Modeling and identifying non-stationary long-term historical condition monitoring data in the presence of noise with non-Gaussian characteristics

Hamid Shiri[1], Pawel Zimroz[1], Agnieszka Wyłomańska[1], Radosław Zimroz[1]
(hamid.shiri, pawel.zimroz, agnieszka.wylomanska, radoslaw.zimroz)@pwr.edu.pl
[1]Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27
50-370 Wroclaw, Poland

Abstract

The advancement of condition monitoring systems has led to an increase in the utilization of long-term monitoring data for diagnostics and prognostics. The effective utilization of such data, collected over extended periods ranging from months to years, is a critical factor in enabling accurate diagnosis and prognosis. However, most of these industrial applications operate under time-varying conditions, making the diagnosis and prognosis approach to condition-based maintenance (CBM) complicated. Furthermore, many machines work in harsh environments, such as mining machines, wind turbines, helicopters, etc. The data acquired from these machines is often affected by noise with non-Gaussian characteristics. Therefore, it is a necessary task to analyse real data and introduce a proper model that could consider time-varying parameters and the effects of non-Gaussian noise. In this research, we first conducted a short literature review on the state of the art in long-term data modelling, focussing on statistical-based models. Then we analysed a well-known benchmark data set collected from the high-speed shaft of a wind turbine. Finally, we present the results for this data set and draw conclusions based on our findings.

1 Introduction

In the recent years, various methods have been developed and published in the field of prognosis , which can be broadly categorised into three main groups. The first group is the data-driven approach, which includes machine learning (ML) based approaches [1] and statistical-based model approaches [2]) . The second group is the physics-based approach, and the third group is the hybrid approach. Data-driven approaches attempt to construct a degradation model using historical data. They are particularly effective for complex engineering systems such as wind turbines, aircraft, and mining machines, where the relationship between degradation processes and physics is challenging to establish. Data-driven approaches fall into two categories: machine learning-based models and statistical-based models.Statistical-based approaches, have several advantages compared to previous methods. They do not rely heavily on extensive degradation data or deep mechanical knowledge of the equipment. Moreover, statistical models possess the potential to incorporate degradation uncertainties. Choosing the appropriate statistical model, stochastic or random, that closely represents the degradation process in a real-world scenario is crucial to effectively address diagnostic and prognostic challenges. Consequently, various approaches based on statistical models have been proposed in the existing literature to address this problem. In our previous work we introduced a framework for modelling and identifying the long-term condition monitoring data this framework [3]. The proposed framework encompasses the separation of deterministic and random components, modelling the time-varying properties of the heavy-tailed random part, and identifying any hidden autodependence within the random sequence. In addition, it addresses the identification of the distribution for the random component. Given the presence of non-linear trends, time-dependent scale (analogous to variance for Gaussian data), and non-Gaussian characteristics in the data, the final formulation of the model becomes complex. Consequently, its identification poses challenges and necessitates the use of specialised statistical methods suitable for heavy-tailed processes.

2 Results

This subsection applies the proposed framework to wind turbine data set. The data set used in this study is collected from a sensor mounted on the high-speed bearing shaft of a wind turbine. The data consists of measurements of the bearing inner race energy calculated every 10 minutes over a period of 50 days. The construction of the health index based on this data set is described in detail in reference [4]. It should be mentioned that during the data collection period, an inner race-bearing fault occurred, which was later confirmed by inspection. This data set, commonly referred to as the wind turbine data set, has been widely used in various studies to predict the Remaining Useful Life (RUL) of wind turbines. The results of the analysis conducted on the wind turbine data set are presented in Fig.1. In panel (a), the trend component (deterministic) is detected using a window of length 51. It can be observed that the detected trend component varies over time. After separating the deterministic trend, the random component of the health index is obtained, as shown in panel (b). The random component exhibits a non-homogeneous sequence with a time-varying scale, indicating that the scale (variance) of the random part changes over time. Panel (c) displays the identified scale of the random component. These results support our assumption of the increasing random component over time, with the scale growing non-linearly. The scale shows different patterns within the regimes of 0-2000 and 2000-5000, after which it becomes more complex. The first ten time-varying AR coefficients are presented in panel (d), with the number of coefficients chosen as a hyperparameter. It can be observed that these coefficients converge to a constant value over time. Particularly, the first coefficient of the AR model holds significant value.The residual of the model is depicted in panel (e), while its empirical autocorrelation function (ACF) is shown in panel (f). The empirical ACF plot suggests that the data can be considered as independent observations.In panel (g), the empirical and theoretical tails are plotted along with non-Gaussian distributions such as the stable and Student's t distribution. The residual series appears to correspond to a non-Gaussian distribution. Furthermore, in panel (h), the probability density functions (PDF) of the theoretical distributions for the residual signal are presented.

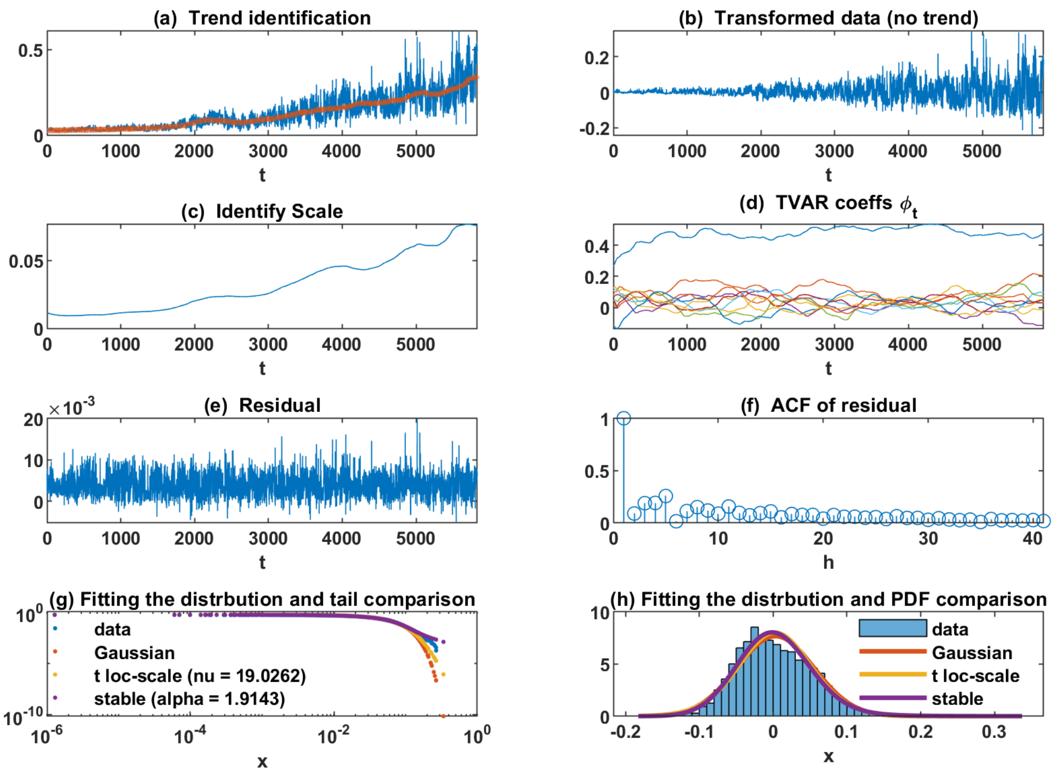

Figure 1: Results of the applied methodology for wind turbine data set.

To further validate our results, we analysed the quantile lines constructed on the basis of the fitted model for the wind turbine data set. The fitted models were synthesised, and the simulation results are presented in Fig. 2. The real data sets are represented by purple and blue lines, while the constructed quantile lines at the 5 and 95 percent levels are obtained from 400 simulated trajectories corresponding to the fitted models.

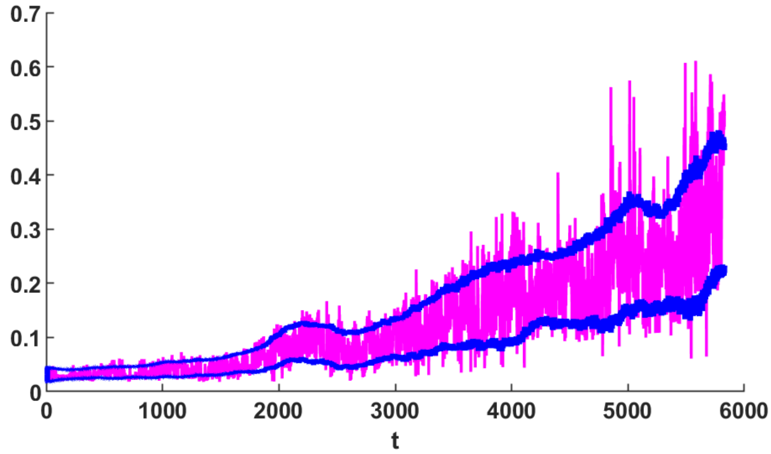

Figure 2: Constructed quantile lines (blue) on the level of 5% and 95% constructed on the basis of simulated trajectories corresponding to the fitted proposed model.

3 Conclusions

In this research, we introduced an approach to model and identify long-term condition monitoring. Our proposed model has the potential to be applied in online applications. We considered that the characteristics of the random component may vary over time, such as the growing scale (variance). Another novelty was the modelling of the random components using time-varying autoregressive time series, which allows for capturing the time-varying dependency of the random component. Also, for estimating the scale and AR coefficient we use MCKF which is a roust version of KF in the presence of non-Gaussian noise.

We applied the proposed procedure to real data sets, namely the wind turbine data sets. The results obtained from data set confirmed the effectiveness of the proposed approach in identifying and modelling the time-varying deterministic part, as well as the time-varying scale (variance) and autoregressive coefficients of the random part.

Acknowledgments

The authors (Hamid Shiri) gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie programme through the ETN MOIRA project (GA 955681).

References

[1] F. Moosavi, H. Shiri, J. Wodecki, A. Wyłomańska, R. Zimroz, Application of machine learning tools for long-term diagnostic feature data segmentation, Applied Sciences 12 (13) (2022) 6766.

[2] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, Remaining useful life estimation–a review on the statistical data driven approaches, European journal of operational research 213 (1) (2011) 1–14.

[3] W. Zulawinski, K. Maraj-Zygmat, H. Shiri, A. Wylomanska, R. Zimroz, Framework for stochastic modelling of long-term non-homogeneous data with non-gaussian characteristics for machine condition prognosis, Mechanical Systems and Signal Processing 184 (2023) 109677.

[4] E. Bechhoefer, R. Schlanbusch, Generalized prognostics algorithm using kalman smoother, IFAC-PapersOnLine 48 (21) (2015) 97–104, 9th IFAC Symposium on Fault Detection, Supervision andSafety for Technical Processes SAFEPROCESS 2015.