

Proficy Advanced Analytics: a Case Study for Real World PHM Application in Energy

Subrat Nanda¹, Xiaohui Hu²

¹GE Global Research, 1, Research Circle, K15A61, Niskayuna NY, 12309.

subrat.nanda@ge.com

²GE Intelligent Platforms, 325 Foxboro Blvd, Foxboro, MA, 02035

xiaohui.hu@ge.com

ABSTRACT

GE monitors a large number of heavy duty equipment for energy generation, locomotives and aviation. These monitoring and diagnostic centers located world-wide sense, derive, transmit, analyze and view terabytes of sensory and calculated data each year. This is used to arrive at critical decisions pertaining to equipment life management - like useful life estimation, inventory planning and finally assuring a minimum level of performance to GE customers. Although a large number of analytical tools exist in today's market, however there is a need to have a tool at disposal which can aid not just in the analytical algorithms and data processing but also a platform for fleet wide deployment, monitoring and online processing of equipment. We describe a Prognostics & Health Management (PHM) application for GE Energy which was implemented using GE Intelligent Platform products, and explore some capabilities of both the application and the analytics tool.

1. INTRODUCTION

GE monitors a very large number of heavy duty equipment for energy generation, locomotives and aviation. The main purpose of this monitoring analysis is to analyze the usage and condition of these equipment components, and to assist the users at Monitoring & Diagnostics (M&D) center to perform proactive maintenance activities, root causing existing problems and assist in planning for future downtime periods. This information can be used to help them to plan parts inventory and logistics, thus ensuring a higher reliability and availability for GE customers.

Various systems and sub-systems require the use of advanced data driven techniques to integrate the large amount of field data captured with the existing empirical and physics based models Jammu, Vinay(2010), et al.,

Nanda, Hu This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

These data driven methods are also used to assist the engineers to unearth hidden relations and patterns in key parameters of interest Vachtsevanos (2006), et al. We present in this paper a platform offering from GE Intelligent Platforms called Proficy Cause+ and Troubleshooter™ (GE Intelligent Platforms), (CSense Systems (Pty) Ltd) and discuss some of the key lessons learnt by applying it to monitor and develop prognostics & health management tools for GE equipment.

This paper is organized as follows: we provide a brief outline of the PHM schema developed and flow of information in Section 2. We then discuss the different sources of captured or derived information and patterns searched and the data pre-processing philosophy that are used by GE M&D in order to develop a robust and reliable PHM system. We also discuss in brief an anomaly detection algorithm to highlight one of the many anomaly detection methods used to monitor critical equipment and related alarm generation. We next introduce GE Intelligent Platforms Proficy platform in Section 3 and describe how it enables one to perform offline analysis, integrate with existing PHM platforms and perform field implementation. Section 4 contains some preliminary methods for anomaly detection that were used for a PHM implementation case study. Finally Section 5 has some pointers to what we are planning to do in near future and conclusions based on this study. The information used in this case study consists of data from GE Energy equipment that was anonymized and scaled to avoid disclosure of proprietary information. This does not in any way affect the validity of the methodology or implementation as described in the paper.

2. PROBLEM DEFINITION

In order to enable GE's Monitoring and Diagnostic Centre to detect insipient anomalies and failures in energy generation turbines and subsequently take corrective action, automated monitoring of the parameters of interest such as the following: operating temperatures, pressure ratios,

power produced, measured variables compared to their set point values, performance levels relative to empirically derived operating profiles, ambient conditions, etc. Hence it is required to automate the process of detecting anomalies in performance levels and pick up sudden shifts, and automatically provide possible root causes for observed anomalies. Further, if future performance levels can be predicted accurately, it would enable field engineers to suggest suitable control settings to power plant personal.

The PHM system developed has the following schema and flow of information (Figure 1)

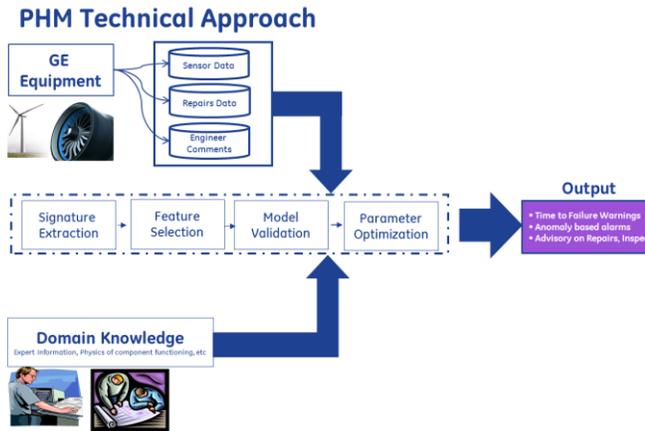


Figure 1: Schema of PHM system developed

In the following subsections, we outline some of the required components for a practical PHM system deployment. This includes capturing multiple data sources, pre-processing and anomaly detection methods. We describe some of these details in the following sections.

2.1 Major Sources of Information captured for PHM

The data captured for remote monitoring & control is of varied types and can be broadly classified into following types:

- Historical operational data: various sensor signals are stored in both central and distributed data bases. These contain equipment performance and component wise data at various time intervals. Different PHM applications may require customized data sampling rates depending upon the specific failure modes in the components for which PHM applications are developed.
- Controller Data: Onboard and centrally located controllers use and generate multiple logical values for accurate controlling process. These calculations are stored and are used for developing predictive or diagnostic methods.
- Engineer observations: Various free-form and structured textual information are recorded by GE field

engineers and M&D personal when responding to customer calls.

- Repair & Maintenance Data: Detailed descriptions relating to previous repairs and maintenance procedures and inspections performed on equipment are stored in various data repositories.

2.2 Types of PHM Signals & Patterns Monitored

A typical PHM system can be used to perform real time or offline processing to provide accurate equipment remaining life estimate thus enabling subsequent decisions to be taken. This requires multiple types of derived values, some of which are mentioned below and are monitored on a continual basis by the onsite controller or central analysis modules:

- Statistical Quantities: Various statistical measures like higher order moments of key parameters, moving statistical calculations, etc. See (Casella, G & Berger R. L, 1990) for more details.
- Evolving physical quantities: In addition to static measures or feature calculations, time evolving nature of the major parameters are critical to detect failures.
- Deviation from expected values: Most engineering parameters have pre-defined set values and are tracked for deviations from their set point values. A significant deviation and the direction of deviation is an indicator for certain failure modes and insipient failures in critical components.
- Model residuals: Increasing residuals between empirically derived models and observed values can give insights into impending failures and isolation using appropriate classification models.

2.3 Data Preprocessing for PHM Modeling

Prior to any subsequent PHM models being developed, raw data captured typically has to be processed to ensure that proper variations are captured, no biases are introduced and the underlying distributions generating real life data are modeled correctly. The following are some of the processes that one might consider during a PHM application:

- Filtering: raw sensor data is filtered on multiple dimensions to include relevant time periods, operating modes of equipment, ambient conditions and failure modes.
- Frame Specific Segmentation: sensor and controller data are sampled based on specific frame types to average out against biases arising from multiple designs and operating ranges of key parameters.

- c. Smoothing: protecting against biases in model development and parameter learning. Also, to prevent faulty rules due to outliers arising out of data quality issues, certain statistical or model based smoothening modules have been implemented.
- d. Data Quality: In addition to the above mentioned data preprocessing, it is also required to resolve data quality issues such as missing data, faulty sensor readings and out of range values.

Due to the nature of multiple key health indicators that are monitored, there was a need felt to implement different anomaly detection algorithms that can capture different failure modes. Different types of data driven methodologies have been developed in order to develop a robust PHM system – including soft computing (Bonissone, P., & Kai Goebel), reliability and remaining useful life estimation (see Ebeling (2005), machine learning methods and the fusion of these methods with physics concepts. In the following section we outline one such anomaly detection method that was implemented for our PHM case study. These anomaly detection methods would isolate abrupt changes in operation patterns, which are then used for subsequent analysis and decision making process.

2.4 Anomaly Detection Algorithm

Both online and batch mode implementations have been tested for anomaly detection, and an example algorithm is described below. Interested user can see following references for advanced analytics methods used: Kumar, Vipin et al(2005), (Russel S & Norvig P, 2002); (Duda R.O, Hart P.E & Stork, D.G 2001)

Multi-variate Hypothesis testing method Hotelling's T-Square was proposed by Harold Hotelling (Hu, Xiao, Qui, Hai and Iyer, Naresh, 2007), (M. Markou & S. Singh, 2007). It is a multivariate technique that captures the changes in data from multiple parameters by using their covariance information. This is a generalization of the Student's t statistic used in multiple hypothesis testing.

Consider a time series as:

$$X(t) = (X_1(t), X_2(t), \dots, X_m(t))^T$$

Where m is the number of variables sampled at each time t , $X_i(t)$ are the parameter sampled at regular intervals. Assuming $X(t)$ is a multi-variate normal $\sim N(\mu, \Sigma)$, where μ and Σ are the multivariate mean and co-variance matrix respectively.

The mean and variance μ and Σ are estimated as follows:

$$\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m)^T \quad (1)$$

$$W = \frac{1}{n-1} \sum_{i=1}^n (X(t) - \bar{X})(X(t) - \bar{X})^T \quad (2)$$

Hotelling T2 Statistic for $X(t)$:

$$T^2 = (X(t) - \bar{X})W^{-1}(X(t) - \bar{X}) \quad (3)$$

The Null Hypothesis states that $X(t)$ is NOT different from previous n samples or that no change occurred. Large values of T-Square statistic imply that the null hypothesis was not true. This was implemented using thresholds in Matlab™ (Mathworks). The test comprised of testing the Hotelling-T Square Statistic if it exceeded a present threshold, thereby confirming a change/anomaly at given time instant.

This statistic uses the statistical distance and incorporates multivariate variance-covariance matrix, to detect significant shifts and linear relationships.

We optimized the various parameter settings and threshold values for the anomaly detection module by analyzing different turbine frame types. This was then run for each of the monitored equipment health indicator and each time it was run, we fused the outputs from multivariate Hotelling-T square with other anomaly detection algorithms. This was done in order to capture both local anomalies (uni-variate sense) and system level anomalies (multivariate sense) across more than a single monitored parameter at same time instant.

2.5 Alarm Generation Process

Due to a large number of equipment that are being monitored and their key parameters, there is a definite need to keep the fleet wide alarm rates to have a high probability of detection with a very low false alarm rate. For this purpose, we implemented a time based alarming process and optimized it with respect to the field data, observed failure rates and deployed the algorithms with an ability to change the alarming settings based on fleet requirements.

3. PROFICY ADVANCED ANALYTICS TOOLSET

Proficy® is a suite of commercial-off-the-shelf software solutions that are designed to help solve the operations challenges in infrastructure and/or manufacturing industries. Proficy software suite offers the depth and breadth of capabilities from control and optimization.

Proficy software Suite provides a whole solution from data collection, data management, data visualization to data analysis for remote monitoring and diagnosis.

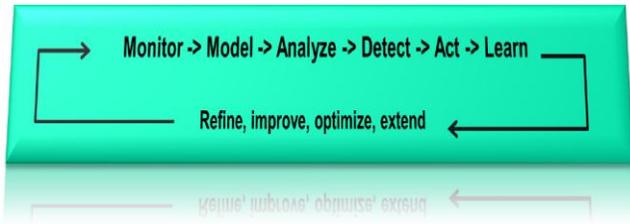


Figure 2. Remote Monitoring & Diagnosis Process

Proficy troubleshooter/cause+ (Figure 3) suite is the key software to perform advanced data analysis and knowledge discovery. It provides a platform to facilitate all steps of advanced analytics

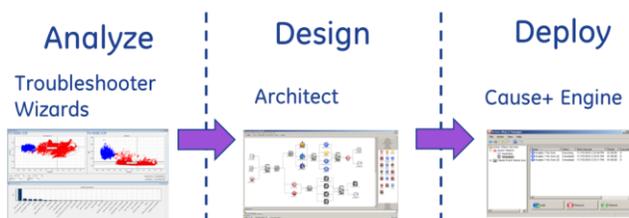


Figure 3. Troubleshooter/Cause+ Functionalities

Analyze: troubleshooter wizards provide rich analytics models and powerful visualization tools for subject matter experts to speed up the process of data exploration and knowledge discovery.

Design: Troubleshooter architect provides integrated development and simulation environment for application engineers to design and debug analytic solutions with pre-built database interface, various data preprocessing/post-processing techniques, and internal or external algorithms/analytics models

Deploy: Cause+ provides a deployment environment for operation managers to run the analytics solutions in real-time, event-triggered, or scheduled manners.

3.1 Offline Analysis

The Troubleshooter Wizards (continuous or discrete) guide users through troubleshooting processes in the developer environment. Using the Wizards, various tools are available to identify the causes of process deviation using historical data. Preparing data is made quick and easy using graph and trend views, and modeling the industrial process is intuitive. From there, knowledge about the process can be gleaned effectively and combined with the knowledge of expert personnel to develop an integrated solution to process problems.

This solution can then be further customized and tested within Architect, and deployed in real-time in the cause+ engine by Action Object Manager.

3.2 Solution Design

The Proficy Architect environment (Figure 4) enables the development of solution blueprints, used to visualize the process in the simulated mode. It contains user-friendly libraries and simple-to-configure blocks with which solution are developed. Various features in the menu and the explorer-type view are available for easy navigation in large solutions. A powerful troubleshooting compiler produces simulation and debugging on the execution ability of a blueprint after development.

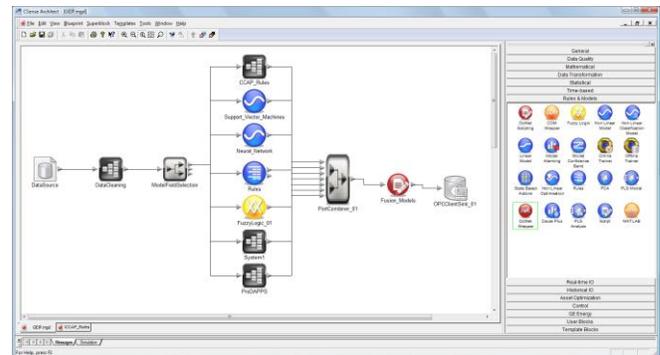


Figure 4: Proficy Architect IDE

This environment enables great flexibility in developing specific condition monitoring and decision-support solutions. The user can choose from high-level rapid prototyping tools to lower-level programming functions in scripting or in any of a number of programming languages that supports Component Object Model (COM, such as Visual C#/C++ and Visual Basic).

3.3 Online Deployment

The Action Object Manager provides a simple and easy way to deploy and monitor Action Objects. It allows users to easily maintain all your Action Objects from one central point of access. An Action Object (AO) is the name given to an executing blueprint. This blueprint could have been created and deployed from a number of services, including Architect, Troubleshooter Wizards, or other Proficy tools.

Proficy Advanced Analytics also provides a toolkit for fleet asset monitoring. The toolkit can apply the same action object to a large number of assets programmatically, which make fleet monitoring easier and more reliable.

3.4 Integration of Proficy with Existing Platforms

The Proficy Advanced Analytics software provides a series of interface tools (Figure 5) to integrate existing algorithms into the system (CSense Systems (Pty) Ltd). Some of the tools are General script, .NET script, COM Wrapper, .NET Wrapper, Matlab script. So users can plug their existing algorithms/ modules directly into the Proficy Advanced Analytics environment and seamlessly integrate with other part of the system. For example, Matlab code can be

plugged into the system directly (Matlab license is needed to run the code). Compiled Matlab objects can also be integrated. The COM Wrapper block allows the user to integrate external COM components within the Proficy Advanced Analytics environment. In this way the user can add custom functionality, own code/libraries, or third party libraries/components to the solution.

Those tools provide great flexibility to end users, who can easily re-use their existing modules develop and deploy PHM platforms. In summary, Proficy advanced analytics provides a series of tools from offline-analysis to online deployment so PHM users can focus on the most creative but challenging part of the job.

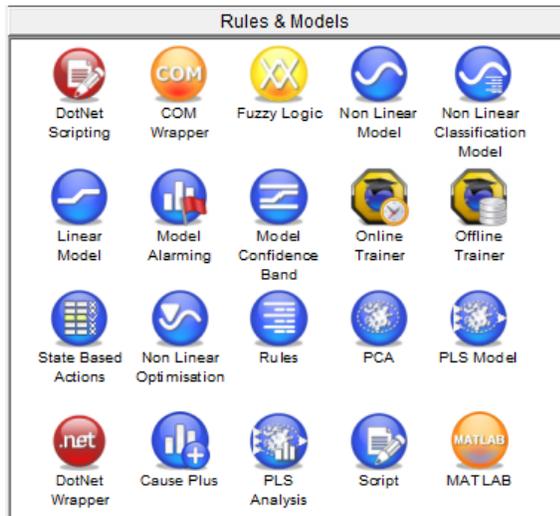


Figure 5: Tools in Proficy Advanced Analytics

4. IMPLEMENTATION

The PHM case study described for GE Energy was implemented on Proficy® platform. We tested the various PHM models developed on 200 units operational data for a six month duration at multiple time resolutions.

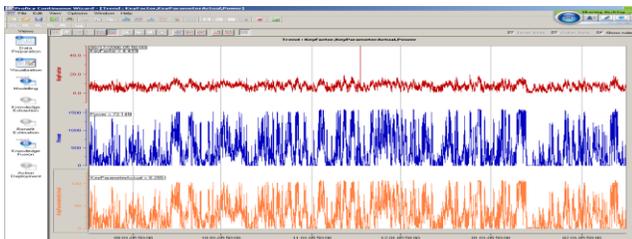


Figure 6: Multiple trending in Proficy®

Some of the data exploratory and model analysis done as explained in Section 2 earlier and was performed to understand the distributions of key parameters of interest, detecting outliers and evolution of key metrics over time. Figure 6 above depicts some of the basic plots explored in Proficy® toolset. As shown, multiple variables can be examined at instantaneous periods of time, understanding

basic distributions and plots such as scatter plots, line plots, overlay plots and histograms.

Deviations of various key parameters from their calculated values and set points can be crucial in a PHM advisory system. An example of such deviations is shown in Figure 7a.

Also, as depicted in Figure7b, some of these critical differentials are tracked over a period of time and early warnings can be picked up to perform enhanced monitoring of high risk units. Some of these can also be used to monitor certain frequently occurring failures and plan for inventory and spare parts.

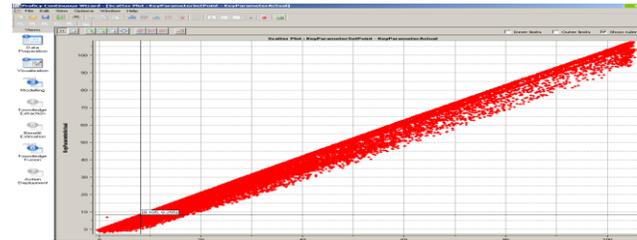


Figure 7a: Predicted Vs Actual Values

A total of over 1 million rows of operational data were analyzed in the above mentioned PHM case study. Initial results indicated a probability of detection over 80%. This is significant as there was little monitoring capability available for some of the turbine generation capability earlier and given that this is work in progress, we are hopeful to increase the probability of detection rates to very high values, keeping the false alarm rates under control.

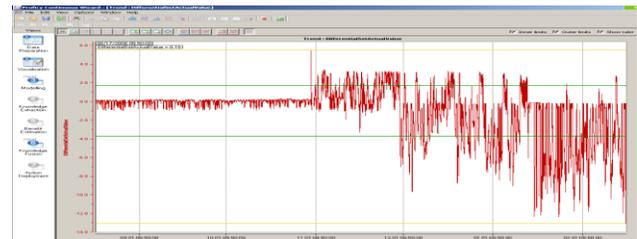


Figure 7b: Tracking evolution of deviations

The main objective is to understand and capture the patterns of increasing deviations and raise appropriate alarms for user to be able to perform exception based monitoring. This would ensure a high productivity and increased reliability.

5. CONCLUSION AND FUTURE DIRECTION

As depicted in this case study for GE Energy, a PHM system using real failures on key equipment was implemented using some of the existing platforms and using GE Intelligent Platform.

On the PHM algorithms side one of the key lessons learnt was to develop PHM algorithms with a high degree of explain-ability to end user: this ensures easy acceptance by field personal and relation of physics of failure with

advanced PHM methods seamless. As this is still work in progress, we plan to improve the quality of results by fusing multiple methods developed individually and using all sources of information available to monitor the equipment.

On the Proficy side, we plan on linking the current GEIP algorithmic capability with GE SmartSignals® algorithms and linking some of the existing legacy algorithms with the Proficy toolset. Also in pipeline is to enhance the native prognostic methods capability within Proficy, thus increasing its analytical power to include advanced methods.

The authors would like to acknowledge that Matlab is a trademark of Mathworks (<http://www.mathworks.com>) and .NET is a trademark of Microsoft (<http://www.microsoft.com>).

REFERENCES

- GE Intelligent Platforms. <http://www.ge-ip.com>
Mathworks: <http://www.mathworks.com>
CSense Systems (Pty) Ltd. User's manual for Proficy Advanced Analytics 5.0 (2011)
Casella, G (1990); Berger R. L Statistical Inference.
Duda R.O (2001); Hart P.E; Stork, D. G: Pattern Classification. John Wiley & Sons.
Hu, Xiao(2007), Qui, Hai and Iyer, Naresh: Multivariate Change Detection for Time Series Data in Aircraft Engine Fault Diagnostics, IEEE.
M. Markou (2003), S. Singh, Novelty Detection : A Review Part 1: Statistical Approaches, Signal Processing, Vol. 83(12), pp2481-2497.
Kumar, Vipin et al(2005): An Introduction to Data Mining.
Russel S(2002), Norvig P; Artificial Intelligence, A Modern Approach, Prentice Hall of India, 2nd edition.
- Jammu, Vinay(2010), et al; Review of Prognostics and Health Management Technologies, GE Global Research
Bonissone, P., & Kai Goebel, Soft Computing Techniques for Diagnostics and Prognostics
Ebeling (2005), C. E., An Introduction to Reliability and Maintainability Engineering, Tata McGraw-Hill Publishing Company, New Delhi.
Vachtsevanos(2006), G., Frank Lewis, Michael Roemer, Andrew Hess and Biqing Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley & Sons, Hoboken, New Jersey, 2006
- Subrat Nanda** is a Lead Engineer at GE Global Research Center. He earned his B.S degree in Production Engineering from Nagpur University in India in 2001 and M.S degree in Autonomous Systems from University of Exeter, England UK in 2003. His research interests are mainly in applying machine learning and pattern recognition methods to industrial problems such as remote monitoring and PHM, Bayesian reasoning, bio-inspired optimization and fusion of data mining methods with physics.
- Xiaohui Hu, Ph.D.** Dr. Hu is Principal Engineer at GE Intelligent Platforms. He earned his B.S. degree in electrical engineering from Tsinghua University, Beijing, China in 1995 and Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, US in 2004. His research interests are remote monitoring and diagnosis, computational Intelligence, swarm intelligence, and data mining and knowledge discovery.