# Enhancing Turbine Performance Degradation Prediction with Atmospheric Factors

Xiaomo Jiang[1], TsungPo Lin[2], and Eduardo Mendoza[3]

[1]*General Electric Company, Power Services, Monitoring and Diagnostics, Atlanta, GA 30339, USA*
*xiaomo.jiang@ge.com*

[2,3]*General Electric Company, Power Services, Performance, Atlanta, GA 30339, USA*

*tsungpo.lin@ge.com*
*Eduardo.mendoza@ge.com*

## ABSTRACT

Heavy duty gas turbine engines are not only ingesting the air, but also eating a myriad of aerosol particles, which may have various negative effects on the turbine operation efficiency as well as the component failure. This paper attempts to develop predictive degradation models for gas turbines by integrating satellite collected atmospheric factors, on-site monitoring data, and physics-based calculated performance results. Multiple variables are analyzed and employed for predictive modeling. The vital variables are identified by using data exploratory correlation analysis and stepwise regression analysis. The performance degradation calculation is obtained from physics based thermodynamic heat balance of gas turbine. It requires balancing mass and energy of gas turbine to match measurement data through thermodynamic cycle matching. The performance degradation prior to the offline water wash is used as the predictor. Artificial neural network modeling is employed to establish the predictive models. A procedure is presented to explain the proposed methodology, and results are discussed. This paper provides an effective methodology and procedure to apply big data for the performance degradation prediction of gas turbines.

## 1. INTRODUCTION

Gas turbine (GT) simple or combined cycle plants are built and operated with higher availability, reliability, and performance in order to provide the customer with sufficient operating revenues and reduced fuel costs meanwhile enhancing customer dispatch competitiveness (Jiang and Foster 2013 & 2014). The availability of heavy duty gas turbines in the plant can be increased through increasing the turbine reliability by maintenance enhancement and recovering performance degradation by remote efficiency monitoring to provide timely corrective recommendations (Balevic et al. 2010, Brooks 2000, and Johnston 2000). In addition, increasing fuel costs requires maintaining the higher efficiency in a gas turbine system. For example, a combined cycle plant with the engineered capacity of 900MW power output can have an annual fuel bill of over 200 million dollars (Meher-homji *et al.* 2001). Therefore, remote performance degradation monitoring, diagnostics, and prognostics of power generation equipment like heavy duty gas turbines has become increasingly important and popular in the energy industry since its introduction in the 90's.

Heavy duty gas turbine engines, however, are not only ingesting the air, but also eating a myriad of aerosol particles, which may have various negative effects on the turbine operation efficiency including performance degradation, compressor fouling, inlet effectiveness and cooling-hole plugging, as well as the component failure such as compressor blade cracks and corrosion in both cold and hot sections. Gas turbine performance is a function of many factors including, but not limited to, turbine design, component technology upgrade, operating mode, and site ambient conditions plus the environmental factors. The turbine degradation accumulates as its operating hour increases. This study is focused on predictive modeling of performance degradation for gas turbines by integration operational variables and atmospheric factors, with the purpose of enhancing customer specific maintenance recommendation, optimal outage planning, asset management, and failure prevention.

The performance degradation can be categorized into two types: recoverable and non-recoverable, as illustrated in Fig. 1. The recoverable degradation can be recovered by proper maintenance actions such as cleaning compressor via regular water washes and parts replacement or upgrade during major inspection. Usually most of the degradation resulted from the equipment operation can be recovered through proper offline water wash of its compressor, while the degradation resulted from the mechanical deterioration (e.g., hot gas path component wear/damage) or parts malfunction (e.g., compressor bleed valve open) can be recovered from an overhaul. On the other hand, the non-recoverable degradation becomes permanent deterioration on the gas turbine even after a major overhaul. The focus of this study is on the recoverable performance degradation.

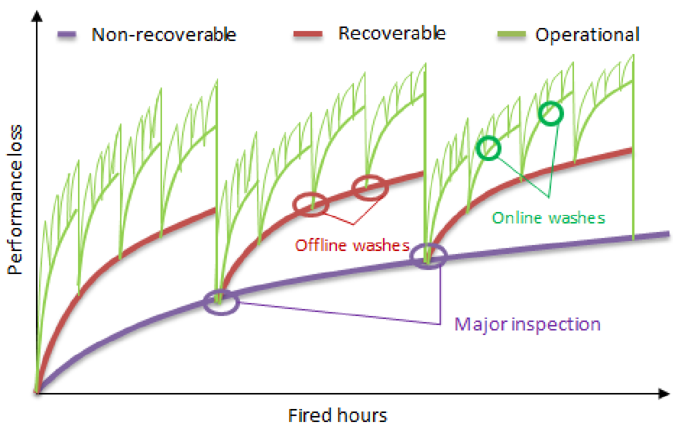


Figure 1. Performance degradation components

In next sections, data used in this study, including the on-site monitoring (OSM) operation and atmospheric (ATM) factors, are first described. The methods used for data pre-processing, performance calculation and neural network predictive modeling are then presented. The physics based thermodynamic heat balance of gas turbine is employed to correct the measured performance to the ISO conditions. It requires balancing mass and energy for the turbine to match measurement data through thermodynamic cycle matching. The procedure to implement the methodology is given. After then, the methods and procedure are applied to the data and to yield results.

## 2. DATA

It has been well recognized that there are a number of time-varying site operational conditions impacting the turbine performance (Meher-Homji *et al*. 2000). These conditions include, but not limited to, ambient humidity, ambient pressure, ambient temperature, inlet filter pressure losses, exhaust system pressure losses, fuel heating value, fuel flow, and fuel temperature. In addition, a vast sea of global information on atmospheric factors is available on public internet. These atmospheric (ATM) pollution factors, which may affect the performance of turbo-machinery, include SO2 gas, Sulfate aerosols, sea salt aerosols, and PM2.5.

The data set used in this study includes 672 performance degradation events collected from 194 different gas turbines. The degradation is calculated right prior to the offline water wash outage. Only GE manufactured F-class gas turbines are used in this paper for demonstration purpose. For each event, there are 37 on-site monitoring (OSM) operation variables (part of them shown in Table 1) and 11 ATM factors.

In this study, the baseload operation time series data is used to calculate the degradation for each event. The data is filtered out to be baseload mode based on three criteria, i.e., the inlet guide vane (CSGV) is full open; the turbine shaft (TNH) is full speed, and the exhaust temperature (TTXM) is close to the reference control temperature (TTRX).

Table 1. Part of OSM performance variables

| OSM Tag | units | descriptions | Example data |
|---|---|---|---|
| **AFPAP** | inHg | Ambient Pressure | 30.2 |
| **AFPCS** | inH2O | Inlet air total press transmitter | 4.8 |
| **AFPEP** | inH2O | Exhaust press transmitter | 12.8 |
| **AFQ** | lbm/s | Compressor Inlet Air Flow | 1422.5 |
| **CMHUM** | #H/#A | Specific Humidity | 0.006 |
| **CPD** | psi | Compressor disch press transmitter | 224.4 |
| **CPR** | ratio | Compressor Pressure Ratio | 16.3 |
| **CSGV** | º | Position feedback IGV (high value selected) | 88.0 |
| **CTD** | °F | Compressor Discharge Temperature | 732.2 |
| **CTIM** | °F | Max Comp Inlet Flange Temperature | 53.2 |
| **DWATT** | MW | Generator watts | 255.3 |
| **FQG** | lbm/s | Gas Fuel Flow | 33.9 |
| **FTG** | °F | Fuel Gas Temperature | 349.3 |
| **TNH** | % | HP Turbine Speed | 99.993 |
| **TTRX** | °F | Temperature Control Reference | 1137.3 |
| **TTXM** | °F | Ex Temp Median Corrected by Average | 1137.3 |

The atmospheric pollution data is obtained from NASA's (National Aeronautics and Space Administration) Earth Observation System (EOS), http://eospso.gsfc.nasa.gov/. The EOS is a coordinated series of satellites meant for long-term global observations of the land surface, biosphere, solid Earth, atmosphere, and oceans.

Satellite data can provide global information on atmospheric factors of interest to turbo machinery. Of particular interest is aerosol optical thickness, which can provide information regarding soot/ash from fires, desert and soil dust, ash/chemical species (SO2) from volcanoes and fossil fuel burning, marine aerosols such as suspended sea salt due to wave action, dimethyl sulfide from phytoplankton, and smog/vog due to pollution/volcanic haze augmented by chemical reactions in atmosphere, including photochemical reactions.

Note that aerosols are the tiny airborne particles that come from forest fires, deserts, volcanoes, breaking ocean waves, and urban and industrial pollution. Aerosols play an important role in the earth system, directly influencing global climate and human health. Satellite remote sensing data shows promise for predicting atmospheric effects on turbo machinery by geo-location, elevation, and season. There are more than 1000 different parameters and over 1600 data types that currently than can be downloaded by the public.

The sample categories likely impacting the degradation of heavy duty gas turbines have been pre-selected by a team of internal subject matter experts as shown in Table 2. These data are pre-processed to be monthly mean data, and can be obtained from the satellite/instrument or chemical transport models maintained by NASA, JPL (Jet Propulsion Laboratory), and NRL (Naval Research Laboratory). In this study, the 12 months averaged data is used for all 11 variables to consider the seasonal effect for each event.

Table 2. Weather and aerosol variables

| **Variable** | **Units** | **Description** | **Example** |
|---|---|---|---|
| PS | Pa | Time averaged surface pressure | 98695.0 |
| QV2M | $kg/m^3$ | Specific humidity 2m above displacement height | 0.0069 |
| T2M | K | Temperature 2m above displacement height | 284.0 |
| DUCMASS2.5 | $kg/m^2$ | Dust Column Mass Density (PM 2.5) | 6221.8 |
| DUCMASS | $kg/m^2$ | Dust Column Mass Density | 16316.8 |
| SSCMASS2.5 | $kg/m^2$ | Sea Salt Column Mass Density (PM 2.5) | 656.5 |
| SSCMASS | $kg/m^2$ | Sea Salt Column Mass Density | 1891.7 |
| SO2CMASS | $kg/m^2$ | SO2 Column Mass Density | 10928.4 |
| SO4CMASS | $kg/m^2$ | SO4 Column Mass Density | 7384.3 |
| OCCMASS | $kg/m^2$ | Organic Carbon Column Mass Density | 947.2 |
| BCCMASS | $kg/m^2$ | Black Carbon Column Mass Density | 7353.4 |

The performance degradation is assessed based on the performance output. To facilitate the data analysis and predictive modeling, the performance calculation is conducted first on the OSM variables and configuration parameters to yield the corrected performance output. The performance degradation percentage is calculated for each event from the corrected output in regard to the baseline value upon its commissioning. As such, the obtained degradation percentage incorporates the performance operational influencing parameters via the physics-based performance modeling. Next, the degradation percentage is further analyzed with the atmospheric factors to identify vital X for predictive modeling.

## 3. METHODS

### 3.1. Performance Calculation

The corrected output and heat rate are needed to be compared with the established baseline values to estimate the level of turbine performance degradation. For comparison and assessment purposes, these typical GT performance parameters, including efficiency, flow, output, and heat rate, need to be continuously corrected to the desired conditions, such as design or ISO (International Standard Organization) conditions. Two sets of performance calculation methodologies are available for performance calculation and correction, namely, thermodynamic modeling and data-driven factor interpolation method. The thermodynamic modeling approach calculates the performance of the gas turbine using physics-based thermodynamic cycle matching. Inputs into this approach include three parts: dynamic OSM data, static condition data, and equipment configuration data. The time-dependent dynamic inputs are first merged with the static data from the database. These static inputs contain the performance of the gas turbine at ISO condition, designed condition and certain accessory options available for the unit. Next all these inputs are merged with the unit configuration data to generate an input file for performance calculation. Then, a data reduction technique is employed to satisfy continuity and conserve energy of the gas turbine using measured OSM data. Finally the performance of the gas turbine is calculated and corrected via thermodynamic cycle matching method at baseload and ISO conditions.

A data-driven adaptive approach, called factor interpolation algorithm, as documented in ASME test procedure (ASME PTC 22-2005) for field test of plant overall performance, may be employed to calculate the corrected output and heat rate using site-specific correction factors, raw OSM plant data, performance test methodology and baseline test information. In this method, a set of correction factor curves are pre-established from the abovementioned physics-based heat balance model given the desirable conditions.

In this study, the thermodynamic modeling is employed. The gas turbine performance is first calculated via a thermodynamic cycle matching to a set of measured parameters such as compressor discharge temperature and pressure, exhaust temperature, fuel flow, and power output. This matching procedure is used to synthesize other non-measured performance parameters, such as compressor and turbine efficiencies, combustor exit temperature, and turbine firing temperature, for the gas turbine operated at the specified boundary conditions. In this method, the turbine mass flow rate ($f_{gt}$) is calculated by utilizing the stage 1 nozzle throat area ($A_{S1N}$), stage 1 nozzle flow coefficient

($\alpha_f$), turbine inlet pressure ($P_{in}$), turbine total inlet temp ($T_{in}$) and the flow calculation function ($g_{air}$), given as follows:

$$f_{gt} = g_{air} \times P_{in} \times A_{S1N} \times \alpha_f / \sqrt{T_{in}} \quad (1)$$

The compressor mass flow rate ($f_{comp}$) follows the law of conservation of mass, i.e.

$$f_{comp} = f_{gt} + f_{ext} - f_{fuel} \quad (2)$$

where $f_{ext}$ and $f_{fuel}$ are the extraction flow and fuel flow, respectively.

The compressor and turbine power values are then calculated via the thermodynamic heat balance with the law of conservation of mass and energy. The tuned gas turbine model is then used to project how this turbine would perform at other operating conditions (assuming nominal GT operating characteristics).

### 3.2. Performance Degradation

The ultimate purpose of this study is to establish a predictive model in order to project the performance degradation to next planned outage or a given fired hours for facilitating critical business decision making such as resource allocation, hardware upgrade, and part procurement. As shown in Figure 2, the projection analytics are usually called degradation prognostics, which utilize the historical performance data, degradation trend, and physics-based simulation to predict the degree of degradation at a given point in time (e.g., next planned outage).

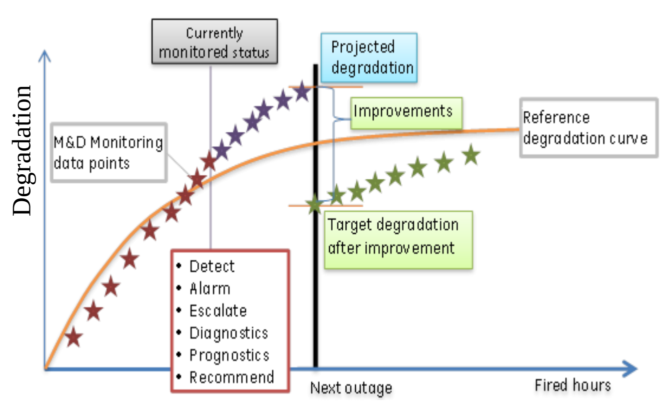


Figure 2. Thermal performance monitoring: Escalation and prognostics

The percentage degradations for power output, %OP_deg, can then be calculated by:

$$\%OP\_deg = (OP\_bl - Cor\_OP)/OP\_bl \quad (3)$$

where Cor_OP and OP_bl are the corrected performance output at the operational time and the baseline output value at zero fired hour, respectively.

After proper actions are taken, the performance degradation is corrected to the target level. The diagnostics and prognostics information continues to be broadcast via multi-channels including telephone, email, and web. Recommendations regarding performance degradation may be projected to the next planned outage for facilitating business decision making such as resource allocation and part procurement. These projection analytics are usually called performance degradation prognostics, which utilize historical performance data, degradation trends, and physics-based simulation. Through performance prognostics from the current status of a given unit or component, the monitoring system enables the trade-off analytics to facilitate critical decision-making regarding hardware upgrades, maintenance scope, and resource allocation. For example, thermodynamics based simulation of the total plant may be employed to quantify how much the performance improvement can be achieved for a certain hardware upgrade such as enhanced inlet filtration, advanced hot gas path, enhanced compressor package, or steam high pressure section upgrade.

As an example, Figure 3 shows the performance output degradation over the fired hours with offline water wash events. A majority of performance degradation is recovered from a clean offline water wash (OFWW), as illustrated by the difference between the pre-OFWW degradation (black dot) and the post-OFWW degradation (green dot).

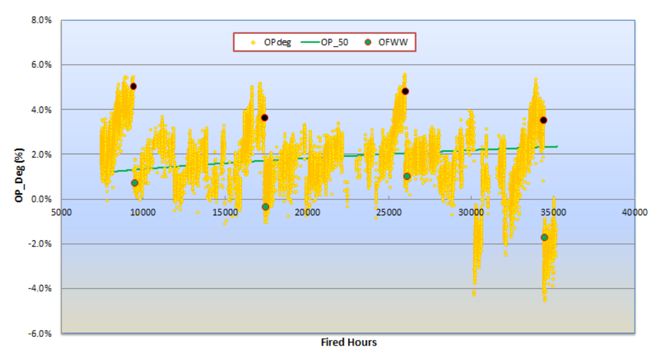


Figure 3. Performance degradation trend with offline WW

### 3.3. Data Preprocessing

It is well recognized that the quality of the input data impacts the accuracy of predictive modeling. The sensed OSM data and the processed atmospheric data usually contain error, missingness, incoherence, and imprecision during the data acquisition, communication, and processing. Therefore, the quality of the input data should be validated prior to further application for performance calculation. Data preprocessing techniques are applied to improve the quality of the data. These techniques include data validation, outlier analysis, and data filtration, which are briefly described below.

#### 3.3.1. Data validation

The aim of data validation is to verify the reliability of sensor data or ATM factors during the normal operational service of the machine. As the easiest way, the widely-used graphical plot should be applied in data preprocessing to visually check the data quality. The simple statistical analysis should be also performed on the data to obtain its minimum, maximum, mean, median, and standard deviation. These values are often used to investigate whether the data points fall in the reasonable, applicable range based on the engineering judgment or physics understanding. As an example, Figure 4 shows the box plot of the atmospheric factor SSCMASS. A few data points seem to be outliers based on the statistical plots. Note that multi-modal distribution form may be used to identify the outliers more accurately.

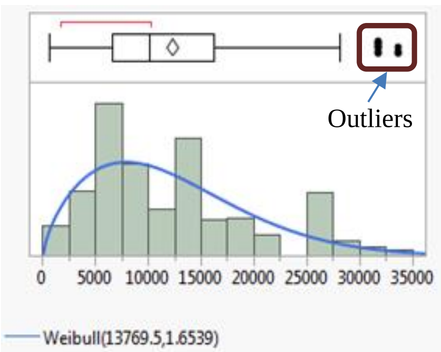


Figure 4. Box plot of SSCMASS data

#### 3.3.2. Outlier analysis

Outliers tend to pull the mean value towards themselves and inflate the variance in their direction. The outliers will largely affect the moment characteristics of the data. Therefore outlier analysis should be conducted on the input data, but outliers should be excluded from the data for further analysis only with proper justification. Some data points may be inconsistent with the expectation of the majority elements of the series. These data points are usually referred to be outliers. These outliers may result from measurement errors and anomaly, which cannot be used to represent the normal operational condition of that unit.

The outliers are often identified via a box-plot. The box-plot invented by Tukey (1977) (also known as a box-and-whisker diagram or candlestick chart) is an exploratory data analysis approach to graphically depict the five-number summary, including the minimum, lower quartile (25%), median, upper quartile (75%), and maximum value, and to indicate the outliers in the data. The outlier is defined as any data observation which lies more than 1.5*IQR (inter-quartile range) and lower than the lower quartile (25%) or 1.5*IQR higher than the upper quartile (75%), in which the IQR is calculated by subtracting the lower quartile from the upper quartile. The box-plot is an important exploratory data analysis technique which is able to visually show different types of populations without any assumptions of statistical distribution. A box plot for SSCMASS shown in Figure 4 indicates a few outliers (marked as dots in the figure).

#### 3.3.3. Data filtration

As a supplementary to the outlier analysis, data filtration is usually performed to ensure that the sensor data represents the unit under normal operation via truncating sensor data in the reasonable operating range. Generally critical variables suggested by the component design team should be used to define the normal operating status of the unit. For example, the turbine HP shaft speed (TNH) with the range of 95% and 105% should be used as one of the critical factors to determine whether that unit is under normal operation or not.

The performance data during the downtime is often treated as noise, and should not be included in the statistics calculation of the performance variables. Various performance variables should have different operating ranges. Before establishing a predictive model, it should be ensured that the proper range for each variable is obtained from OEM (Original Equipment Manufacturer) manual, material standard, or design team.

### 3.4. Vital X’s Identification

In order to decide the vital X’s or predictors, a combination of best engineering assessment of the underlying physics of the failure mechanism, statistical analytics, tribal knowledge, expert opinion, and available data may be used comprehensively. The analytics methods may include analysis of variance (ANOVA), correlation analysis, although not all methods are needed in the identification of the vital X’s. In this study, the correlation analysis and stepwise regression analysis are employed to identify the vital X’s for predictive modeling.

#### 3.4.1. Correlation analysis

Correlation analysis is another exploratory data analysis method widely used to measure the dependence between two variables, quantitatively or qualitatively or both. The measurement scales used should be at least interval scales (e.g., [0, 1] or [-1, 1]). The simplest way to find out qualitatively the relationship between two variables is to plot the data. However, the graphical approach is feasible only when there are only a few variables in the problem. When there are many variables (e.g., >5), the graphical method becomes both labor and time consuming. In addition, the qualitative method cannot provide objective information on the judgment of the correlation between two variables. Therefore, quantitative correlation analysis may be pursued to more accurately identify the relationship between two variables.

Pearson correlation is the most familiar quantitative approach to measure the dependence. It divides the covariance of the two variables by the product of their standard deviations. The value of Pearson correlation coefficient may have the range between -1 and 1. A larger value near +1 or -1 implies the stronger interdependence between two variables, while a smaller value near 0 indicates little dependence between them. The probability of un-correlation between two variables, obtained by t-statistics and called p-value, is also used to quantitatively assess the correlation. The two variables are usually judged to be correlated if p-value is smaller than 0.05 or the correlation coefficient is greater than a predefined value (e.g., 0.50).

The most widely-used type of correlation analysis is Pearson R coefficient, also called linear or product- moment correlation. Pearson correlation assumes that the two variables are measured on interval scales (e.g., $-1.0 \le R \le 1.0$). It determines the extent to which values of the two variables are "proportional" to each other. The value of correlation (i.e., correlation coefficient) does not depend on the specific measurement unit of each variable; for example, the correlation between PS (Surface Pressure) and QV2M (Specific humidity) will be identical regardless of which units are used for the two variables. Proportional means linearly related; that is, the correlation is high if it can be "summarized" by a straight line (sloped upwards or downwards). Given two data sets $X_1$ and $X_2$ collected from two variables $x1$ and $x2$, respectively, the Pearson correlation coefficient between the two variables, $R_{X_1,X_2}$, can be obtained by dividing the covariance of the two variables by the product of their standard deviations, expressed as follows:

$$R_{X_1,X_2} = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]}{\sigma_{X_1}\sigma_{X_2}} \qquad (4)$$

where $\mathrm{cov}(X_1, X_2)$ is the sample covariance of the two variables, and $\mu_{X_i}$ and $\sigma_{X_i}$ are the sample mean and standard deviation values of the variable $x_i$, respectively. In general, the coefficient $R_{X_1,X_2}$ falls in the range between -1.0 and 1.0 with the magnitude and the sign of $R_{X_1,X_2}$ representing the *strength* and *direction* respectively of the dependence between the two variables.

A hypothesis testing is often used in the Pearson correlation analysis to test whether the two variables are correlated in terms of the sample data. In this context, the *null hypothesis* asserts that the two variables are not correlated, while the *alternative hypothesis* asserts that the variables are correlated. A T-statistic is used to test the hypothesis. The observed value of T-statistic, called T-value, can range between -∞ (infinity) and +∞. A T-value near 0 is to support the null hypothesis that there is no correlation between the two variables, while a T-value far from 0 (either positive or negative) is to support the alternative hypothesis that there is correlation between the variables. The T-value (or T-statistic) is defined as:

$$T_R = R_{X_1,X_2}\sqrt{(N_X - 2)\big/(1 - R^2_{X_1,X_2})} \qquad (5)$$

Clearly, if the correlation coefficient *R* is either -1 or +1, the T-value is represented by NULL. Usually a *p*-value is used to represent the T-statistic. The *p*-value is the probability that the absolute value of the T-statistic at the significant level (e.g., $\alpha$ = 5%) would equal or exceed the observed value, i.e., $T_\alpha \ge T_R$ when the null hypothesis is true. A small *p*-value is to judge that the null hypothesis is false and the two variables are, in fact, correlated.

#### 3.4.2. Stepwise regression analysis

Stepwise regression analyses is further performed to identify the vital X's, particularly for multivariate analysis. In stepwise regression method, predictive variables in a regression model are selected through an automatic procedure using a sequence of F-tests. Other techniques may be used to select the predictive variables, such as t-tests, adjusted R-square, Akaike information criterion (AIC), Bayesian information criterion (BIC), Mallows' $C_p$, or false discovery rate. In the multivariate analysis, the predictive response is continuous variable (i.e., degradation level). Therefore, the linear is employed in the stepwise regression analysis to identify vital X's. Both AIC and BIC are employed as the criteria in the parameter analysis.

### 3.5. Neural Network Predictive Modeling

Neural network model is established to project performance degradation level given a certain period of operation. The model is formulated as a function of vital Xs identified previously.

Artificial neural networks (ANNs) are a family of statistical learning models inspired by biological neural networks, the central nervous systems of animals in particular the brain, as shown in Figure 5. They are usually used as a non-parametric approach to estimate or approximate functions that depend on a set of inputs and outputs. An ANN model is generally presented as a system of interconnected neurons which communicate messages to each other. The connections have numeric weights which are tuned based on the data and experience, making the neural net adaptive to inputs and capable of learning. The ANN method is widely used to model complex problems with implicit relationship among the variables which cannot be clearly explained mathematically.

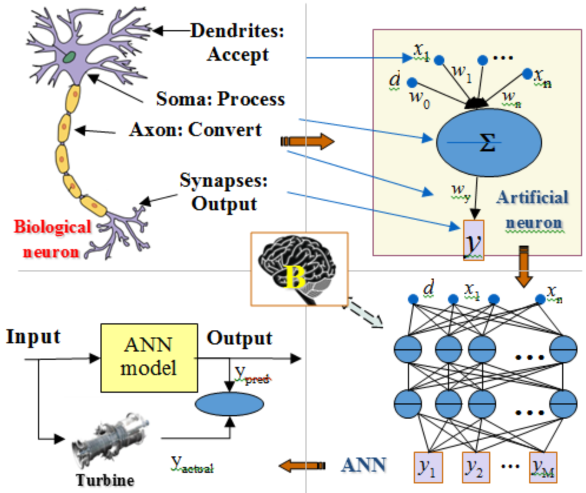

Figure 5. Illustration of neural network

There are two major learning paradigms, namely, supervised learning and unsupervised learning. In the supervised learning, a set of data pairs (*X*,*Y*) is given and its aim is to find a function *f*: *X*->*Y* matching the data set. The commonly used cost, mean-squared error is usually used to minimize the average squared error between the ANN output, *f*(X), and the target value Y over all the modeling or training data set. The supervised learning ANN is employed in this study.

In the unsupervised learning, only a set of input data *X* is given to minimize the cost function, which is employed in this paper. Priori assumptions are usually needed for the model parameters and observed variables. The neural network model may be in general represented as follows:

$$\hat{y}_k = \sum_{i=1}^{M} w_i \sum_{j=1}^{D} \varphi\left(X_{kj}\right) + \sum_{j}^{D} b_j X_{kj} + d \qquad (6)$$

where

$Xi = \{x_{i+1}, \ldots x_{i+n}, y_{i+1}, \ldots, y_{i+n}\}$ = Input vector

$\hat{y}_k$ = Predicted response quantity

$D$ = Input dimension

$M$ = Number of functions

$w$, $a$, $b$ = Parameters to be estimated

$\varphi(.)$ = Nonlinear activation function, a logistic function used in this study

## 4. IMPLEMENTATION PROCESS

Figure 6 shows the process to implement the methodology described previously. It consists of four main parts: data collection and preparation, performance calculation, predictive modeling, and results analysis & reporting. Refer to the above sections for details.

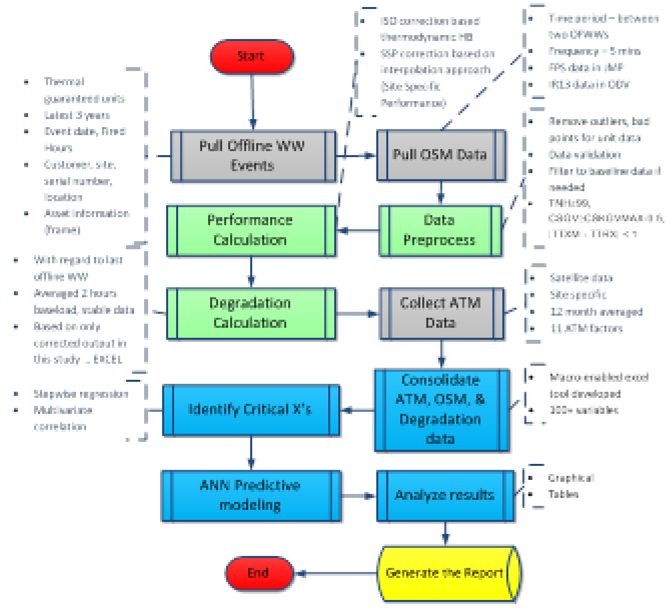

Figure 6. Implementation process

## 5. APPLICATION

Following the process shown in Fig.6, the performance calculation is conducted on all units under investigation to obtain the performance degradation for each unit prior to offline WW event. After data consolidation and outlier analysis, 126 units with 104 parameters are used in this application, of which there are 11 ATM factors and the rest are either operating factors or calculated performance factors. The performance degradation prior to offline WW is employed as the predictor, while 103 parameters are used as potential influencing factors. The correlation analysis is used to identify the independent variables from the ATM factors, while the stepwise regression method is employed to identify the critical variables from the other factors. As an example, Figure 7 shows the data versus the operation time at offline WW in terms of performance degradation, compressor discharge temperature (CTD), fuel flow (FQG), and specific humidity 2m above displacement height (QV2M).

### 5.1. Correlation Analysis

The correlation analysis is applied to identify the independent variable from the 11 ATM factors and the other OSM factors separately. For instance, from the multivariate Pearson correlation analysis by using Eq. (4), we can obtain 5 relatively independent ATM variables, PS, QV2M, DUCMASS, SO2CMASS, and OCCMASS, which will be included for further analysis. For demonstration purpose,

Figure 8 shows the color map of the correlation coefficients, where the deep red color indicates strong correlations between two factors, while the light color indicates weak correlations.

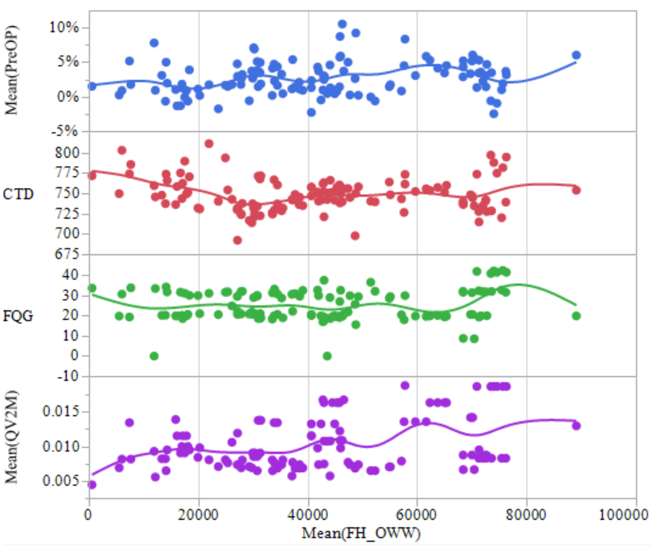


Figure 7. Data versus operation time at offline WW

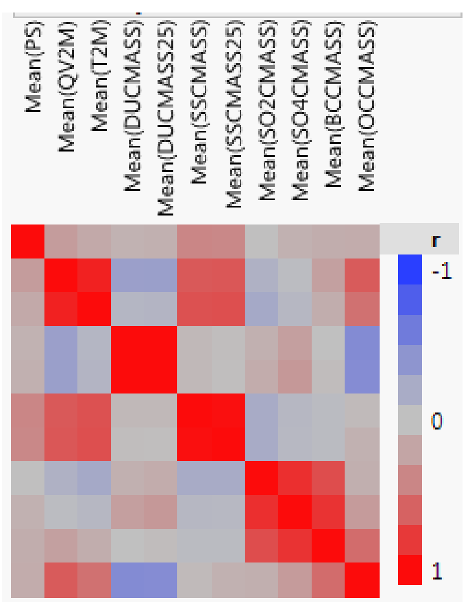


Figure 8. Color map of correlation coefficients

## 5.2. Stepwise regression

The stepwise regression analysis is further employed to identify the critical factors influencing the performance degradation. The trend of AIC and BIC criteria is shown in Fig. 9. The list of 10 factors with minimum BIC value is identified as the vital X's from the OSM variables

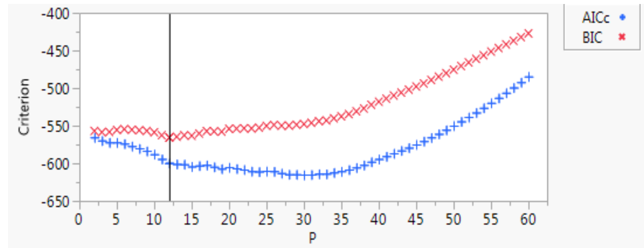


Figure 9. AIC and BIC Criterion trend in Vital X's Selection

## 5.3. Neural Network model

In the neural network modeling, a learning rate of 0.1 is used to fit an additive sequence of models. The 6-folder validation strategy is employed in the model generation. The 114 data points are randomly selected for model training and then the rest 12 data points are used to validate the model. We select 2-hidden layers neural network structure first, and then determine the number of hidden nodes by a trial and error approach.

Figure 10 shows the ANN structure with 15 input nodes (factors), 15 first hidden nodes, 5 second hidden nodes, and 1 output node representing for the performance degradation data. The R-squared value of 0.905 is obtained for model training, while 0.977 for model validation, implying the model produce the high prediction accuracy. Figure 11 shows the plots of actual degradation versus the predicted results for a) model training and b) model validation. It is observed that most data is clustered around the 1:1 plots in both cases, indicating an acceptable predictive ANN model.

## 6. SUMMARY

Performance assessment and prediction of a gas turbine provides valuable identification and understanding of which components were suffering from rapid or severe degradation and required technical assistance to resolve the deficiencies. This paper attempt to develop predictive degradation models for GTs by integrating satellite collected atmospheric factors, OSM data, and physics-based calculated performance data. Total 127 assets with 100 variables are analyzed and employed in this study. Outlier analysis and preprocessing are used to clean the data for modeling. Fifteen vital X's, including 5 ATM factors and 10 OSM factors, are identified by using correlation analysis and stepwise regression analysis. The performance degradation calculation is obtained from thermodynamic heat balance of gas turbine with the measurement data. The performance degradation prior to the offline WW is used as the predictor. Artificial neural network modeling is employed to establish the predictive models. A procedure is presented to explain the proposed methodology. This is a multivariate-input-single-output predictive problem. The results show that the ANN model presents acceptable prediction accuracy.

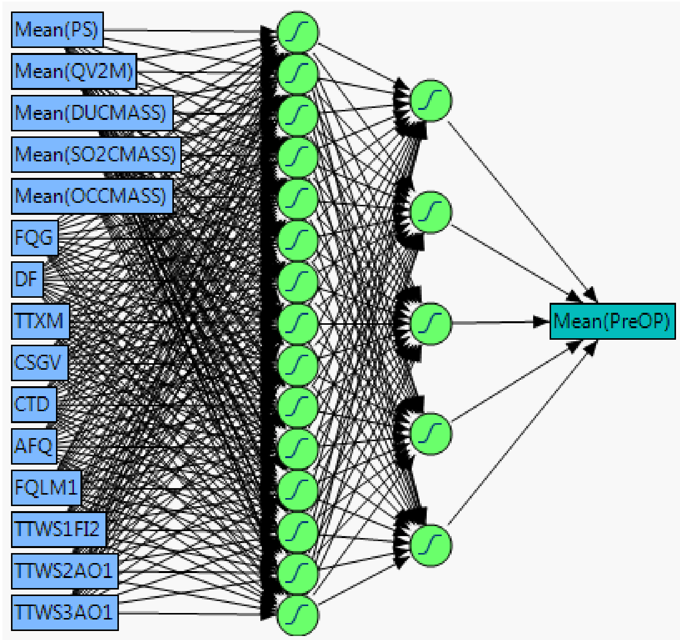


Figure 10. Multilayer neural network for performance degradation prediction

Current study is focused on the yearly averaged data as inputs, and the degradation is treated as a static output. Further research may be conducted to develop time-dependent degradation predictive model. In addition, only 10 ATM factors have been investigated in this study, more ATM factors if available may be investigated in future, associated with other predictive modeling techniques. The developed predictive model would be applied for outage optimization and predictive maintenance of assets, in order to reduce the operational and maintenance cost for customers.

## REFERENCES


ASME (2006), Performance Test Code on Gas Turbines. American Society of Mechanical Engineers (ASME), ASME PTC 22-2005.

Balevic D, Hartman S and Youmans R (2010), Heavy-Duty Gas Turbine Operating and Maintenance Considerations. GER-3620L.1, GE Energy, Atlanta, GA.

Brooks FJ (2000), GE Gas Turbine Performance Characteristics. GER-3567H. GE Power Systems, Schenectady, NY.

Jiang X, Foster C (2013), “Remote Thermal Performance Monitoring and Diagnostics – Turning Data into Knowledge,” Proceedings of the ASME 2013 Power Conference, July 29 – August 1, 2013, Boston, Massachusetts, USA.

Jiang X, Foster C (2014), “Plant Performance Monitoring and Diagnostics – Remote, Real-Time and Automation,” Proceedings of ASME Turbo Expo 2014: Turbine Technical Conference and Exposition, June 16 – 20, 2014, Düsseldorf, Germany.

Johnston JR (2000), Performance and Reliability Improvements for Heavy-Duty Gas Turbines. GER-3571H, GE Power Systems, Schenectady, NY.

Meher-Homji CB, Chaker MA and Motiwala HM (2001), “Gas Turbine Performance Deterioration,” Proceedings of the 30th Turbomachinery Symposium, 139-175, Houston, TX.

Tukey, JW (1977). Exploratory Data Analysis. Addison-Wesley, Reading, MA.


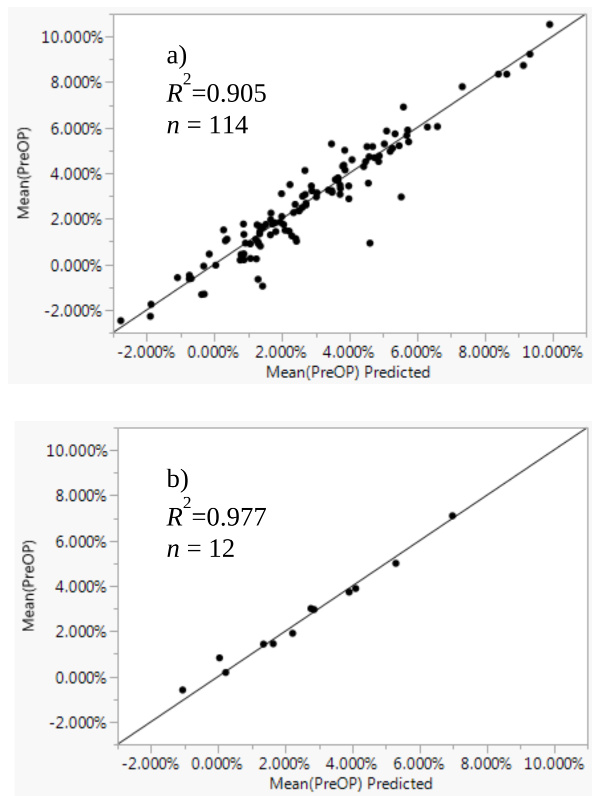


Figure 11. Actual degradation and predicted results: a) Model training, and b) model validation

## BIOGRAPHIES

**Xiaomo Jiang** received his Ph.D. in Structural Engineering from the Ohio State University in 2015 and M.Eng. from National university of Singapore in 2000. From 2005 to 2007, Dr. Jiang conducted post-doctoral research at Vanderbilt University. Currently Dr.

Jiang is a Technical Leader of Analytics in Monitoring & Diagnostic center at GE Power, where he is leading the advanced analytics development, integration, execution, and validation efforts for next generation Asset Performance Management platform, with applications in heavy duty power generation products such as gas turbines and steam turbine, as well as other OEM assets in coal and nuclear plants and liquidated natural gas manufacturers. Dr. Jiang is an expert at Bayesian statistics, wavelet signal processing, and intelligent algorithms. He has 20 years of experience in condition based diagnostics, predictive analytics, prognostics, risk and reliability modeling, performance monitoring, and model validation for engineering systems under uncertainties. Dr. Jiang has authored 1 book, 3 book chapters, and more than 40 research Journal papers and 25 international conference papers in the cross-disciplinary fields of computer science, engineering, and applied statistics. He holds multiple U.S. and foreign patents. He has been invited to serve as a peer reviewer, editorial board, or associate editor for over 40 different international journals and conferences.