Proceedings of ASME Turbo Expo 2011
GT2011
June 6-10, 2011, Vancouver, British Columbia, Canada

GT2011-45991

EARLY WARNING OF GAS TURBINE FAILURE BY NONLINEAR FEATURE EXTRACTION USING AN AUTO-ASSOCIATIVE NEURAL NETWORK APPROACH

[1]Thomas Palmé, [1]Peter Breuhaus, [1]Mohsen Assadi, [2]Albert Klein, [2]Minkyo Kim

[1]Department of Mechanical- and Structural Engineering and Materials Science
University of Stavanger
N-4036 Stavanger, Norway

[2]Alstom Power
Baden, Switzerland

ABSTRACT
This study investigates the application of nonlinear Principal Component Analysis (PCA), implemented through the use of Auto-Associative Neural Network (AANN), for early warning of impending gas turbine failure. The study is based on a real operational data set that includes a compressor failure. The analyzed data set consists of measured operational parameters whose identity are unknown, hence this study presents a purely data driven approach to the problem of early warning. In this case study, the use of AANNs for early detection of abnormal engine behavior could have provided the operator with a warning a few days prior to the fully developed failure, which resulted in a forced shut-down and extensive maintenance. Furthermore, a comparison is made between the nonlinear PCA by AANNs and the standard PCA model, which is an inherently linear method. The result shows that the AANN provides a more reliable detection of the failure by a higher residual generation during failure mode as well as fewer false indications prior to the failure. Consequently, this study shows that nonlinear PCA as performed with AANNs can be a valuable data driven tool for early warning of gas turbine failure.

Keywords: nonlinear principal component analysis, auto-associative neural networks, failure detection, industrial gas turbines

1 INTRODUCTION

Increasing competition in the electricity market forces operators to continuously improve the overall economics of their power plants, determined except for the efficiency by reliability and availability. Early detection of impending failure in gas turbines can improve plant availability since it provides the operator with the opportunity to take appropriate actions before any serious malfunctioning has occurred. Early warning detection ability should not be underestimated, since the economical consequences of a failure can be drastically reduced since additional time is given for plant operators to take appropriate maintenance actions.

Monitoring is usually performed by evaluation of the measured parameters at different locations along the gas path of gas turbines. Most of the parameters have a variation in the measured values caused by different operational conditions. In a gas turbine, this variation may be caused by different ambient conditions and load levels or operational modes such as frequency control where the Inlet Guide Vanes (IGV) is constantly varying. In condition monitoring, the main focus is to differentiate between normal variation and variation caused by failure, degradation or e.g. sensor faults. There are mainly two different approaches to approach this problem. One is to build a mathematical model of the system, and compare the actual readings against model predictions. The second one is to use operational data to build a so-called nonparametric model or data-driven model of the system. Regardless of the approach, a monitoring system should be easy and fast to develop and use as well as provide sufficient accuracy to detect impending failures. It should also be efficient, in the way that it should provide early detection, but at the same time minimizing the number of false alarms. Another important feature of monitoring systems is processing speed, which should permit on-line applications for real time equipment surveillance. In addition, a monitoring system is normally an ad hoc installation using the existing standard engine measurements already in place. Modern Monitoring & Diagnostics (M&D) is usually centralized for entire fleets e.g. in support centers and to efficiently inform experts and guide them to the critical areas, M&D systems need to provide a clear message on the status on the component of concern.

The decision between a mathematical modeling approach and a data driven modeling approach may depend on what kind of information is available for model development. A mathematical model, like a heat and mass balance program or something similar, requires detailed specifications, component characteristics and process knowledge of the GT system. Some of these required specifications and characteristics may not be easily obtainable for plant operators because of their Original Equipment Manufacturer (OEM) proprietary nature. One example is the compressor map, necessary to model a gas turbine at off-design operation. The well-known Gas Path Analysis methodology, which is used in conjunction with a Kalman filter [1], is normally accompanied with a normalization procedure, which requires the compressor map for single shaft gas turbines, as well as sensor noise levels for each sensor to be executed correctly. However, one resource that is available, normally in excess and not used to its full potential, is operational data, corresponding to normal fault-free operation. Since this is the only requirement in a data driven modeling approach, this alternative provides an attractive cost efficient approach. Different methodologies exist for nonparametric modeling, ranging from statistical methods such as principal component analysis (PCA) [2], similarity based modeling, which is a type of Kernel approach [3], neural network based models [4-5], and Auto-Associative Kernel Regression (AAKR) [6].

In this study, the model development is performed using an auto-associative neural network [7-8], which permits nonlinear feature extraction from high dimensional data. The models are developed from a data set collected during the normal operation of the gas turbine and thus represent a baseline model of the gas turbine. It is then applied on a data set containing a real failure, collected after the training data. This is actually a unique situation that allows verification of the model under real conditions. To verify the need of a nonlinear model a comparison is made to the well-known PCA, which performs linear feature extraction. The performance of the feature extraction models are highly dependent on the number of extracted features and a comparison is made between different models with different numbers of extracted features. The result shows that the AANN provides a lower training error for a certain number of extracted features as well as a higher residual prior to the fault, which verifies the nonlinear modeling approach. In addition, it is seen that the training error can be used as guideline for selection of model configuration, which is necessary when the model development is based on normal operational data.

2 CORRELATION MONITORING METHODOLOGIES

In this section, a general introduction to monitoring by nonparametric correlation modeling will be given. The methodology applied is as follows.

- A data set containing only healthy gas turbine data is used to produce a so-called baseline model. This baseline model incorporates the gas turbine characteristic in normal operating condition, i.e. without any fault.
- New data are compared with predicted data from this model, and a residual is calculated. By comparing the actual residual and the residual obtained in the training data set, a decision as to whether or not a problem exists can be made.

Figure 1 show the principle, where a new measurement vector is compared against the historical data. This comparison is done to see if a similar input as the one that is currently measured can be found in the history, which indicates that the current operation is in-line with normal operation.

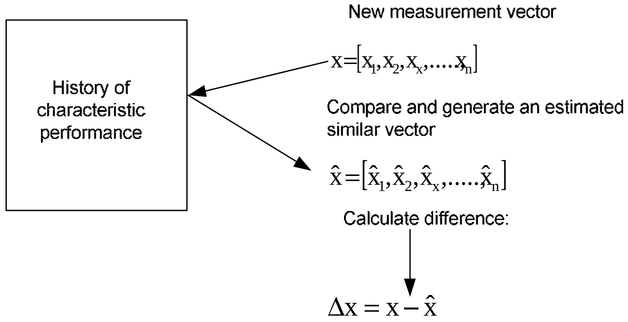

Figure 1 Principle for nonparametric correlation modeling

Considering on-line monitoring, this should be performed for each new reading in time applying the same data resolution as in the data used for model development. The actual calculations are for most data driven methods a fast process since the model parameters are set once the model is developed. The main idea behind evaluating all parameters together is that only certain combinations of parameter values correspond to normal operation due an interrelationship between the parameters. This interrelationship can be termed correlation or dependency between the parameters. Consider figure two, which illustrates linear and nonlinear dependency, or correlation, between two different parameters, *x* and *y*. Since they are correlated, it is only certain combinations of the parameters that belong to normal operation.

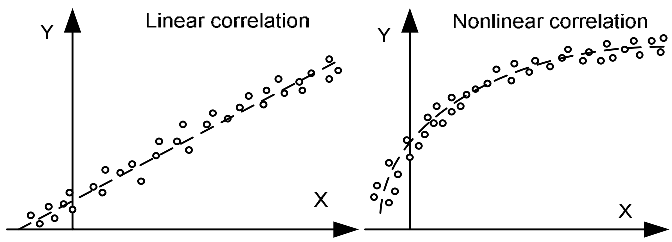

Figure 2 Illustration of linear and nonlinear parameter correlation

When all parameters are used as input parameters, the model is said to work in an auto-associative mode. This is in contrast to a dedicated input-output model, or hetereo-associative mode, where output parameters are predicted based on certain inputs. For some systems that are highly interconnected with other systems, as well as have a high degree of correlation between the monitored parameters, it might be more suitable to evaluate the correlation between the parameters, due to difficulty to selection of appropriate input parameters or the absence of parameters that define the operational condition.

Principal Component Analysis is the most common technique for dimensionality reduction. Given a set of data on

n dimensions, PCA aims to find a linear subspace of lower dimension such that the data points mainly lie on this linear subspace [2]. This subspace attempts maintain most of the variability of the data by applying the Mean Square Error (MSE) objective function. Data compression methods such as PCA have been used in process engineering for a long time, see for example ref [9-13]. PCA is however designed to model linear variabilities in high-dimensional data, which means linear correlations between the parameters. However, if the data are nonlinearly correlated, the high-dimensional data lies on or close to a nonlinear manifold and not a linear subspace. Then, PCA cannot model the data variation accurately which imply that nonlinear PCA are needed to model the data correctly. To clarify, nonlinear PCA is in this context data compression into nonlinear manifolds, and not a method. Nonlinear PCA can be performed with different methods, such as Kernel PCA [14] and principal curves [15] and [16] and some application is shown in e.g. [17]. However, in these cases the number of parameters is limited to few such as 3 in [17]. It might be stated that the high dimensional modeling in combination with nonlinearity provides a challenging case. One method that can cope with both nonlinearity as well as high dimensional systems is the AANN, which is based on a three hidden layer multi-layer perception. The AANN's main difference to PCA is that a nonlinear optimization problem has to be solved during the model building process since there is no closed form solution for calculation of the model parameters. However, the AANN provides a flexible data compression methodology, which means that arbitrary correlation can be learned without restriction to certain correlations. In addition, no a priori information about the correlations is required. This approach is more suitable for real world a system that in many cases includes a combination of linear and nonlinear correlations between parameters.

3 PRINCIPAL COMPONENT ANAYSIS AND AUTO-ASSOCIATIVE NEURAL NETWORKS

PCA [2] is a nonparametric method to extract relevant information from complex data sets. Assume a data matrix X has n number of observations and m number of parameters, or variables. PCA allows a linear mapping from $\mathfrak{R}^m$ to $\mathfrak{R}^p$ where $p < m$, i.e. data compression:

$$X = TP^T + e \qquad (1)$$

Where T is the score matrix with dimension $n \times p$, p is the number of principal components, $< m$, of X and P is the loading matrix with dimension of $m \times p$. The Euclidian norm, or the MSE, of the residuals matrix is minimized for the given number of principal components. If $P^T P = I$ the linear of PCA is given by (2):

$$T = XP \qquad (2)$$

Where X represent a row of X, a single data vector and T represents the corresponding row of T. The loadings P are the coefficients for the linear transformation, and essentially define the orientation of the principal components with respect to the original m-variables. The information lost in this mapping is assessed by the reconstruction back to $\mathfrak{R}^m$:

$$X' = TP^T \qquad (3)$$

Where X' is the reconstructed vector [2]. In PCA, any nonlinearity between the variables is lost through the compression steps. Nonlinear Principal Component Analysis (NLPCA) allows arbitrary nonlinear mapping from $\mathfrak{R}^m$ to $\mathfrak{R}^p$. Consider the mapping:

$$T = H(X) \qquad (4)$$

Where H is a nonlinear vector function of p individual nonlinear functions, $H = \{H_1, H_2, \ldots\ldots, H_p\}$:

$$T_i = H_i(X) \qquad (5)$$

Reconstruction of the original data is accomplished by a second nonlinear function, $G = \{G_1, G_2, \ldots\ldots, G_p\}$:

$$X_j' = G_j(P) \qquad (6)$$

It should be noted here that G is the inverse of H. Figure three shows a three hidden layer AANN, which implements nonlinear principal component analysis. The bottleneck layer represents the dimension of the number of principal components extracted. The AANN is trained in a so-called supervised mode, and the targets used to train the network are simply the input vectors themselves. The network attempts to map each input vector onto itself though the reduced dimension in the bottleneck layer. The numbers of principal components is determined by the size of the bottleneck layer, while the nonlinear complexity of H and G is determined by the size of the mapping layers. In addition to PCA, where only the number of extracted principal components has to be determined, the AANN has two independent variables, number of neurons in the bottleneck layer and in the mapping layers. Even though the name implies nonlinear principal component analysis, there are some differences worth mentioning. In a PCA model, the principal components are orthogonal, which is not a requirement for the three hidden layer AANN. The AANN tries to decode the information in such a way that it can be transferred through the bottleneck layer and still retain accurate reconstruction capability. It does not require the extracted features in the bottleneck to be orthogonal.

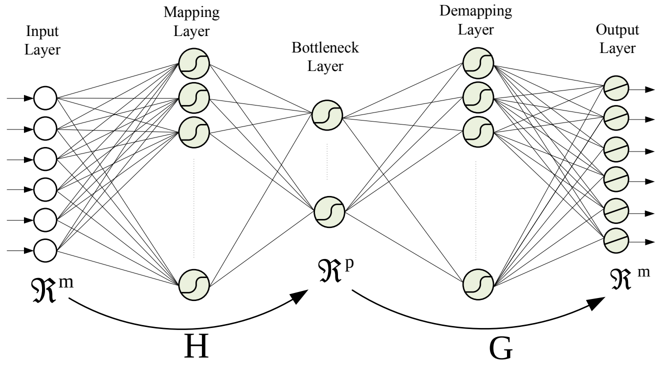

Figure 3 Nonlinear principal component analyses by AANN

The number of adjustable weights for an AANN depends on the number of parameters as well as on the size of the different layers. The number of adjustable parameters is:

$$w_{3H}(n_1, n_2, p) = 2 \cdot p \cdot n_1 + 2 \cdot n_1 \cdot n_2 + 2 \cdot n_1 + n_2 + p \quad (7)$$

Where p is number of parameter, n_1 is number of neurons in the mapping layers and n_2 is number of neurons in the bottleneck layer. To assure generalization, it is recommended to have more data patterns than weight values, even though no exact guidelines are available in this matter. This will avoid over fitting of the network, which means that, the AANN learns the specific training parameters. When the network is trained with a reduced dimension in the bottleneck layer, over fitting is less likely to happen since the correlation between the parameters has been implemented during training. In [18], Malthouse describes the limitations of AANN, which basically are that they cannot model functions that intersect themselves such as circles or parameterizations that have discontinuous jumps. However, it should be noted that additional hidden layers might weaken the argument made by Malthouse, as indicated in [19]. These conditions may or not present a practical problem, but they should be considered when setting up the model for a certain system. For example, if the operation modes are distinguished by two different conditions which cause a discontinuous jump in the data, it would be more feasible to make two models, one for each operational mode due to the limitations of the AANN. For more information and details on AANNs, the two articles by Kramer in 1991 and 1992 [7-8] are recommended, and [20] which give a detailed mathematical description of the data compression issue.

4 EXAMPLE: 3D – 2D – 3D

In this section an example of data compression with AANNs will be shown for an arbitrary nonlinear function, which illustrates the nonlinear PCA performed by AANNs. Consider a functional relationship, $y = f(x, y)$, which is a three dimensional function in the Cartesian coordinate system. Since this is a surface, the same information could be expressed in only two dimensions. We might think about this as stretching out the surface, see ref [21] and [22] for details on this matter. For simple geometries, such as a circle or a sphere, it is possible to use an analytical expression for transformation from the Cartesian coordinates to a polar or the spherical coordinate system. This can provide a dimensionality reduction. For this transformation, the issue is to perform an arbitrary coordinate change by data compression. Figure four shows a schematic illustration of data compression for surface.

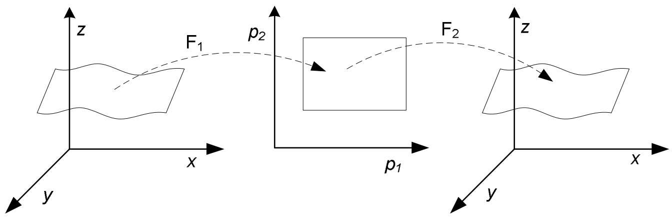

Figure 4 Nonlinear feature extraction and reconstruction of 3-dimensional surface

Consider the nonlinear function:

$$z = 1.3356(1.5(1-x) + e^{(2x-1)} \sin(3\pi\,(y-0.6)^2)..$$
$$..+e^{(2y-1)} \sin\big(4\pi(x-0.9)^2\big)\big) \quad (8)$$

The surface generated by formula (8) is shown in figure five.

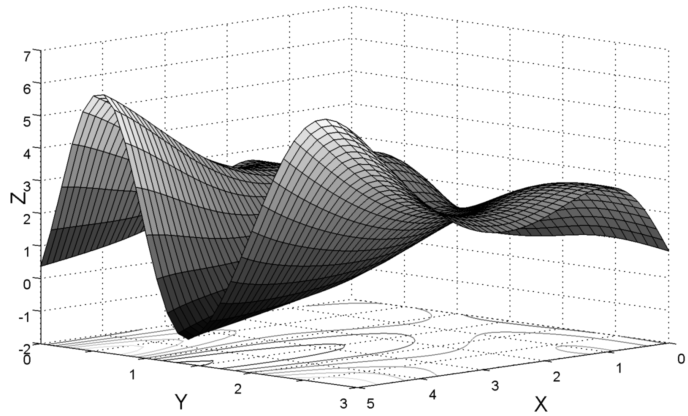

Figure 5 Simulated nonlinear function

This highly nonlinear function is arbitrarily selected; it could also be any nonlinear continuous function that does not intersect itself. However, since this is a surface, it should be possible to express the same information in only two dimensions. To verify this, several different AANNs with different number of neurons in the hidden layers is trained to reproduce this surface and the MSE for each network configuration is shown in figure six.

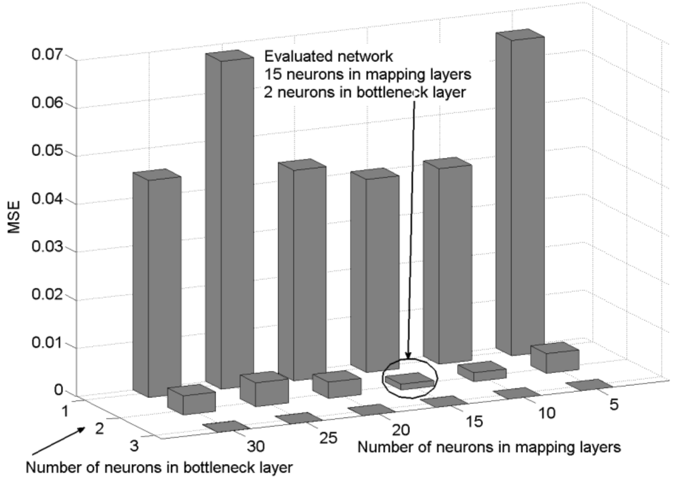

Figure 6 Training error as a function of neurons in hidden layers

In figure six it can be recognized that two neurons are enough to provide almost a perfect reconstruction of the surface. Three neurons in the bottleneck layer perform a perfect reconstruction, but in this case no dimensionality reduction is performed. With one neuron in the bottleneck layer the data reconstruction error is substantially higher since the information in the function (8) cannot be explained in one dimension. A PCA with two principal components would create a plane on the function, placed in such a manner that the location would minimize the variance in a MSE fashion. Figure seven shows the original surface as well as the network-reproduced surface for the network with 15 neurons in the mapping layer and two in the bottleneck layer.

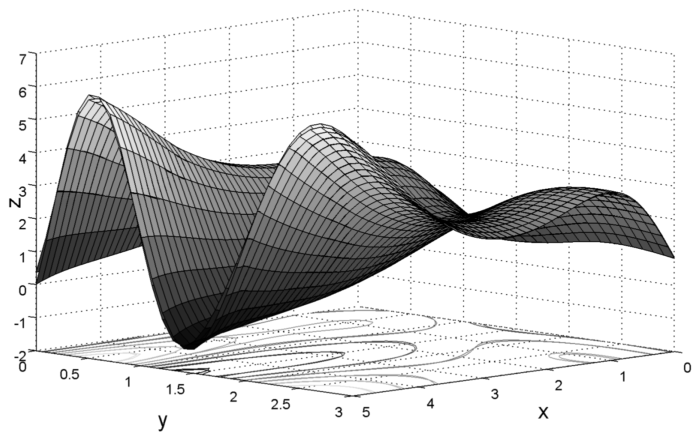

Figure 7 Simulated surface and network reproduced surface on top of each other

Since the AANN is constructed with two neurons in the bottleneck layer, the network has decoded the information into only two dimensions, thus it has been forced to perform nonlinear principal component analysis on the modelled function.

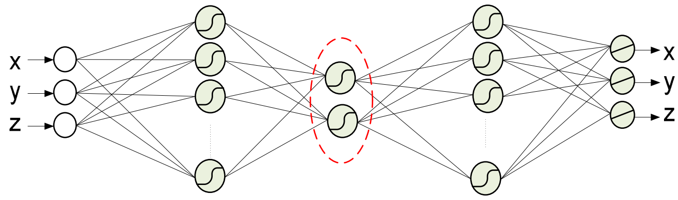

Figure 8 Schematic illustration of AANN as a feature method from 3D to 2D and back to 3D

One may ask why this should be useful. Pretend that the generated function (8) represents the healthy characteristic behavior of a specific component, and all three variables, x, y, z, are measured parameters. The network would only provide a good match between input and output vectors when presented with input vectors that are located somewhere on the surface/function. If the characteristics behavior, i.e. the imbedded function (8) would change and generate values of x,y,z which does not belong to the surface, a residual between the input and output values would be seen. Thus, the dependency between the parameters at each operational point is considered. It is not required that the parameters are correlated as long as there is a functional dependency between them; however, a high correlation normally implies a higher data compression capability.

5 CASE STUDY

A data set containing a real gas turbine failure was used for this study. It consists of 25 measured parameters along the gas path of the specific gas turbine. This specific industrial gas turbine was operated in a so-called frequency mode, which means that the power output is constantly controlled by varying the Inlet Guide Vanes (IGV) angles in order to support the grid frequency. As mentioned previously, the specific gas turbine as well as the identity of the parameters is unknown. The data was divided into two sets, one set called training data and a second called failure data. In the training data, operational data from normal engine operation is included which are used to develop the baseline models. The failure data consists of data collected after the training data time window, and includes the real failure event in the end. The failure that occurred was a compressor blade failure, which caused an automatic shut down of the gas turbine. Unfortunately, any additional information about the failure is not available. The training data set consist of 6500 data patterns, while the failure data consist of 5262 data patterns. Each data pattern consists of a 5-minute average value from each specific sensor, which corresponds to normal data resolution used for gas turbine monitoring and data archiving. In figure nine a schematic illustration of the case study is shown.

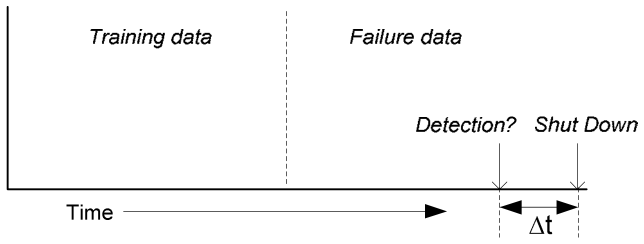

Figure 9 Schematic illustration of the case study

6 MODEL DEVELOPMENTS – CONFIGURATION AND TRAINING

In this section, the process of building a model based on the training data set will be described. Since the knowledge of the data set is limited, a systematic trial and error approach is adopted to decide the correct model configuration. The first step is to train a model of the gas turbine by using the training data set. Since the actual parameters are unknown, no assumption about which parameters to include can be made, and therefore all parameters are selected. A three hidden layer AANN has two independent parameters, which are the number of neurons in the bottleneck layer and the number of neurons in the mapping layers. The goal is to compress the data as much as possible, while at the same time preserving the data reconstruction capability. This lowest possible dimension at which this is possible is called the intrinsic data dimension. Due to the nonlinear optimization problem during model development, the optimization algorithm can be trapped in a minimum higher than provided by the network complexity. Because of this, at least two different networks for each configuration are developed. However, assuming that each network converges to the global minima, the network reconstruction error can be specified as:

$$MSE = f(\boldsymbol{x}, \boldsymbol{y}), \quad y < x \tag{9}$$

Where **x** is number of neurons in the function approximation layers and **y** is the number of neurons in the bottleneck layer. The neurons in the bottleneck layer represent the number of extracted nonlinear features. Fewer neurons in the bottleneck layer imply higher data compression, since the information in the measured parameters is explained in a lower dimension provided that the reconstruction error does not increase. Each parameter is individually linearly rescaled, before presented to the network to give theme equal importance and to fit transfer functions in the network. During training, the training data is divided into three different sets called train, cross validation (CV) and test. The train data set is used to numerically adjust the network weights, while the CV data set is used as a benchmarking set to ensure network generalization. The test is used to validate the final network. The performance should be similar in all three sets. For training of the AANN models, the Scaled Conjugate Gradient (SCG) algorithm by Møller [23] has been used as the training algorithm. It is fast, and works well with large data sets and requires no user-defined parameters in contrast to the common Gradient Descent (GD) based optimization algorithm.

In the normalized data space, the difference between each parameters measured value and reconstructed value is defined as:

$$A_i = |x_i^{norm} - \hat{x}_i^{norm}| \tag{10}$$

While total network error, the residual, between the network input and output is defined as:

$$A_{tot} = \sum_{i=1}^{i=p} A_i \tag{11}$$

In contrast to PCA, the principal components in the AANN are not required to be orthogonal. In fact, in the training phase of the AANN, the network is trained to reproduce the input data vector as closely as possible at the output layer, given the constraints implemented by the number of neurons in the bottleneck layer. Thus, the network could group some highly correlated parameters through some of the bottleneck layer neurons and pass some other parameter or parameters through the bottleneck layer without correlation to the other parameters. Because of this, a validation test has to be performed in order to assure that the parameters are correlated to each other. This is done by contaminating each input parameter at the time in the training data set with faulty values, according to:

$$x_{i-fail} = x_i \pm \Delta_i \tag{10}$$

Where Δ_i is individually calculated for each parameter according to:

$$\Delta_i = \frac{\max(x_i) - \min(x_i)}{10} \tag{11}$$

Where 10 is arbitrarily selected. In literature the standard deviation of the measured values are commonly used. However, this is not suitable for non-Gaussian distributed parameters and therefore we divide with 10, just to make sure that the faulty values are within the operational range for each parameter. This is performed for all patterns in the test data and then an average of the network residual is calculated. A parameter which does not produces any residual when contaminated with these faulty values should be removed, since it is not correlated to the other parameters; it is just transferred through the network. Using this methodology, the correlation between the parameters can be validated. Of course, a standard linear correlation coefficient can be calculated, but this indicates only linear correlation and can be decisive when a parameter is nonlinearly correlated to the other parameters.

In the PCA model, the data are first normalized into a linear scale, the same with the AANN model, and then rescaled according to each parameters standard deviation. This implies that the data are centered on zero with a standard of one, which is the normal PCA procedure. Thus, by monitoring the residual in the normalized linear data scale, a comparison between the AANN and PCA models with the same number of extracted components as well as the same residual scale can be performed.

6.1 MODEL RESULTS

Figure 10 shows the data reconstruction error for different configurations that have been trained using the training data set. As seen, the main dominating parameter is the bottleneck dimension.

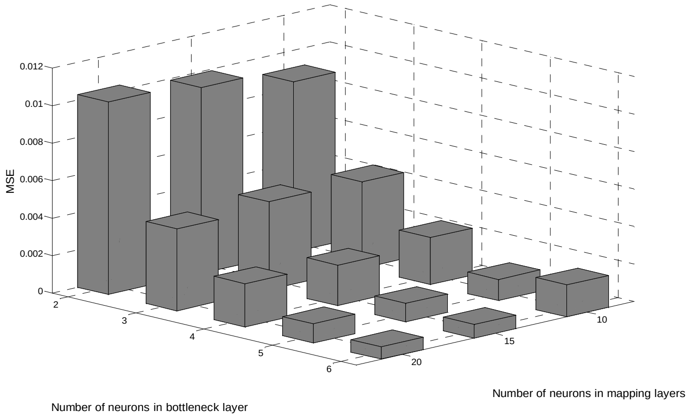

Figure 10 MSE for different network configurations, first run

The network with four neurons in the bottleneck layer and 20 in the mapping layers is selected and the parameter correlations are evaluated. One parameter, number nine, was seen to be uncorrelated to the other parameters since any failure in this parameter was simply transferred through the network. This was validated by examining different levels of failure in the parameter as well testing of different configurations of AANNs.

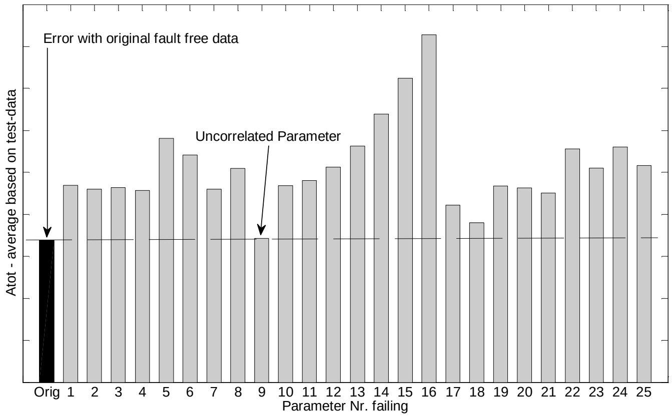

Figure 11 Error A_{tot} for failing input parameters

Figure 11 shows the error A_{tot} when the different parameters were contaminated by failing values according to (10). As seen, parameter nine is transferred through the model unchanged which means that one neuron in the bottleneck layer is dedicated to pass this parameter through the network. In figure 12 the error A_i is shown for each parameter when parameter one is contaminated with failing values and figure 13 show the result for parameter nine, clarifying the result in figure 11. This approach can also be performed for different Δ to specify the exact sensitivity to each parameter.

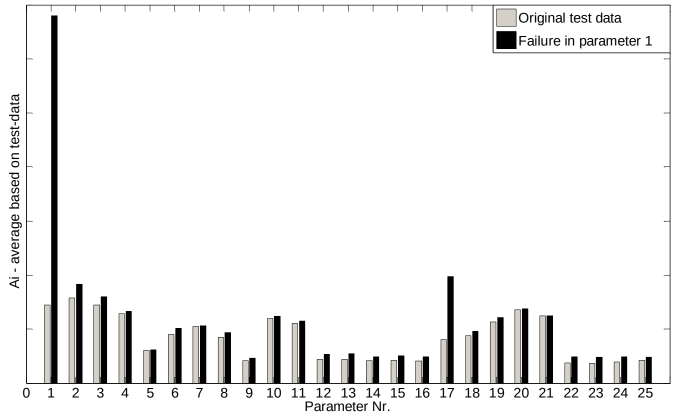

Figure 12 A_i for each parameter when parameter one is failing

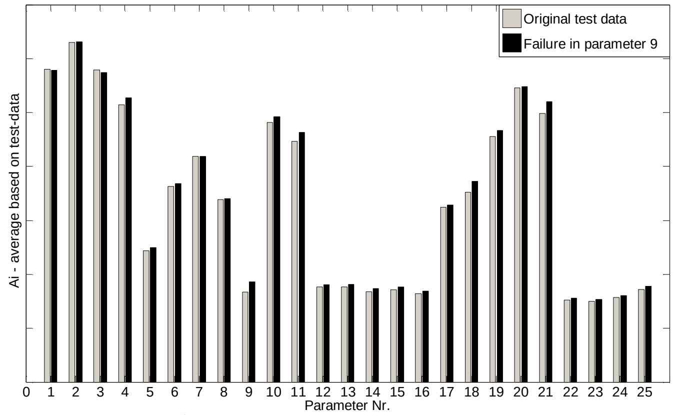

Figure 13 A_i for each parameter when parameter nine is failing

Since parameter nine does not contribute to residual generation during failure, this parameter can be removed. A new training variation is performed and in figure 14 the new training performance is shown. When the bottleneck layer is decreased from six to three neurons, the effect in the reconstruction capability is rather small and the big difference occurs when going from three to two neurons. Hence, it might be possible to assume that the intrinsic dimension, the lowest possible dimension to preserve the information, is a three dimensional space.

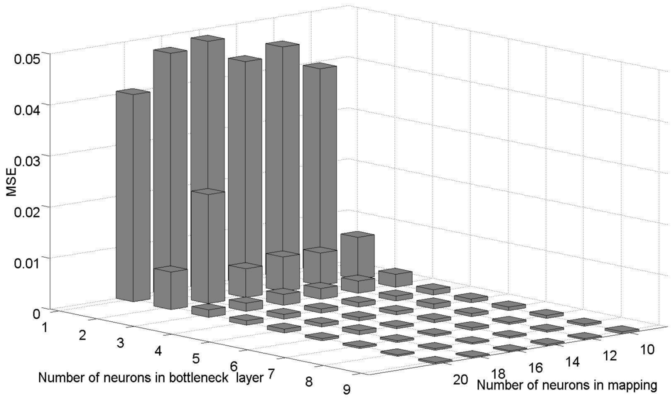

Figure 14 MSE for different network configurations, parameter nine omitted, second run

The network with three neurons in the bottleneck layer and 20 neurons in the mapping layer is selected for parameter correlation validation. Figure 15 shows the result and is clearly indicated that a failure in any of the parameters causes a residual and the dependency between the parameters is verified.

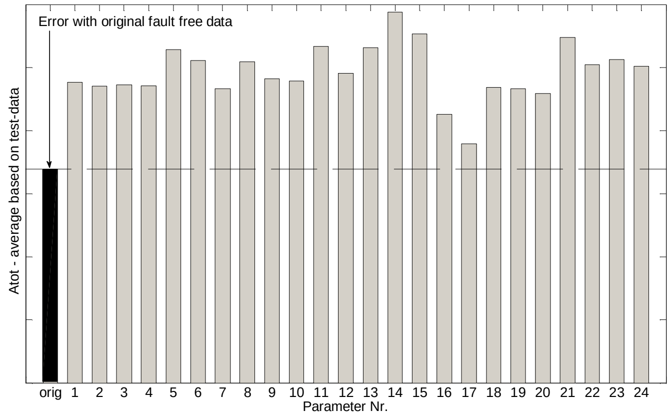

Figure 15 Error A_{tot} for failing input parameters, second run

To show the data reconstruction capability when applying three neurons in the bottleneck layer, figure 16 shows the result for an arbitrary parameter and for a certain time interval. The measured data values are rather similar to the network reconstructed values, even though this 24 dimensional data set has been compressed to and decompressed from a three dimensional space which indicates a high correlation between the parameters.

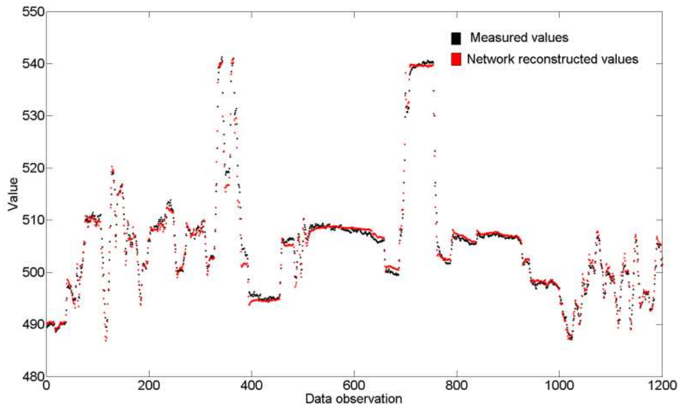

Figure 16 Measured values and network-reconstructed values for parameter 19

Since the performance of the AANN models will be compared against PCA models, figure 17 show the MSE as a function of used principal components applying the PCA procedure which indicates that five or six principal components are required to capture the variance in the data.

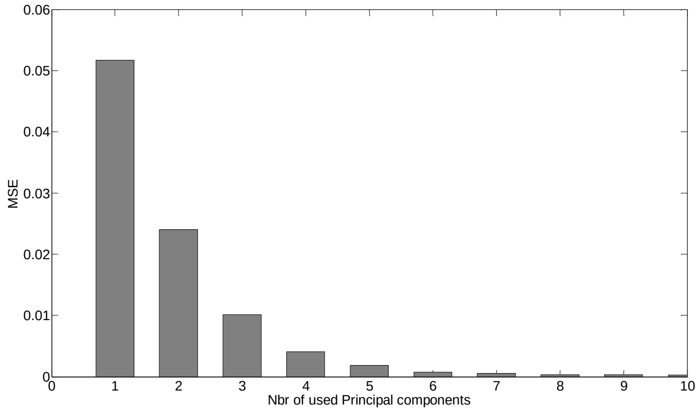

Figure 17 MSE as a function of principal components, PCA

To evaluate the different models performance on the failure data, four different AANN models and four different PCA models respectively, the AANN models contain one to four neurons in the bottleneck layer and the PCA models one to four principal components. One important issue is, is it possible to decide the optimal data compression level, or the intrinsic dimension, based on the training data?

Figure 18 show A_{tot} for the different AANN models. The first observation is that when two neurons are used in the bottleneck layer the residual is substantially higher than for the other models. Furthermore, increasing the number of neurons from three does not substantially improve the data reconstruction error.

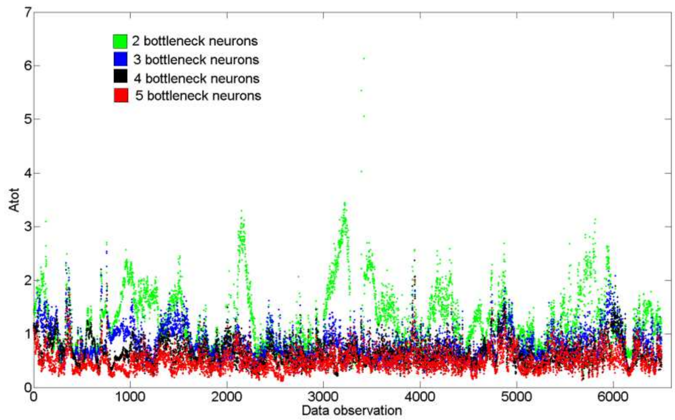

Figure 18 A_{tot} for the AANN models with different number of bottleneck neurons, training data

To illustrate the different residuals for the different configurations, figure 19 show the residual histogram for each configuration. The result shows that the measured parameters can be explained in three dimensions and additional dimension does not improve the result. Thus, it can be concluded that the intrinsic dimension for this data set, or system, is three dimensions.

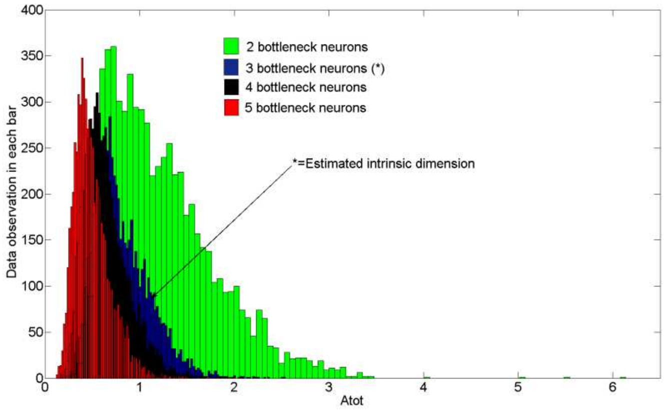

Figure 19 Histogram of A_{tot} for the different numbers of bottleneck neurons, training data

The same analysis is performed for the PCA models, and figure 20 show the result for the different PCA models. Compared to the AANN models, the PCA models produce a substantially higher residual with two and three principal components while the residual for four and five principal components are substantially lower than the previous.

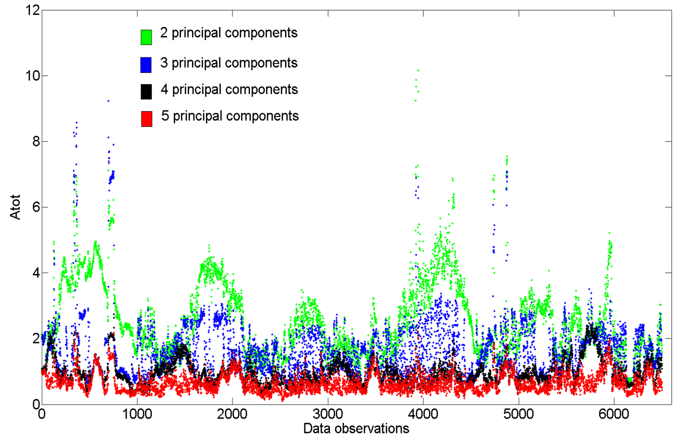

Figure 20 A_{tot} for the PCA models with different number of principal components, training data

From figure 21 the histogram of the different PCA models are shown, compared to the AANN models the intrinsic dimension seem to be four, since the difference between five and four principal components are rather low while the change between four and three principal components is substantial. Already here the nonlinear correlation between the parameter is verified, since the AANN model could process the data through a three dimensional space with higher accuracy.

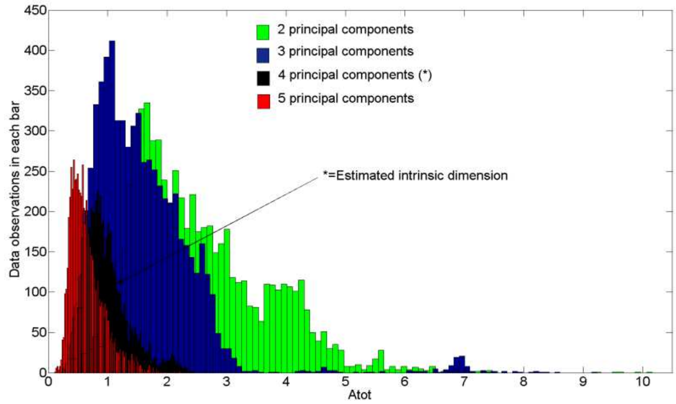

Figure 21 Histogram of A_{tot} for the different numbers of principal components, training data

7 EVALUATIONS WITH FAILURE DATA

In this section, the residual generation on the failure data is investigated. The important question is: how early could the operator have been notified so that corrective measures could have been taken? In addition, how does the residual generation depends on the model configuration and is the assumption of the intrinsic dimension based on the training data the most optimal considering high residual generation during failure mode?

The last hours of operation data in the failure data set is removed, since these data observations produces such high residuals that a detection could have been done by a simple max/min threshold on each parameter. A_{tot} is used as the monitored parameter, but it could also be possible to monitor A_i. However, A_{tot} provides one simple measure to monitor and includes the total residuals. In addition, more complex algorithms are available for residual evaluation such as consistency check, CUSUM etc., but the main point is that the model should produce a low residual when there is no fault and a high residual during failure mode. For this reason, the model residual is directly plotted versus time. It should also be mentioned that the authors does not know exactly when the fault occurred and hence the analysis is done based on the observed residual. This would simulate a real case where a model is trained on normal operational and then applied for on-line monitoring.

From figure 22, showing the residual generation for the different AANN models, it can be recognized that the gas turbine operation seems to be in-line with the training data performance until around data observation 4000. At this point, an increase in the error can identified which indicate an abnormal operation. However, shortly after, the residual decrease and is further on followed by sharp jump where the residual increases for all investigated models. An explanation can be, the failure occurred at the first indication of change in operation but the operational condition changed, or the control system intervened, and thereby the effect of the failure was less indicative on the performance.

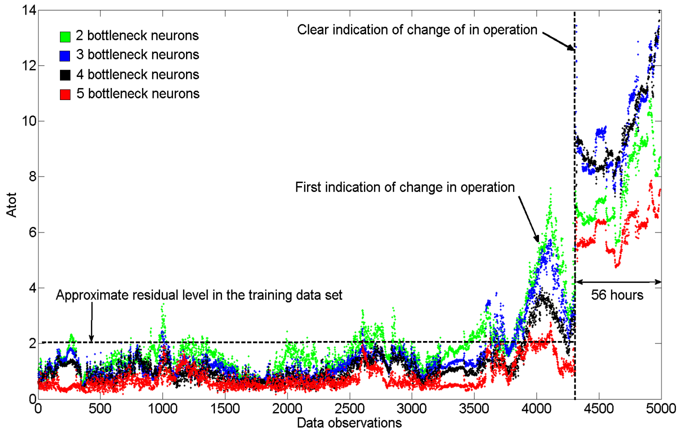

Figure 22 A_{tot}, failure data set, AANN models

The same analysis is performed for the PCA models, and visualized in figure 23. In this case, the PCA models with two and three principal components reveal a rather scattered residual prior to the assumed failure, while the PCA models with four and five principal components produces a rather low and constant residual. The sharp jump in A_{tot} is detected for all PCA as well, while the first indication of a change in the performance as seen with the AANN models are not visible (around data observation 4000). The pikes in A_{tot} produced by the PCA models with two and three principal components can be attributed to modeling error, since the same behavior was seen in the training data for these two models.

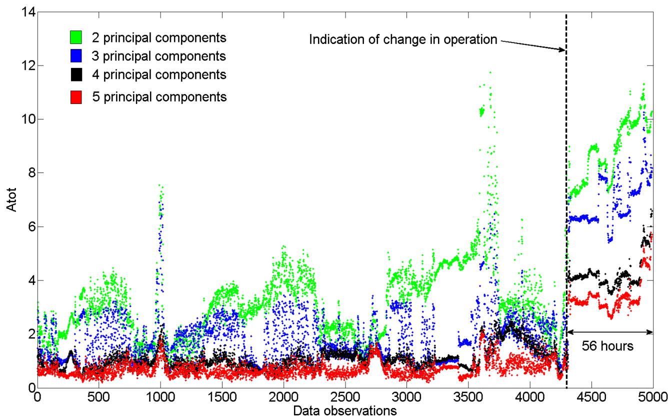

Figure 23 A_{tot}, failure data set, PCA models

The time period in which a clear change in the operation has been identified for both the AANNs as well as the PCA models are further studied. At this point it can be assumed that the failure has occurred and is expected to progress, i.e. it should produce an increasing residual. The AANN models with three and four number of neurons in the bottleneck layer produces a similar, increasing, residual after the failure has been detected, see figure 24. In addition, the residual generation is also higher for these models compared the model with two bottleneck neurons as well as the model with five bottleneck neurons. Two conclusions can be drawn at this point: selection of the appropriate bottleneck layer size produces a higher residual during failure and this can be determined from the training data, (as in figure 19).

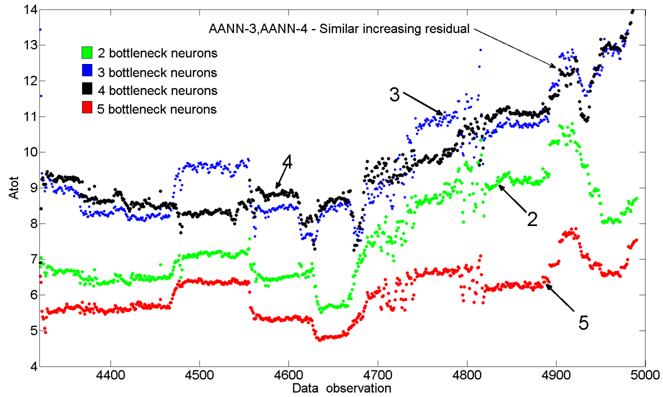

Figure 24 Close up of failure data set, AANN models, between data observations 4320 and 5000

The result, see figure 25, for the PCA models is rather different, in two aspects. First, the PCA model, which was assumed to contain the intrinsic dimension, four principal components, does not cause the highest error; instead this is seen for the PCA models with two and three principal components. Secondly, the residual generation is lower and does not reveal the same progressive trend as the AANN models. It could also be argued that the PCA models with four and five principal components produce a higher residual than in the training data set and is of course also valid as monitoring models. However, the AANN seem to be more sensitive to change in the performance through a higher residual generation and thereby provides a higher confidence in the results.

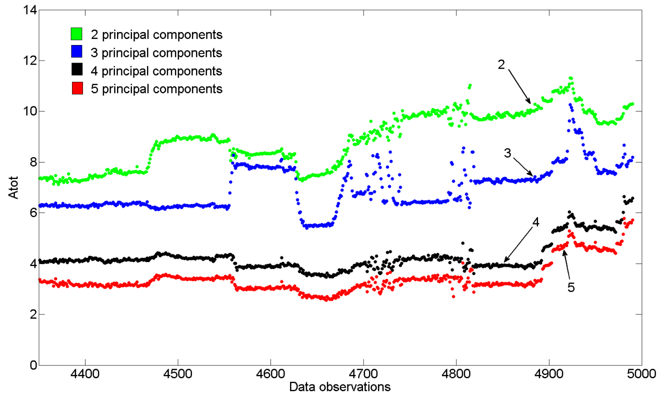

Figure 25 Close up of failure data set, PCA models, between data observations 4320 and 5000

8 DISCUSSIONS AND CONCLUSIONS

The aim of this study was to investigate the usefulness of AANNs for early warnings of gas turbine failure by imbed the characteristic operational performance in a feature extraction model which can be used to validate new observed data. When the interrelationships between the parameter changes due to failure, the model will produce a difference between the measured data and the model estimate, hence this difference can be viewed upon as a residual used to detect changes in the operation related to abnormal data readings.

One may ask what kind of failures that a feature extraction model can detect; the answer is that any failures that change the interrelationships between the parameters will be detected. One of the main tasks during the model development phase is to decide upon the model configuration, which in the case of AANNs requires selection of two variables, number of neurons in the bottleneck layer and number of neurons in function approximation layers. The bottleneck layer is the most important since it determines the number of extracted features, or the dimension in which the measured parameters can be explained without information loss. Number of neurons in the mapping layers should be selected such as the transformation from the measured dimension (i.e. number of parameters) to the compressed dimension can be performed satisfactory. When the so called intrinsic dimension (the smallest dimension that can explain the same information as in the measured data dimension) is not known a priori, this can be determined from a trial and error approach where the smallest possible dimension should be selected. In this study, different AANN with different number of bottleneck neurons was selected to evaluate the residual during failure and it was seen that a correct selection of number of neurons in the bottleneck layer produce a more sensitive model which respond with a higher residual to failure. In this study a real compressor failure was investigated and it was seen that the most appropriate feature extraction could be selected by evaluating A_{tot}, and specifically the difference in A_{tot} for different configurations.

The main purpose to apply an AANN is the assumption of nonlinear correlation between the parameters, since a PCA model can applied for linear correlations, thus the improvement provided by an AANN depends on the nonlinearity within the system. An AANN requires a higher computational effort to develop since the model training requires a nonlinear optimization procedure. One AANN model takes approximately one hour to develop while a PCA model for the same data set may take in the range of one minute. In many cases, a system can be assumed to behave linearly at a certain operational point but if a full operational envelope is considered, it most often behaves nonlinear. Thus, an advantage provided by the AANN is that one model can incorporate the full operational behavior.

In this case applying a gas turbine data set, a comparison between the PCA and AANN was performed for different configuration, i.e. number of extracted features. The AANN model could express the data set by approximately three extracted featured, while the PCA with three principal components resulted in a higher data reconstruction error. Thus, this verifies that the PCA assumptions fail and that the AANN performs a nonlinear PCA on the training data set. The PCA model with three principal components showed a few pikes in training data set where the error was substantially higher than the average, which can be explained by operational conditions that imply a nonlinear behavior. All models identified a time 56 hours prior to the fault, where the residual indicated that a shift in the performance occurred. However, the residual by the AANN model were both higher as well for the AANN model 3 and 4 number of extracted features indicated a progressive residual which is in line with what is expected. The AANN models with 2 and 5 numbers of extracted features revealed a lower residual, thus this indicated that a correct selection of the bottleneck layer is required for optimal performance.

In summary, this study shows that AANNs can applied as valuable tool for early warning, or novelty detection, of abnormal operation when correctly configured. Compared to PCA, it permits nonlinear system to be modelled and thereby provides a wider applicability to different power plant systems.

ACKNOWLEDGMENT
ALSTOM Power is gratefully acknowledged for financial support which made this study possible.

NOMENCLATURE

AAKR	Auto-Associative Kernel Regression
AANN	Auto-Associative Neural Network
CV	Cross Validation
GD	Gradient Descent
IGV	Inlet Guide Vanes
M&D	Monitoring and Diagnostic
MLP	Multi-Layer Perceptron
MSE	Mean Square Error
OEM	Original Equipment Manufacturer
PCA	Principal Component Analysis
SCG	Scaled Conjugate Gradient

REFERENCES

[1] Urban, L. A., Volponi, A.J “Mathematical Methods of Relative Engine Performance Diagnostics” Hamilton standard Division of United Technologies Corporation

[2] H. Hotelling. Analysis of a complex of statistical variables into components. J. of Educational Psychology, 24:417–441, 1933.

[3] Herzog, J. P., Hanlin, J., Wegerich S.W., Wilks, A. D., “High Performance Condition Monitoring of Aircraft engines”, ASME Turbo Expo 2005, GT2005-68845

[4] Singer, I. R. M., Gross, K. C., King, R. W., Wegerich, S., ”A pattern-Recognition-Based, Fault-Tolerant Monitoring and Diagnostic System”, Prceedings of 7th Symposium on Nuclear Reactor Surveillance and Diagnostics, Vol.2, Avignon, France 18-23 June 1995

[5] Radhakrishnan, V. R, Mohamad, A. R. (2000) Neural networks for the identification and control of blat furnace hot metal quality. Journal of Process Control 10 (6), 509-524

[6] Wand and Jones (1995) Kernel Smoothing, Monographs on Statistics and Applied Probability, 359-372 (1964)

[7] Kramer, M.A., (1991) “Nonlinear Principal Component Analysis Using Auto associative Neural Networks”, Laboratory for Intelligent Systems in Process Engineering, Dept. of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

[8] AUTOASSOCIATIVE NEURAL NETWORKS, M. A. Kramer, Computers chem. Eng, Vol 16, No. 4 pp. 313-328, 1992

[9] J.E, Jackson “Quality control methods for several realated variables” Industrial Quality Control, 7:2-6, 1956

[10] J.E Jackson, “Quality control methods for several related variables, Technometrics, 1:359-377, 1959

[11] K. A. Kosanovich, M. J. Piovoso, K. S. Dahl, J. F. MacGregor, P. Nomikos “Mulit-way PCA applied to an industrial batch process” In Proceedings of the American Control Conf, pages 1294-1298, Piscataway, New Jersey, 1994, IEEE Press

[12] M. J. Piovoso, K. A. Kosanovich, R. K. Pearson. “Monitoring process performance in real time” In Proceedings of the American Control Conf, pages 2359-2363, Piscatawy, New Jersey, 1992, IEEE Press

[13] B. M. Wise, N. B., Gallagher, “The process econometrics approach to process monitoring and fault detection” Journal of Process Control, 6:329-348,1996

[14] Schölkopf, B., Smola, A. J., & Muller, K. (1998) Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10(5) 1299-1399

[15] Hastie, T. J. & Steutzle, W (1989) “Principal curves”, Journal of American Statistical Association, 84,502-516

[16] Dong, D & McAvoy, T. J., (1996) “Nonlinear principal component analysis – based on principal curves and neural networks”, Computers and Chemical Engineering, 20, 65-78

[17] Cui, P, Li, J., Wang, G., “Improved kernel principal component analysis for fault detection”, Expert Systems with Applications 34 (2008) 1210-1219

[18] Malthouse, E.C., “Limitations of Nonlinear PCA as Performed with Generic Neural Networks”, IEEE Transactions on Neural Networks, Vol. 9, No. 1, January 1998

[19] Ikbal, M.S, Hermant, M., Yegonanrayna, B., “Analysis of auto associative mapping neural networks”, International Joint Coference on Neural Networks, (ICJNN’99); Wahsington DC;USA;10 July-16 July 1999, pp. 3025-3029,1999

[20] Marseguerra, M., Zoia, A.,”The auto associative neural network in signal analysis: I The data dimensionality reduction and its geometric interpretation” Annals of Nuclear Energy 32 (2005) pp. 1191-1206

[21] Hermann, R., 1973. Geometry, Physics and systems. Marcel Dekker, Inc., New York

[22] Sastry, S., 1999. Nonlinear Systems, Analysis, Stability and Control. Springer, Berlin

[23] Møller, M. F., 1993, "A scaled conjugate gradient algorithm for fast supervised learning". Neural Networks, Vol.6, pp.525-533.