a critical limit so anomalies can be easily spotted. This it not only possible when the system user is actively looking at these charts, SR::SPC can also send emails automatically with a summary to the responsible users in case there are any anomalies detected.

Additional to the steps discussed so far in this paper (raw data, Digital Twin based on first principle model or Digital Twin based on supervised learning), Figure 9 shows how an unsupervised learning can be established: First, the sensor data of a reference period is imported. Second, via preprocessing, and unsupervised machine learning a model of the Digital Twin can be generated. Third, current sensor data can once again be used to be compared to the Digital Twin and by automated error analysis a comprehensive overview can be automatically prepared.

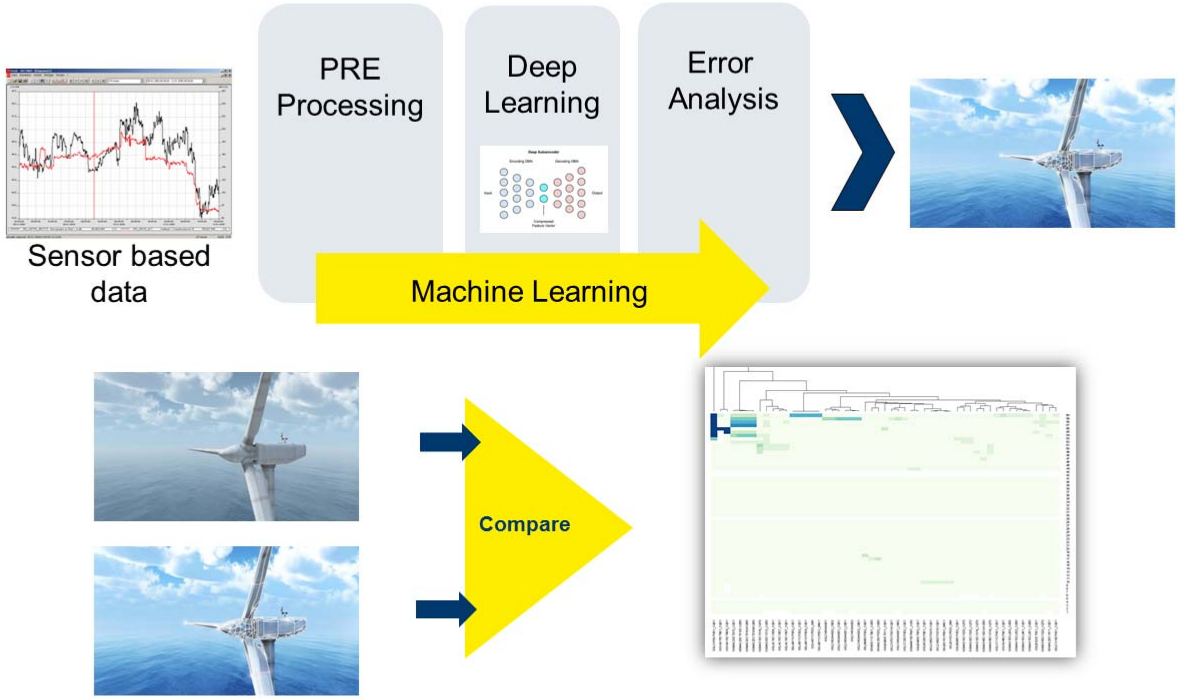

Figure 9: From Sensor Data to digital twin and

In the following Figure 10, the overview will be discussed on more detail:

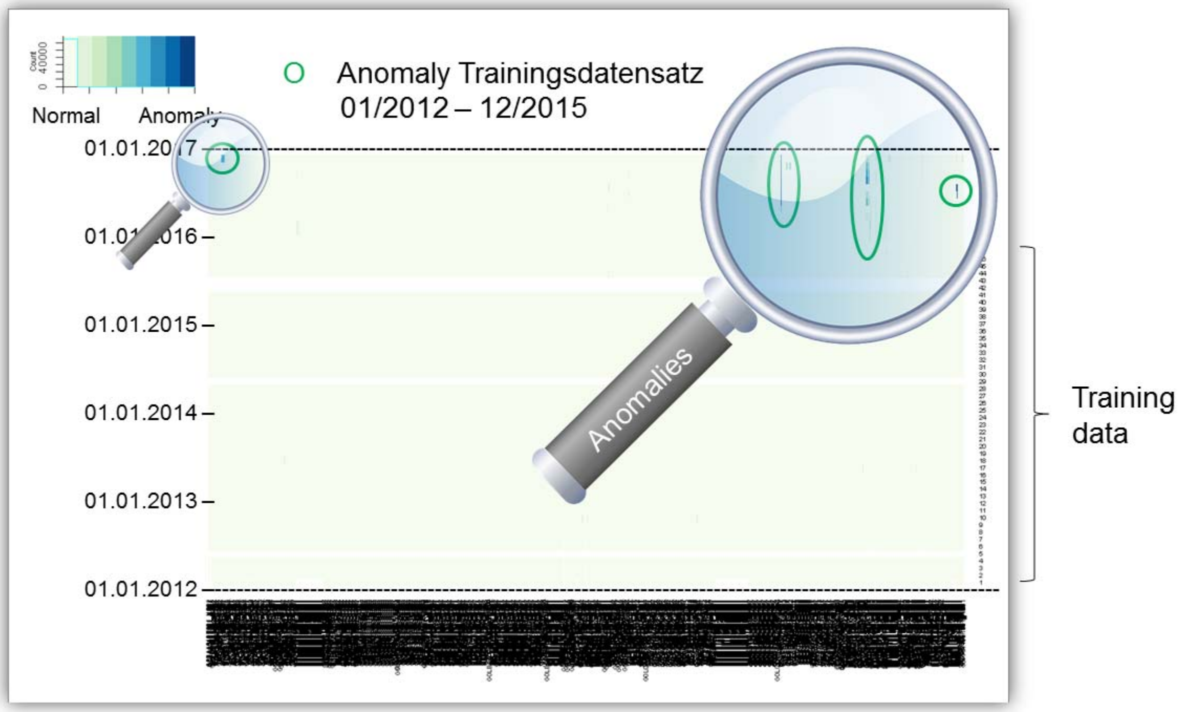

Figure 10: Hundreds of measurements in one view, five years data for a set of data for anomaly training from 01/2012 to 12/2015

As highlighted in Figure 10, the unsupervised learning was started with input data of four years.

The last year (2016) of this example shows a total of four groups of anomalies (highlighted by the magnifying glasses).

In terms of precision of the anomaly detection (avoiding false alarms), that is a very good result: Considering the hundreds of measurements and data of full four years, **only** few areas are significant. That means, based on the data that was used to train the system, the unknown data of the last year is evaluated reliably.

By selecting only the channels with anomalies in the original report, individual KKS groups become recognizable.

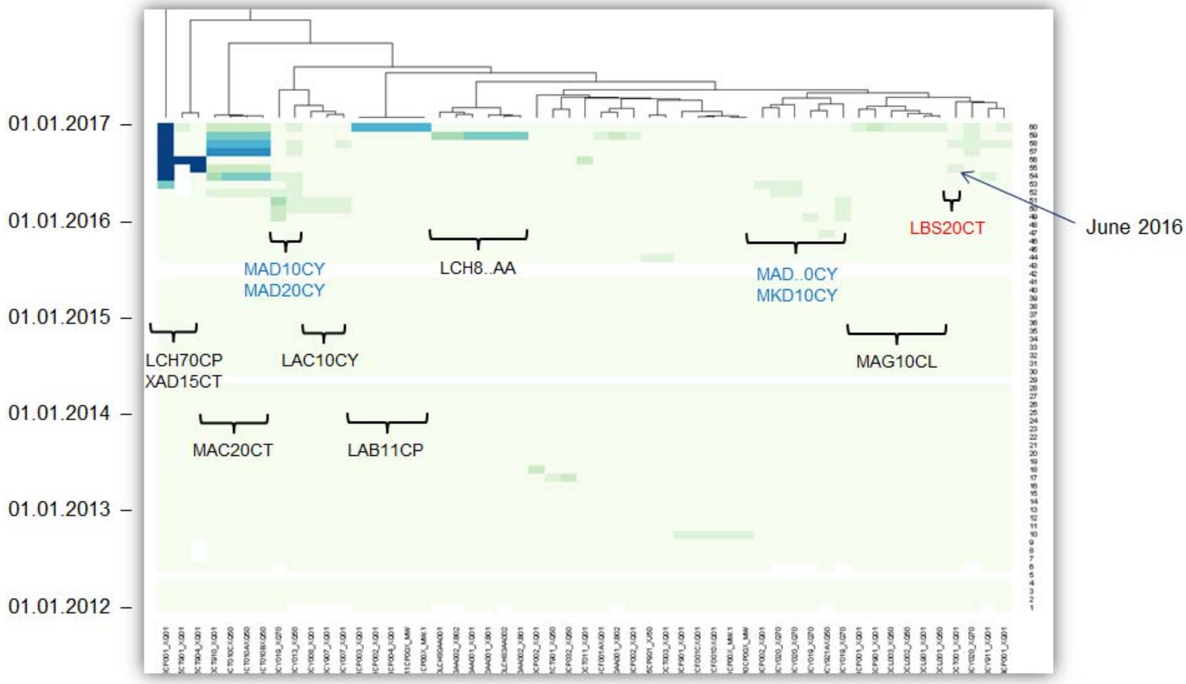

Figure 11: Heatmap for Turbine, training data 01/2012 – 12/2015 (incl.), anomalies above a certain limited commented with their respective KKS groups

Based on this information, the responsible engineer can then start the analysis for the root cause of the anomalies. Additional information, for example from data historians, the DCS archive or process monitoring systems can be used to narrow down the actual issue. On-site inspections and maintenance jobs can therefore be planned efficiently.

5. Summary

As shown in this paper, modern IT has the potential to significantly support operators of plants to meet their goals.

Also, advanced methods of analyzing operational data have successfully left the realms of academia and in real-world projects they were able to provide valuable feedback to operators and resource scheduling.

The setup and maintenance of the system are in good hands of an experienced engineer who can become an expert for these applications within a few days of advanced training.

Finally, operators should be considered to be “customers” of the system: evaluating the results and procure user-level steps for analysis are sufficient for them unless otherwise requested. A good exchange and cooperation with the system expert seem to be the road to a successful and beneficial long term usage of such a system. Also from our experiences, having a black-box system which does not allow the users and engineers of the customer to analyze and verify how certain results are obtained makes it hard to gain the trust in the system that is needed. White-box systems seem to be better accepted.

Finally, the paper showed at different applications that different approaches can lead to various but equally helpful results in terms of alarming for predictive maintenance and anomaly detection.

However, the effort to set these systems up (HQ-KPI, BigData) and to evaluate the alarms differs tremendously.

Based on the recent project experiences, the following intermediate step – a smart combination of the advantages of both approaches **and** empowering the engineer who sets-up the system whilst using modern IT-tools to take care of the relevant specifics of the respective seem to be the best compromise. The following Figure 12 expands the version from Figure 2 and brings together two worlds in the IT branch as well as it shows an approach that has proven to provide valuable feedback in real world applications:

	Advantage	Limitation
“Expert HQ KPI“	• Systematic configuration per KPI and individual training period allow for high sensitivity • Easy interpretation of the results • Also good detection of cyclical changes incl. projection	• Initial engineering effort higher than for Big Data approach • Faults for which no KPI has been defined are not detected.
“Smart Data“		
“Big Data“	• Lower efforts when creating the system • Anomalies no longer go unnoticed because no KPI has been defined.	• Attribution of detected anomalies to possible errors/faults requires analysis by engineers. • Selection of the learning period influences the result; more data / longer periods are required.

Figure 12: Smart Data combines the strengths of HQ-KPI approaches with the benefits of a big-data approach

In our understanding “Smart Data” means using existing engineering knowledge to set up individual models with optimized selection of the input data. That could mean a “Big Data” model for a specific part of the unit is set up instead of pouring all available data into the model and making the significance of alarms as well as the evaluation of them more difficult than necessary.

For the large scale conventional power plants huge effort was appropriate to develop highly sophisticated software tools to support engineers. The engineers were able to accumulate a lot of know-how, on the one hand about the plants but also on the other hand about the software tools and how to use them efficiently. These tools and the know-how is now available for the benefit of other industries which otherwise would not pursue such a high-end approach, for example due to the possible savings and losses to avoid. However, the potential for applying the high-end solution to these new industries is significant.

These industries besides large scale conventional plants, which are often much smaller in terms of installed capacity - for example smaller wind parks or other renewable energy projects (PV, solar thermal, geo thermal etc.) – can profit from the developments in the conventional plants and also use modern methods and systems which might not be available if they were only meant to be developed for these smaller units and wallets.

That allows these owners, operators and engineers to use those state of the art systems to also provide valuable feedback and make the best use out of it.

Finally, based on this in-depth knowledge about the true condition and the possibility to organize maintenance effort accordingly the plan resource scheduling can be more precise and avoid unpleasant surprises such as “seemingly” sudden issues that can lead up to an emergency shutdown and extended periods for repair.

Dennis Braun, M.Sc.
System Technologies / Sales
Telefon +49 201 801-4113
Mobil +49 160 90747391
dennis.braun@steag.com

STEAG Energy Services GmbH
Rüttenscheider Str. 1-3
45128 Essen
Germany
www.steag-energyservices.com