Statistical and Semantic Multisensor Data Evaluation for Fluid Condition Monitoring in Wind Turbines

Ayetullah Günel[1], Ankush Meshram[2], Torsten Bley[1], Matthias Klusch[2], Andreas Schütze[1,3]
[1] *Zentrum für Mechatronik und Automatisierungstechnik GmbH (ZeMA), Saarbruecken, Germany*
schuetze@mechatronikzentrum.de
[2] *Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbruecken, Germany*
[3] *Saarland University, Lab for Measurement Technology, Saarbruecken, Germany*

Abstract:

Condition monitoring of complex systems is usually based on various sensors to gain information on the system status and identify potential problems at an early stage. However, evaluation and interpretation of multiple sensor data is often a major problem due to complex interdependencies between measured sensor data and actual system condition. We have addressed this problem for the example of a fluid condition monitoring system in wind turbines, i.e. an oil filter and cooling subsystem equipped with various sensors to monitor relevant temperatures, differential and absolute pressure drop across the filter as well as specific oil conditions such as dielectric constant and viscosity to monitor chemical oil degradation. Data were recorded from two wind turbines in a field test over more than two years. Data evaluation was performed with both statistical as well as semantic analysis methods with the goal of predicting the remaining filter lifetime and identifying unusual data patterns indicating potential problems for the overall system. Results indicate a large potential for the hybrid combination of statistical and semantic sensor data analysis for condition monitoring and have especially indicated the potential for learning data patterns with one specific system and transferring the results to other systems. This would allow using operational experience in larger networks such as offshore wind parks to improve the performance and reliability of intelligent condition monitoring systems.

Key words: wind turbines, fluid condition monitoring, multisensor data evaluation, statistical data evaluation, semantic data evaluation.

Introduction and Motivation

Wind turbines play an important role for converting our energy system to sustainable energy sources, especially in offshore wind parks. However, offshore systems pose large challenges for their efficient operation under unfavorable weather conditions, essentially making repairs and maintenance impossible during winter months for wind parks in e.g. the North Sea. But also onshore systems would greatly benefit, i.e. achieve a higher economic gain, if efficient scheduling of maintenance operations as well as early diagnosis of wear and prediction of critical system conditions are state of the art. Condition monitoring systems allowing early detection of faults and wear as well as condition based maintenance are thus an important aspect for achieving reliable and cost efficient operation. Within the complex wind turbine system, the main gear transferring the rotation from the main rotor to the generator is one critical component. To ensure reliable operation, an oil conditioning system is integrated which filters the oil continuously and additionally cools the oil at higher operating temperature to certify consistent lubrication and prevent fast degradation of the oil.

Condition monitoring of the main gear system is today addressed mainly by systems monitoring vibrations to recognize wear in gears and bearings using complex signal processing and interpretation [1, 2]. Other methods address low cost condition monitoring by trying to monitor the output power and rotational speed, i.e. data available without additional sensors [3]. This approach reflects not only the reduced cost of systems without additional sensors, but also acknowledges the fact that increasing the complexity of the system can actually lead to more maintenance and increased downtime, e.g. due to a defective sensor. Wind park operators are therefore reluctant to install additional sensors without a clearly proven benefit for the operation. This benefit also includes a clear indication of the detected fault or predicted system condition, thus requiring a comprehensive interpretation of the complex recorded sensor raw data.

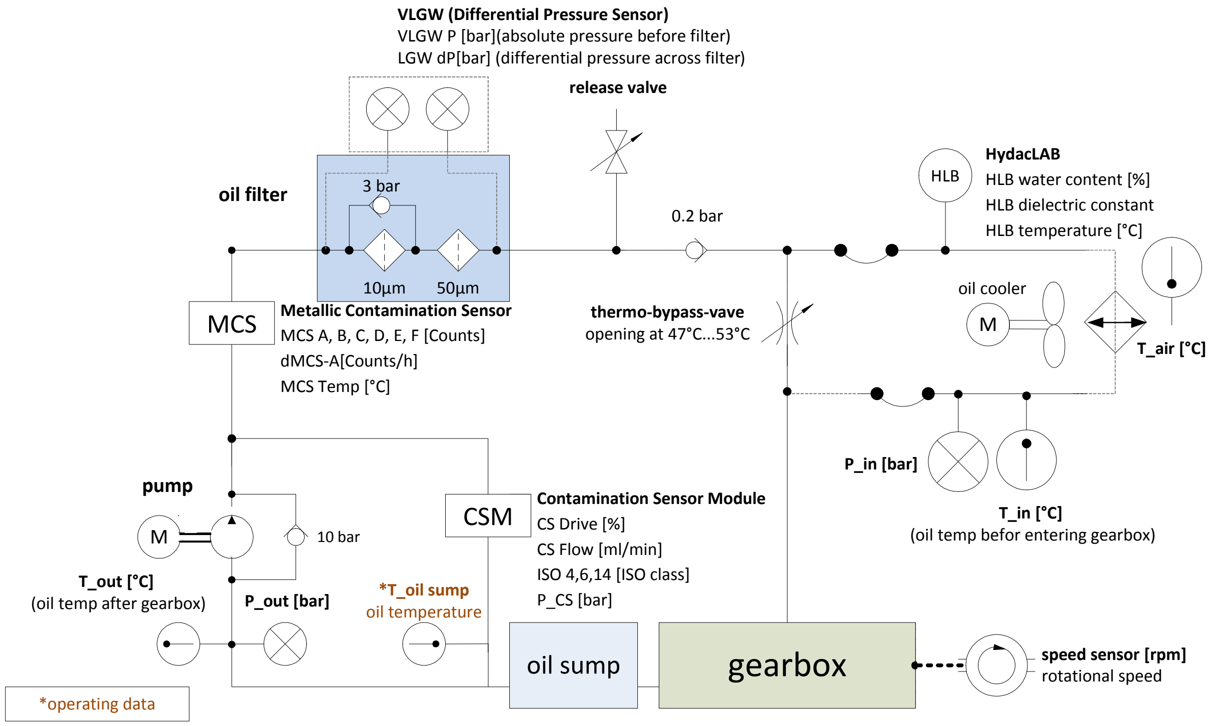

Fig. 1. Schematic of the fluid condition monitoring system showing the oil filter and cooling loop with multiple sensors for monitoring physical parameters (temperatures at different points, absolute pressure, differential pressure across the filter, gear speed) as well as various fluid parameters (metallic contamination, particle contamination, water content, dielectric constant).

We therefore propose a combination of statistical and semantic data evaluation to allow better interpretation of the data recorded from various sensors, but also to make use of (hidden) redundancy in the sensors to increase the reliability of the overall system. In a first study we concentrated on the oil conditioning system for the main gear. Data evaluation was studied based on data collected over several years of normal operation for two identical wind turbines.

Experimental setting

To study the potential for condition monitoring of the fluidic subsystem, a multi-sensor system combining various classical sensors (absolute and differential pressure at the oil filter, temperature monitored at different points in the oil loop, gear speed) with specific fluid condition sensors (dielectric constant, particle contamination sensors) as shown in Fig. 1, was installed in two identical wind turbine systems (GE 1,5 sl) [4]. Fig. 2 shows a 3D sketch of the oil conditioning system indicating the compact integration of the various sensors with the oil filter and cooler system.

In a first approach, we concentrated on predicting the condition of the oil filter and allowing a prediction of its remaining useful operating life. The main sensors used for this approach were the differential pressure sensor (VLGW in Fig. 1) as well as oil temperature recorded near the oil filter (MCS Temp in Fig. 1) to reflect the fact that oil viscosity and thus pressure drop across the filter depends greatly on the actual oil temperature. This was combined with the rotational speed of the generator and with the cumulated power generated by the wind turbine as an indicator of the operational lifetime of the system. Cumulated power was obtained from operational data recorded by the wind turbine operator. The raw data were reduced, i.e. filtered, to small but typical wind turbine speed (1000-1100 rpm at generator) and oil temperature (55-56°C) ranges, to allow comparison of similar operating conditions.

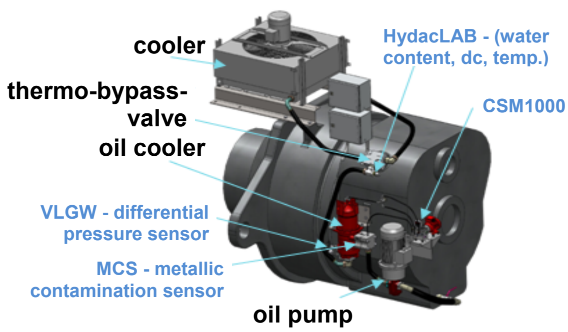

Fig. 2. 3D sketch of the standard oil conditioning subsystem with filter, cooler and oil pump equipped with various additional sensors indicated in blue.

Fig. 3 shows an example for the differential pressure recorded for wind turbine B over a period of almost 100 weeks starting in January 2000 filtered to typical operating conditions.

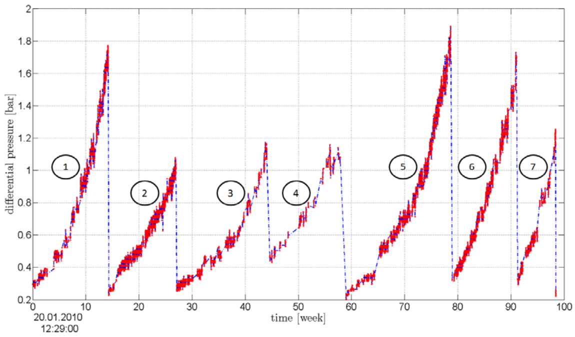

Fig. 3. Differential pressure recorded in wind turbine B over a period of nearly 100 weeks. Data were filtered to show only typical operating conditions with respect to wind turbine speed and oil temperature. Numbers indicate seven filter change cycles showing the strong increase of the differential pressure caused by increased loading of the filter. This differential pressure signal is today the main indicator for the filter change.

Statistical data evaluation

For the statistical data analysis, Linear Discriminant Analysis (LDA) [5] was used to classify the actual filter age based on the measured sensor data as well as additional information obtained from the wind turbine operating data, e.g. generated power. In addition, secondary features were calculated from the sensor data, e.g. differential pressure increase over time and statistical variation of the differential pressure. Based on the obtained high dimension data vectors, an LDA projection was calculated based on data from the first filter change cycle of wind turbine A classifying the filter condition in three stages: new, medium and advanced aging. This classification was based on the differential pressure only, i.e. $\Delta p < 0.6$ bar is classified as new, $0.6 < \Delta p < 1.0$ as medium and $\Delta p > 1.0$ as advanced aging.

Applying this LDA to all recorded data from wind turbine A results in a distinct trend for subsequent filter changes reflecting the aging of the overall system as shown in Fig. 4. In this figure, only the group centroids are shown for better clarity. Note that this aging of the system is already evident after each filter change, i.e. the projection of the data reflecting low differential pressures also shows this general trend.

With this LDA, data from a second wind turbine B were also projected to test the transferability of the calibration from one complex system to another. The results, examples shown in Fig. 5, proved that a correct prediction of the filter status of the second wind turbine was possible, also reflecting the aging of the overall system with an increasing shift along discriminant function (DF) 1 and, to a lesser degree, DF2 with every filter change. Note that the aging is

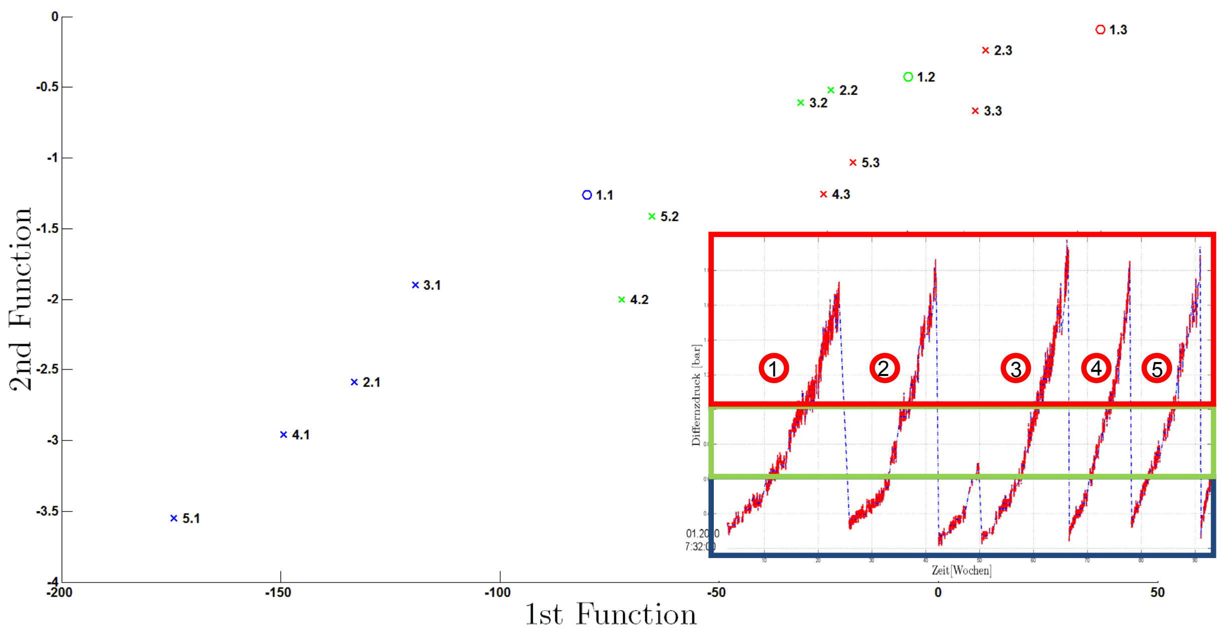

Fig. 4. Projection of all sensor data of wind turbine A recorded over five filter change cycles (insert) trained with data from the first filter change. Blue indicates new, green medium and red advanced aging. Here, only the centroid for each group is shown, the numbers indicate the filter change cycle and the group within the cycle, i.e. 1.1 reflects data from the first filter change classified as new, 5.3 data from the fifth filter change classified as advanced aging. These centroids show a shift towards the lower left corner of the LDA plot indicating the aging of the overall system. This aging can already be seen for the data immediately after each filter change as indicated by the five blue group centroids.

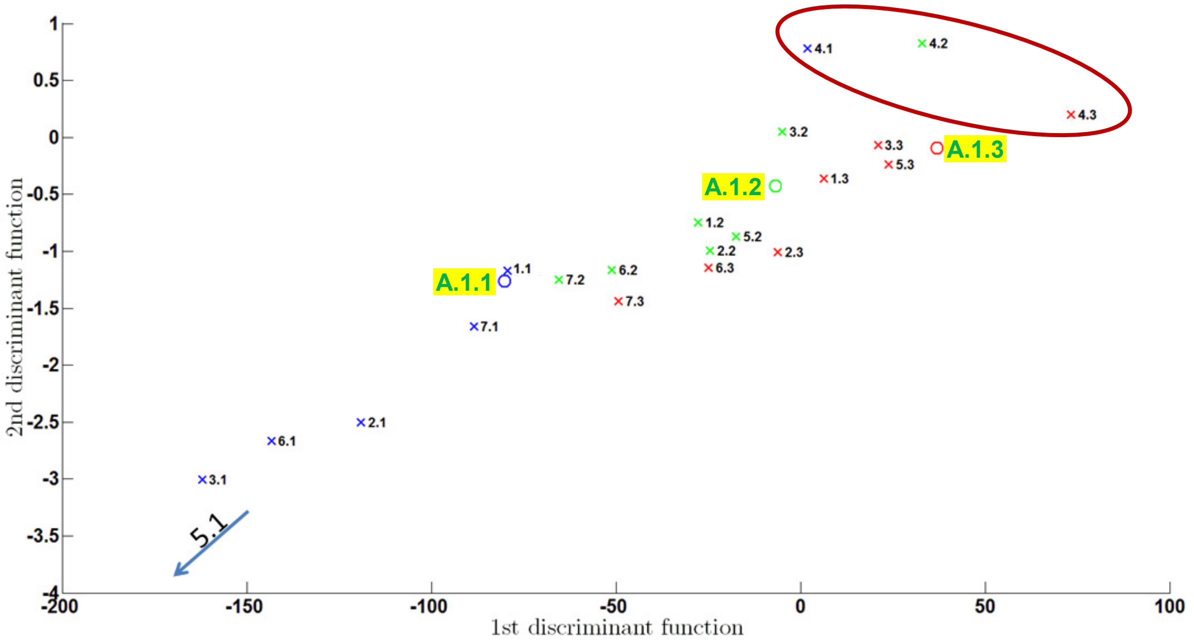

Fig. 5. Projection of all sensor data of wind turbine B recorded over seven filter change cycles (cf. Fig. 3) trained with data from the first filter change of wind turbine A (yellow markers). Blue indicates new, green medium and red advanced aging. Here, only the centroid for each group is shown, the numbers indicate the filter change cycle and the group within the cycle, i.e. 1.1 reflects data from the first filter change classified as new, 5.3 data from the fifth filter change classified as advanced aging. The projection shows that the general trend is comparable for both systems, i.e. increased filter age leads to a shift to the right and up, while the aging of the overall system, i.e. subsequent filter changes, shifts data to the left and down. The data for filter change 4 (red ellipse) and 5.1. show clear deviations from the general trends indicating an abnormal system behavior. For more details see text.

more clearly seen with increasing filter loading, but due to statistical variations, individual group centroids show deviations from this trend.

However, two deviations from the overall trend are immediately obvious in Fig. 5: For cycle 4, the centroids are shifted to the right and up against the general trend. In addition, over the filter cycle, the shift along DF2 is reversed compared to the other filter cycles, i.e. going down instead of up. In this filter cycle 4, a filter with a smaller mesh (5 µm instead of the usual 10 µm) was installed. This “unusual operating condition” is easily identified by the clearly different projection of the data obtained by the LDA. Note that the differential pressure data for this filter cycle are not notably different from the other cycles (cf. Fig. 3) except that Δp is significantly higher immediately after the filter change.

In addition, the data for filter cycle 5 show a very large shift towards the lower right immediately after the filter change (group centroid 5.1 with a DF1 value of -350). It is actually not clear what has caused this shift in the projection, but this might be connected to the sudden increase of differential pressure during this filter cycle approx. six weeks after the filter change (cf. Fig. 3). However, such deviations from normal behavior could be used as an indicator to check the system and to learn the specific causes. In this way, over the lifetime of the systems and with increased data base as well as experience in data interpretation, different unusual, potentially harmful or even hazardous conditions could be identified.

This statistical analysis alone, however, while allowing important insight for component and system manufacturers, is difficult to interpret for the end user and wind turbine operator without additional specific domain knowledge-based interpretation. For this purpose, the ICM-Wind system, as depicted in Fig. 6, provides a second part for semantic sensor data analysis allowing the operator to issue high-level queries on sensor concepts, machine components, their logical interrelations and conditions based on the encoded data. A hybrid analysis module of the system combines the statistical with semantic analysis which allows for an even more advanced sensor data interpretation none of both analysis modules is able to provide.

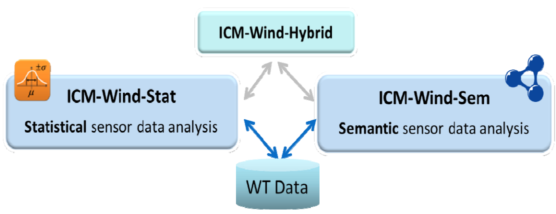

Fig. 6. ICM-Wind system architecture for individual and combined statistical and semantic analysis based on wind turbine sensor and operating data.

Semantic data evaluation

The semantic analysis of sensor data leverages semantic Web technologies and is based on (a) the semantic representation of the specific domain, (b) semantic data encoding, and (c) semantic data reasoning for intelligent query answering by the system. The domain knowledge of human experts about the application field is represented in a specific machine-interpretable, so-called formal domain ontology. In our case, a wind turbine domain (WTD) ontology was created as an extension of the *generic semantic sensor network ontology* published by the W3C expert group in 2011 [6]. This WTD ontology represents conceptual and factual knowledge about wind turbines (including the fluid condition monitoring subsystem) in the W3C standard ontology language OWL2 with formal logic-based semantics of defined concepts, relations and data types. Our WTD ontology is of reasonably small size with 91 concepts and 27 relations in total and encoded in machine-readable XML-RDF notation.

For example, the concept of a contamination sensor can be logically defined as a sensor which is attached to a gearbox and measures certain properties like specific ISO classes, flow and drive of contamination. Such conceptual knowledge is encoded in the concept base of the ontology while factual knowledge about physical components and measurements is represented in compliance with the defined concepts and relations in terms of RDF triples of the form (Subject Relation Object), where RDF is a W3C standard format for describing data semantics. For example, the RDF triple (wtd:CS1000 rdf:type wtd:ContaminationSensor) represents the factual knowledge that the sensor *"CS1000" is an instance of the class of contamination sensors* which is defined in the concept base of the WTD ontology. At the same time, the data values for properties measured by this sensor are RDF-encoded in the fact base as well. The complete WTD ontology (concept and fact base) is maintained in the RDF triple store SwiftOLIM (cf. Fig. 7). This knowledge base also includes the implicit factual knowledge of all logically deduced facts. For example, our 30K data records of 20 days of sensor measurements encoded in 518K RDF triples amount to 2 million triples in the knowledge base of the ontology. Semantic reasoning over such volumes is reasonably manageable.

Once the WT data has been semantically encoded and stored in the knowledge base, the user can make use of it by issuing different types of semantic analysis queries which are processed by the ICM-Wind system with different semantic data reasoners such as Pellet [9], STAR, and SPARQL-SPIN [10].

Conceptual queries are checking for logical subsumption relations between concepts with certain functionalities. One example is the simple Boolean query *"Is there a sensor type measuring the temperature and contamination of oil with particle size bigger than 400 µm?"*. In case of a positive („*true*") result, the system returns the respective sensor type (e.g. the concept of *MetallicContaminationSensor* in the ontology is logically equivalent with the query concept). The user can identify alternative types of sensors, i.e. sensors which provide similar or identical measurement functionalities, in case of a sensor failure, with conceptual queries like "*Which sensor is most specifically similar to the (new) oil temperature sensor?"* and the auto-

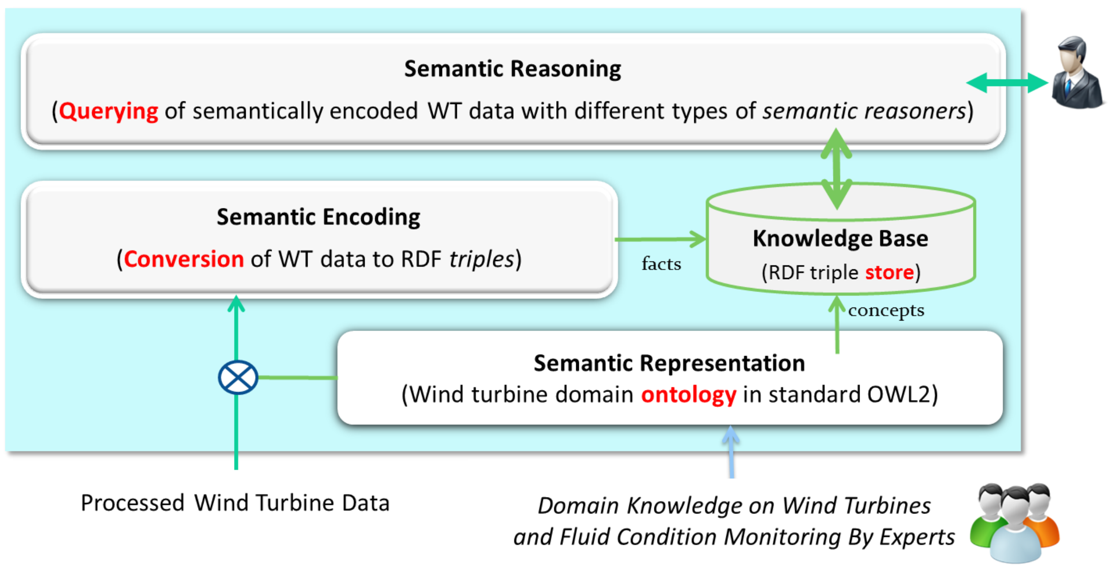

Fig. 7. Architecture of the semantic sensor data analysis part of the ICM-Wind system comprises of (i) the semantic representation of the domain by a wind turbine ontology together with (ii) the semantic encoding of sensor data in compliance with the ontology in a knowledge base, and (iii) the semantic reasoning on these data.

mated retrieval of actual sensors of this type in the fact base. This can make the overall sensory system more robust and reduce the maintenance effort for increased system complexity due to the integration of additional sensors. Combined with statistical data analysis this allows to additionally identify the degree of correlation between different sensors.

Data-centric queries are checking for objects with certain properties or the logical relation between a given set of objects. A simple example for the first type is the semantic query: "*Which sensors produced faulty measurements for which properties on 04.09.12?*" The system would check which of the individual sensors had readings outside of their specified range. For object relational queries like *"How are the sensors CSM_MarpI, SpeedSensor_MarpI related to each other?"* the returned (simplified) answer would be that both sensors are part of the same wind turbine *MarpI* together with a logical explanation based on the minimal property path between both objects in the knowledge base.

The system provides a web-based graphical user interface with format- and rule-based rewriting of informal user queries and result visualization which gives a simple indication of the monitored system condition and early warning of critical conditions or unusual sensor data requiring further analysis or maintenance.

Hybrid data evaluation

Hybrid analysis of sensor data by the ICM-Wind system is combining statistical and semantic data analysis where each would fail individually. We started to investigate the potential of such combination in both directions: Bottom-up where a semantic analysis result is required to perform statistical data analysis, and vice versa (top-down). For example, the query "*What is the condition of oil filter Filter_4?*" is answered first by the semantic analysis part ("top") which in turn exploits the statistical LDA result from the statistical analysis part ("down") to check the filter condition and finally return, for example, "Filter_4 is *faulty filter element*". In turn, if the LDA analysis ("bottom") of the oil temperature reveals that it was measured by some possibly faulty sensor (*S_Temp* property), the semantic analysis part answers the derived query "*Which property measured by which sensor can be used instead of S_Temp?"* followed by a re-run of the LDA analysis ("up") with the new data.

Conclusion and Outlook

We presented a first approach to statistical and semantic data analysis as well as their combination for fluid condition monitoring in wind turbines. Our experiments with implemented modules of the respective system ICM-Wind revealed that both new kinds of sensor data analysis are feasible and promising. Our future work on the ICM-Wind system is concerned with leveraging the principal component analysis [5] and neural networks [8] for improved, unsupervised statistical analysis of non-linear data as well as continuous semantic querying of sensor data streams produced by FCM sensor networks [6, 11], and scalable semantic reasoning on very big volumes of sensor data in the knowledge base. Further research addresses the hybrid data analysis and expanding the fluid condition monitoring to other subsystems in wind turbines. Finally, this approach can be applied to other fields both in renewable energies as well as general industrial applications.

Acknowledgement

Generous support for this project by HYDAC Filter Systems GmbH, Sulzbach, Germany, and ABO Wind AG, operator of Windpark Marpingen, Germany, is gratefully acknowledged.

References

[1] Z. Hameed, Y.S. Hong, Y.M. Cho, S.H. Ahn, C.K. Song, Condition monitoring and fault detection of wind turbines and related algorithms: A review, *Renewable and Sustainable Energy Reviews,* 13 (1), 1–39, 2009; doi: 10.1016/j.rser.2007.05.008

[2] B. Lu, Y. Li, X. Wu, Z. Yang, A review of recent advances in wind turbine condition monitoring and fault diagnosis, Proc. *IEEE PEMWA 2009 (Power Electronics and Machines in Wind Applications)*, 1-7, 2009.

[3] W. Yang, P.J. Tavner, C.J. Crabtree, M. Wilkinson, Cost-Effective Condition Monitoring for Wind Turbines, *IEEE Trans. Industrial Electronics* 57 (1), 263 – 271, 2010.

[4] Wind turbines installed in wind park Marpingen, operated by ABO Wind GmbH, http://www.abowind.com/com/operational-management/germany-marpingen-wind-farm.html

[5] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd ed., New York: Wiley, 2000.

[6] W3C Semantic Sensor Network Incubator Group, Semantic Sensor Network Ontology, (2011); available online: http://purl.oclc.org/NET/ssnx/ssn

[7] OWL2:http://www.w3.org/TR/owl2-overview/; RDF: http://www.w3.org/RDF/

[8] A. Zell, Simulation Neuronaler Netze, R. Oldenbourg Verlag, 1997

[9] Pellet: http://clarkparsia.com/pellet/

[10] SPARQL 1.1 query language for RDF: http://www.w3.org/TR/sparql11-query/; SPARQL-SPIN http://www.w3.org/Submission/2011/SUBM-spin-sparql-20110222/

[11] A. Sheth, C. Henson, Semantic Sensor Web. IEEE Internet Computing, 2008.