Effects of Sampling Rate on the Accuracy of the Gas Turbine Performance Deterioration Modeling

Houman Hanachi
Queen's University
houman.hanachi@queensu.ca

Jie Liu
Carleton University
jliu@mae.carleton.ca

Avisekh Banerjee
Life Prediction Technologies Inc.
banerjeea@lifepredictiontech.com

Ying Chen
Life Prediction Technologies Inc.
cheny@lifepredictiontech.com

Abstract - Logging the operating parameters of a gas turbine engine (GTE) is essential for the real-time performance monitoring and future health state prediction. Sampling rate of the operating data is restricted to the available resources, especially when they are logged manually by the operators. In some recent research works, the authors have introduced a physics-based approach to develop unified indicators for gas turbine performance monitoring. "Heat Loss index" and "Power Deficit index" were two of the indicators introduced to provide metrics for the health state of the gas turbines using the logged data from the engine control system. In this paper, prediction models are established on the time series of the indicators, and the modeling accuracy is investigated by considering with the sampling rate and the time window within which the model is established. As a result, this paper provides an insight into the uncertainty of the performance prediction model, with respect to the sampling rate and the length of the time window.

Keywords - gas turbine health monitoring; performance deterioration; modeling uncertainty; sampling rate; sampling decimation.

I. INTRODUCTION

Modern approaches for health management of the gas turbine engine (GTE) tend to account for the real-time health condition of the machine for maintenance decisions, which is called condition based maintenance (CBM). Such strategies call for unscheduled maintenance actions for imminent failures, and prevent replacing the healthy parts based only on prescheduled maintenance tables. In this way CBM improves the reliability and availability of the GTE, and reduces the maintenance costs using the actual health information from diagnostic and prognostic analyses [1].

Prognosis of the future performance deterioration of the GTE, variation of the GTE performance needs to be investigated by trending, for which different statistical and inference techniques can be used. Residuals are generated, which are the differences between the measurements and the predicted value by the model. A well fitted modeling curve can be achieved by minimizing a length of the residuals [2]. For a prediction model fitted on numerical data, the rate of data sampling is a key factor for the modeling accuracy, which needs to be carefully investigated. In the case of the GTE performance, the short-term performance is expected to deteriorate steadily, which implies that the performance prediction model has a non-periodic nature. As a result, the minimum sampling frequency cannot be defined by the Nyquist-Shannon sampling theorem [3].

In recent research work [4, 5], physics-based approach has been taken by the authors to develop performance indicators to provide metrics for the health level of the single shaft GTEs for diagnostics. Two of the introduced performance indicators are "heat loss index" (*HL*) and "power deficit index" (*PD*), which showed a promising robustness to the variations of the operating conditions. *HL* is defined the normalized difference between the measured exhaust gas temperature (*EGT*) and the ideal *EGT*, i.e., the *EGT* calculated given the measured operating conditions. *PD* is defined the normalized difference between the ideal power and the measured power of the GTE. Both the indicators are calculated using the measurements obtained from the GTE control system. Therefore, these indicators are time series signals with the same time index of the measurements. The variation trends of the indicators are then modeled by polynomial curve fitting [6], which is well accepted for performance deterioration modeling [7].

When a model is established using the sampled data, the prediction accuracy of the model will depend on the sampling rate. Given a particular statistical characteristic for the introduced performance indicators, the corresponding sampling rate should be investigated in regards to the required accuracy of the performance prediction. The dependency of the modeling accuracy to the sampling rate of the indicators is further investigated in this research. The study is intended to figure out the required data logging frequency from the GTE

control system. The results will help the GTE operators to optimize the logging periods as well as help in subsequent performance monitoring. The results can also be utilized for state estimation frameworks using sequential hybrid filters, where the modeling uncertainty is a necessary parameter.

II. UNIFIED PERFORMANCE INDICATORS

The control system of the GTE provides measurements on the gas path parameters of the machine, which are dependent on the GTE operating conditions. Therefore, there is no single parameter which can individually provide a measure on the GTE performance. Having a unified performance indicator that provides the health level of the machine in a single scalar can be much beneficial. Given the measurements available from the GTE control system, dependency of the measured parameters can be modeled by thermodynamic modeling of the GTE gas path, what can provide an insight into the health status of the GTE.

A. Available Operating Data

The formulation of the GTE thermodynamic model depends on the available turbine design data, as well as input and output parameters. The gas turbine in this study is a single shaft 5 MW GTE, which was in service in a local power plant for 38 months. The basic technical data of the GTE are provided in Table I. The operating parameters were measured once every two hours during the entire operating period. There are 19 readings from the cycle parameters, which can be used for thermodynamic modeling. For the same period, the ambient conditions like air pressure and relative humidity of air were also acquired from the historical weather records for the plant location. After performing sensitivity analysis for developing GTE performance model, three ambient conditions, i.e., inlet air temperature, pressure and relative humidity (T_i, P_i and ϕ_i), and three operating parameters, i.e., shaft speed (N), EGT and power (PW) were selected. The Fig. 1 shows variations of the operating data during the operating period. The gaps in the plots correspond to the GTE down time. It can be concluded from the plots that the health state of the GTE cannot be readily acquired from an individual parameter.

B. Introducing the Performance Indicators

The only available measurements to define the performance indicators were presented in the previous section. In a recent research work, a comprehensive nonlinear model for the single shaft GTEs using humid air as the working fluid was constructed [8]. The model combines the submodules of the compressor, combustion chamber(s) and the turbine sections, using heat balance and mass conservation principles. The mathematical forms of the model to predict power and EGT are:

$$PW_M = PW_M(T_i, P_i, \omega_i, N, f)\,, \tag{1}$$

$$EGT_M = EGT_M(T_i, P_i, \omega_i, N, f)\,, \tag{2}$$

TABLE I. BASIC TECHNICAL SPECIFICATIONS OF THE GTE

Parameter[a]	Value
Nominal power (PW_D)	5 (MW)
Shaft speed (N_D)	16500 (RPM)
Thermal efficiency (η_D)	30.2 (%)
Pressure ratio (PR_D)	14.0
Exhaust gas temperature (EGT_D)	785 (K)

a: On design data points

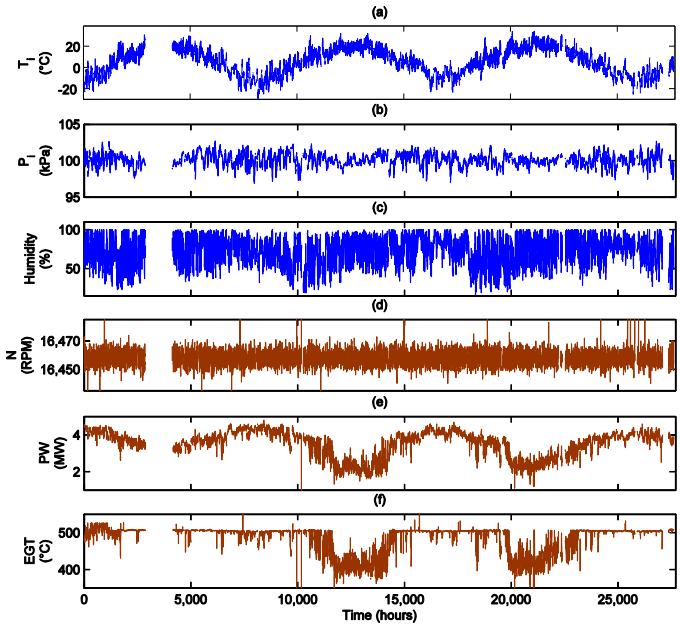


Fig. 1. Variations of the ambient conditions (blue), and the operating parameters (brown): (a) intake air temperature; (b) intake air pressure; (c) relative humidity; (d) shaft speed; (e) power and (f) *EGT*.

where ω_i is the specific humidity of the intake air and f is the fuel to air mass ratio. ω_i is a function of ϕ_i at a given temperature and pressure, and it was shown that f can be eliminated from the equations by taking into account the measurements on EGT and power in (1) and (2) respectively. As a result, the model equations reduce to:

$$PW_M = PW_M(T_i, P_i, \phi_i, N, EGT), \tag{3}$$

$$EGT_M = EGT_M(T_i, P_i, \phi_i, N, PW)\,. \tag{4}$$

Once the model is calibrated with the design parameters of a GTE, (3) and (4) predict the expected values for the power and the EGT in the ideal brand new condition. During the GTE operation, internal conditions of the machine deviate from their ideal level and the performance gradually deteriorates. Consequently, the model prediction will no more fit the measurements. At a given power, the higher performance deterioration accounts for the more energy waste through a warmer exhaust gas. Similarly, at a given EGT, a lower power is expected from a degraded GTE, compared to that of a brand-new GTE. Using this concept, two robust performance indicators were introduced [4] to quantify the level of the performance deterioration. The heat loss index is defined the difference between the measured EGT and the model prediction (EGT_M), normalized by the EGT at the design point (EGT_D):

$$HL = (EGT - EGT_M)/EGT_D. \tag{5}$$

The power deficit index is defined the shortage of the measured power from the model prediction (PW_M), normalized by the nominal power (PW_D):

$$PD = (PW_M - PW)/PW_D. \tag{6}$$

Fig. 2 shows the processes through which HL and PD are calculated from the GTE measurements.

C. Model Implementation Results

To investigate the GTE performance deterioration, the established model in Fig. 2 is used at each time step, and as a result, it creates a time series for each performance indicator. Variations of HL and PD throughout the operating time are plotted in Fig. 3. There are two distinct trends in the plots; a short-term sequential saw-tooth trend with gradual growths and sudden falls, and an overall slow growth during the entire operating time. The short-term trends in the indicators reflect the deterioration in the performance, which has recovered at the end of each segment.

The times of the sudden falls were checked with the GTE service logs, which revealed they are coincident with the dates of the compressor washes. There are seventeen distinct segments observable between consecutive compressor washes, which are given in Table II. As an example, Fig. 4 shows HL in the 5^{th} time segment. To predict the performance as a function of time, trend modeling techniques, such as curve fitting can be utilized in each time segment.

III. PERFORMANCE PREDICTION MODELING

A. Performance Trend Modeling by Curve Fitting

In curve fitting, there are always residuals, which are the differences between the data points and the predicted values by fitted curve. The root-mean-square residual (RMSR) provides the fraction of the standard deviation of the data that is explained by the fitted curve, and it can be used as a measure of accuracy to compare forecasting errors of different models for a particular variable, though not between different variables, because of its dependency on the scale [9].

Given a set of data values Y_t with n population, the fitted model estimates the values by $\hat{Y}_t$, which creates the residuals $r_t = \hat{Y}_t - Y_t$. The RMSR of the fitted model is:

$$RMSR = \sqrt{\frac{1}{n}\sum_{t=1}^{n} {r_t}^2}. \quad (7)$$

To utilize RMSR of a model for comparing modeling errors between different variables, the scale of the variables should be

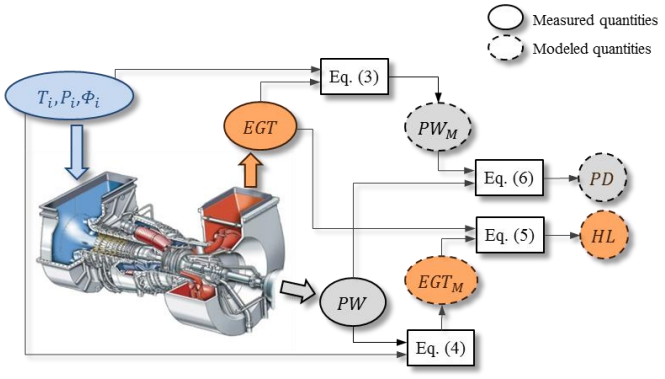


Fig. 2. The process flow to find HL and PD from the GTE measurements.

removed. As a result, the coefficient of variation of the RMSR is defined the RMSR normalized by the mean of the variable.

$$CV(RMSR) = RMSR/\bar{Y} \quad (8)$$

The best-fit curve is defined as a model, whose parameters are set such that the corresponding mean squared residual reaches to its global minimum. As a result, for a given type of function for the fitted curve, the modeling error increases if the

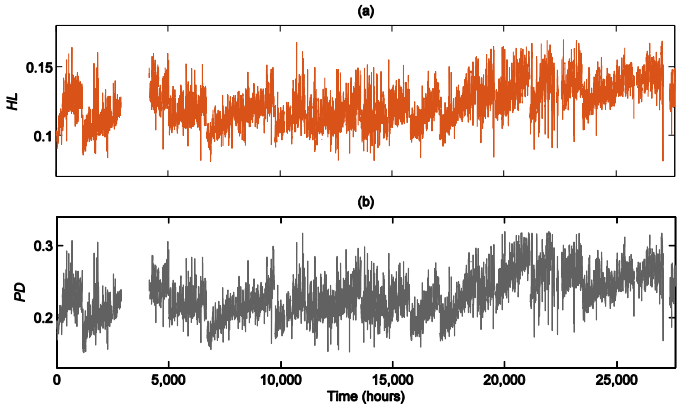


Fig. 3. Variations of the indicators:(a) HL; (b) PD.

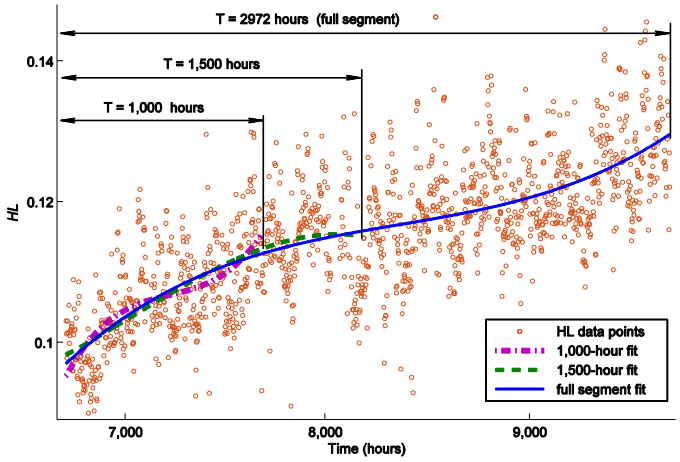


Fig. 4. The best-fit curves evolve upon new observations.

TABLE II. SEGMENTS OF SHORT TERM PERFORMANCE VARIATIONS

Segment	Period (hours)
1	2 – 1132
2	1150 – 2872
3	4128 – 5002
4	5012 – 6672
5	6714 – 9700
6	9746 – 11098
7	11158 – 13572
8	13584 – 15748
9	15786 – 17104

Segment	Period (hours)
10	17112 – 18782
11	18786 – 19524
12	19606 – 21122
13	21136 – 22128
14	22146 – 23434
15	23444 – 25790
16	25900 – 27082
17	27382 – 27612

curve deviates from the best-fit. For the short-term performance deterioration between two consecutive washes, the performance indicators gradually increase. To fit prediction models on the time series of the indicators, polynomial curves with different degrees were examined. Eventually, the third-degree polynomial curve was adopted, which comparatively showed smaller modeling errors.

It is known from the Bayesian inference that a subjective degree of belief will update to account for the newly observed evidences [10]. In this application, it means: given the indicator time series in an observation window and the corresponding best-fit model, the model will no more remain as the best estimator, if new data are added to the observation window. Assuming T the length of the observation window between the last compressor wash and the present time, the expected indicators $\widehat{HL}$and $\widehat{PD}$are estimated by the best-fit curve on the available data. When time elapses, Texpands and more data become available, what accounts for a new curve fitting. As a result, when the accuracy of the fitted model is subject to study, the length of the time window is an effective factor to be investigated. Fig. 4 shows the difference of the models, set on HL data in different observation windows in the 5th time segment.

B. Change of Sampling Rate

It is believed that higher sampling rate provides more updated data and a better understanding of the real-time state of a system. However, it accounts for more resources, especially when the labor is involved. When the sampling rate decreases, variations of the measured parameter between two measurements will be missed, and the corresponding fitted curve may deviate from the best-fit, which could have been fitted on potentially more samples. For a set of data with read out frequency of f_D, the decimation factor of M stands for sub-sampling with the frequency of $f_s = f_D/M$. Therefore, with regards to the original data points, $M-1$points will be lost between two consecutive sub-sampling, which may have contained useful information about the variation trend. At the same time, sub-sampling can have started from any of the 1st, 2nd,… or Mth time steps, which leads to$M$ independent subsets for the sub-sampled data points, i.e., $\{m, M+m, 2M+m, \ldots$, where$m=1, 2, \ldots, M$.

Fig. 5 depicts the effects of data decimation and the choice of the sub-sampled subset on the resulting fitted model. It shows the fitted models in the observation window of the first 960 hours (40 days) in the 5th time segment for HL. The orange bold line is the best-fit model on the entire record with f_s= 1/2 hr^{-1},within the time window. Then HL data is sub-sampled by the frequencies 1/12hr^{-1}. The green dash-dot line shows the fitted curve on the 1st subset and the purple dashed line is fitted on the 4th subset. Deviation of the fitted model from that of the original data indicates the dependency of the modeling accuracy on the sampling rate, and takes place in the cost of losing accuracy with lowering the sampling rates.

C. Modeling Error

As explained, sampling with the frequency of f_D/Mcreates M decimated subsets. In the observation window of T, we define $\hat{Y}_{T,M}^{(m)}$the set of n estimated values by curve fitting on the mth subset. For such subset, there will be a modeling residual $r_{T,M_t}^{(m)} = \hat{Y}_{T,M_t}^{(m)} - Y_t$ at each time step t, and the modeling error is:

$$\mathcal{E}_{T,M}^{(m)} = CV(RMSR)_{T,M}^{(m)}$$

$$= \sqrt{\frac{1}{n}\sum_{t=1}^{n} r_{T,M_t}^{(m)\,2}} \,/\, \bar{Y}, \qquad (9)$$

All decimated sub-samples have equal chances to be picked up. Therefore, the expected modeling error regardless of a specific decimated subset will be:

$$\mathcal{E}_{T,M} = \sqrt{\frac{1}{m}\sum_{m=1}^{M} \mathcal{E}_{T,M}^{(m)\,2}}. \qquad (10)$$

The expected modeling error $\mathcal{E}_{T,M}$ is independent from both the population size and the scale. As a result, it can be used to monitor the error when the observation window expands, the decimation factor changes, and even when it is intended to compare the modeling errors between two different indicators.

IV. RESULTING ERRORS AND ANALYSIS

Equation (10) shows the modeling error as a function of the sampling period and length of the observation window, which is depicted in Fig. 6 for selected time segments between compressor washes. It is seen that for both *HL* and *PD*, the modeling error steadily increases with the growth of the sampling period. At the same time, when the observation window expands, the modeling error becomes less dependent on the length of the window, and it gradually stabilizes.

Fig. 7 shows the variation of the average modeling errors for all the 17 time segments between consecutive compressor washes. It can be seen that modeling errors vary similarly for both indicators *HL* and *PD*. It means, there is no tangible difference between the time models established on either data to predict the indicator values. The plots show that the modeling errors are dependent on the observation window length, when the length is short. For the windows longer than

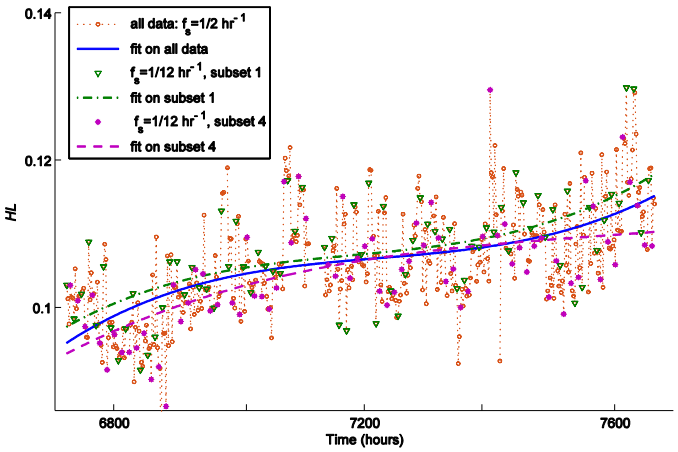


Fig. 5. Sampling rate affects the established model.

480 hours, the errors practically do not change any more, if the sampling frequency is fixed. At the same time the modeling errors are highly dependent on the sampling frequency, i.e., low sampling frequencies generate large errors, while the higher sampling frequencies lead to smaller errors. This dependency is more significant especially when the observation window is short. For instance, for $T = 120$ hours, with sampling frequency of less than 0.06 hr^{-1}, the modeling errors exceed 0.10, and the errors fall below 0.02, when the sampling frequency rises above 0.3 hr^{-1}. It is interesting that, for the sampling frequencies more than 0.08 hr^{-1}, the modeling errors become independent from the observation window length. The modeling errors are defined the normalized

standard deviation of the residuals. Given a normal distribution for the residuals, $\mathcal{E}_{T,M} = 0.02$ means that about 68% of the observations are predicted by the mode within a tolerance of ±2% of the average index value. If the model is built on the data with 0.03 hr^{-1} sampling frequency, the mentioned errors reach to 0.07. It means that the tolerance band should expand over 7% of the average index in order to include the same 68% of the observations.

V. CONCLUSION

In this paper, the effect of sampling frequency on the error of the prediction models for *HL* and *PD* indicators is studied. It is quantitatively shown that increasing the sampling frequency improves the modeling accuracy. It is also shown that establishing the model based on larger number of data points improves the modeling accuracy, however, the improvement is limited, i.e., after a certain number of data points, the accuracy will not increase any further. The results of the research work can be utilized to optimize the sampling frequency of the measured data if a certain level of accuracy is desired for performance monitoring models. The study also provides a measure for model prediction uncertainty, which is necessary when the model is employed in sequential data-model fusion frameworks for state estimation [11].

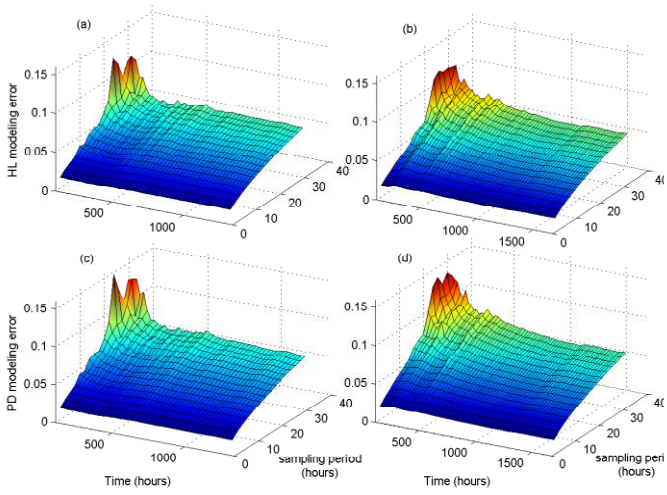


Fig. 6. Dependency of the modeling error ($\mathcal{E}_{T,M}$) on the length of observation window (T) and sampling period (f_s): *HL* modeling error in segments 9 (a) and 10 (b); *PD* modeling error in segments 9 (c) and 10 (d).

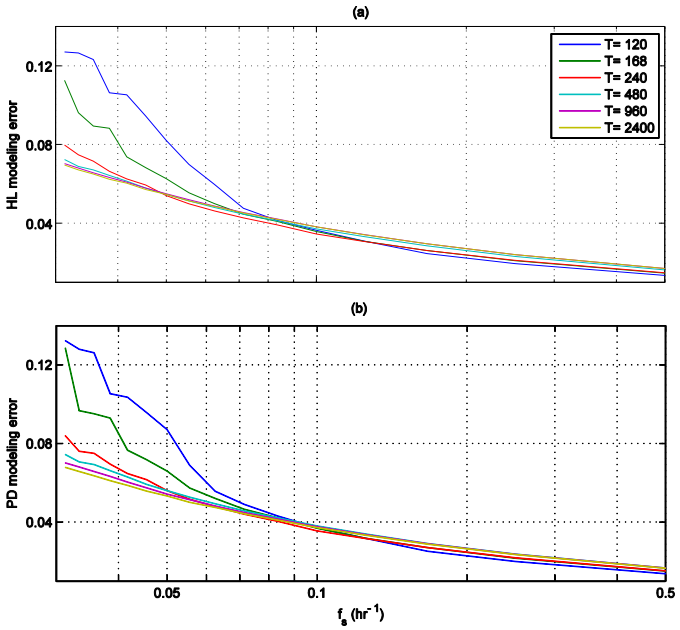


Fig. 7. Average modeling errors show similar behavior: (a) *HL* model error; (b) *PD* model error.

ACKNOWLEDGMENT

This project was funded and supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and Life Prediction Technologies Inc. (LPTi), Ottawa, Canada.

REFERENCES

[1] A. K. Jardine, D. Lin and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, pp. 1483-1510, 2006.

[2] M. Li, "Minimum description length based 2D shape description," in *Computer Vision, 1993. Proceedings., Fourth International Conference On*, 1993, pp. 512-517.

[3] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, pp. 10-21, 1949.

[4] H. Hanachi, J. Liu, A. Banerjee, Y. Chen and A. Koul, "A Physics-Based Modeling Approach for Performance Monitoring in Gas Turbine Engines," *Transactions on Reliability*, vol. PP, 2014.

[5] H. Hanachi, J. Liu, A. Banerjee, Y. Chen and A. Koul, "A physics-based performance indicator for gas turbine engines under variable operating conditions," in *Proceedings of ASME Turbo Expo 2014*, Düsseldorf, Germany, 2014.

[6] H. Hanachi, J. Liu, A. Banerjee and Y. Chen, "Effects of sampling decimation on a gas turbine performance monitoring," in *Prognostics and Health Management (PHM), 2014 IEEE Conference On*, Cheney, WA, USA, 2014, pp. 1-6.

[7] D. Sánchez, R. Chacartegui, J. Becerra and T. Sánchez, "Determining compressor wash programmes for fouled gas turbines," *Proc. Inst. Mech. Eng. A: J. Power Energy*, vol. 223, pp. 467-476, 2009.

[8] H. Hanachi, J. Liu, A. Banerjee and Y. Chen, "Effects of the intake air humidity on the gas turbine performance monitoring," in *ASME Turbo Expo 2015: Turbine Technical Conference and Exposition*, Montreal, Canada, 2015, pp. (accepted article in Turbo Expo 2015, Montreal, Canada).

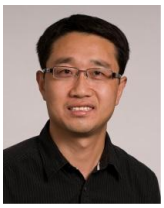
[9] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, pp. 679-688, 2006.

[10] J. Joyce, "Bayes' theorem," 2008.

[11] J. Sun, H. Zuo, W. Wang and M. G. Pecht, "Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 28, pp. 585-96, 04, 2012.

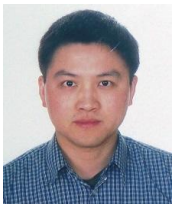
AUTHOR BIOGRAPHY

Houman Hanachi received his B.Sc. and M.Sc. in mechanical engineering from Sharif University of Technology, Iran and his Ph.D. in mechanical engineering from Carleton University, Ottawa, Canada. He worked as a professional engineer for thirteen years and is currently a postdoctoral research fellow and Term Adjunct Assistant Professor at Queen's University, Kingston, Canada. His research focus is on diagnostics, prognostics and health management, and hybrid techniques for state estimation, specialized in gas turbine engines.



Jie Liu received his B.Eng. in electronics and precision engineering from Tianjin University, China; his M.Sc. in control engineering from Lakehead University, Canada; and his Ph.D. in mechanical engineering from the University of Waterloo, Canada. He is currently an Associate Professor in the Department of Mechanical & Aerospace Engineering at Carleton University, Ottawa, Canada. He is leading research efforts in the areas of prognostics and health management, intelligent mechatronic systems, and power generation and storage. He has published over 30 papers in peer-reviewed journals. Prof. Liu is a registered professional engineer in Ontario, Canada.

Avisekh Banerjee, Ph.D, P.Eng., is the Systems Development and Services Manager at Life Prediction Technologies Inc. (LPTi). He manages the development of diagnostics and prognostics tools for turbo-machinery and avionics at LPTi. His broad areas of research interest are performing physics based prognostics case studies, ENSIP, data trending for failure prediction, development of parts life tracking systems and the development of PHM framework. He works extensively with end users requiring prognostics services.



Ying Chen is a senior Thermal Engineer at Life Prediction Technologies Inc. (LPTi). He received his B.E. and M.E. degrees in the Department of Power and Energy Engineering from Huazhong University of Science and Technology, Wuhan, China. He received his Ph.D. degree (2007) in the Department of Mechanical and Aerospace Engineering from Carleton University, Canada. He is in charge of the thermal aspects of all the projects at LPTi, and his research interests include gas turbine performance monitoring, conjugate heat transfer (CHT) based computational fluid dynamics (CFD) of turbomachinery components, thermal analysis of electromechanical systems and sustainable energies.