Data Descriptor

Prognosease: A Data Generator for Health Deterioration Prognosis

Tarek BERGHOUT * and Mohamed Benbouzid

Posted Date: 26 January 2023

doi: 10.20944/preprints202301.0473.v1

Keywords: Data generator; dataset; deep learning; health index; machine learning; prognosis and health management; remaining useful life

Data Descriptor

PrognosEase: A Data Generator for Health Deterioration Prognosis

Tarek Berghout [1] and Mohamed Benbouzid [2,3,*]

[1] Laboratory of Automation and Manufacturing Engineering, University of Batna 2, 05000 Batna, Algeria; t.berghout@univ-batna2.dz
[2] Institut de Recherche Dupuy de Lôme (UMR CNRS 6027), University of Brest, 29238 Brest, France
[3] Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China
* Correspondence: mohamed.benbouzid@univ-brest.fr

Abstract: This paper presents PrognosEase; a software that provides an easier way to produce different types of run-to-failure data mimicking real-world conditions to simplify prognosis studies in terms of data collection and improvement in ML degradation modelling process. Different types of degradation types made available to meet different types of applications. Besides, some primary ML tests were performed to ensure that complexity patterns of real systems could be observed in the training/testing predictions attitude. This paper also presents the impacts, limitations and potential improvements of the data generator.

Keywords: Data generator; dataset; deep learning; health index; machine learning; prognosis and health management; remaining useful life

---

Metadata

Nr	Code metadata description	*Please fill in this column*
C1	Current code version	*v1.0.0*
C2	Permanent link to code/repository used for this code version	*https://www.mathworks.com/matlabcentral/fileexchange/119743-prognosease*
C3	Permanent link to reproducible capsule	*none*
C4	Legal code license	*Batna 2 university*
C5	Code versioning system used	*none*
C6	Software code languages, tools and services used	*MATLAB*
C7	Compilation requirements, operating environments and dependencies	*MATLAB ≥ r2018b*
C8	If available, link to developer documentation/manual	none
C9	Support email for questions	berghouttarek@gmail.com

 doi:10.20944/preprints202301.0473.v1

1. Motivation and Significance

Prognosis and health management (PHM) is a discipline dedicated to study health deteriorations of systems under operating conditions [1,2]. Thus, it plays a crucial role in scheduling Condition-Based Maintenance (CBM) tasks while reducing downtime through early failure detection. Remaining Useful Life (RUL) is the primary health indicator, upon which the prognosis process depends to assess the spread of damage during system operating conditions. Indeed, this is the time between the current state of health (SoH) (i.e. the health state prediction time) and the time when the failure could occur. Logically, the run-to-failure RUL labels are obtained from real degradation cycles. However, for some systems (e.g. aircraft engines) it is impossible to achieve such conditions as a higher level of criticality and damage could be achieved, including financial, reputational loss and of life [3].

As an alternative, accelerated aging experiments and simulation models are the available data source used to build data-driven methods [3–5]. In this case, acceleration and simulation will not maintain RUL synchronization as in real conditions. Accordingly, other health indicators such as Health Index (HI) and Health Stage (HS) should be used to identify SoH of the system [1,2]. The HI index whose deterioration function is declined in different trends (e.g. linear and exponential) gives information on the current performance of the system. HS indicates the health level of the system at the time of SoH assessment based on some specific divisions (e.g. healthy, critical, and unhealthy SoH) defined by ML developers. Data obtained from accelerated tests suffers from missing patterns and labels, and also suffers from higher-level non-stationarity due to the harsh conditions imposed by the experiment [6]. Moreover, simulation models lack real data patterns. In addition to lacking authenticity at some point, the most significant drawback of both data collection methods is the cumbersome timelines and financial costs that make replication difficult to do.

In this context, it is important to provide an easy way to collect massive data needed to produce and study ML algorithms faster and more accurately as easily as possible. Accordingly, PrognosEase is introduced in this article with the aim of overcoming the shortcomings of accelerated aging and simulation models, simplifying ML studies and speeding up data generation process. The philosophy of PrognosEase depends on the generation of complete life cycles based on specific measurement types and the corresponding HI. These measurements are equivalent to sensor measurements in real applications and show a variety of trends, for example, linear degradation trends like in fuel cells [7], exponential like in turbofan engines [8,9], sinusoidal with exponential growth as in bearings [10], cyclic linear degradation trends as for Li-ion batteries [11]. Since the signals are generated according to some specific patterns, hence they lack RUL timings. In this case, HI and HS can be used. HI from PrognosEase comes with two different types, namely exponential and linear deterioration trends. Non-stationary and changing working conditions are generated as noise and distortion in a kind of randomly injected pulses in the generated measurements. Accordingly, this article is dedicated to presenting all these signals in relation to sensor and HI measurements. Also, it is dedicated to presenting some ML experiments indicating the training attitude and providing similar conclusions to accelerated aging and simulation experiments. In addition, it presents the impact of PrognosEase in the field of PHM, the limitations and potential improvements.

This paper is organized as follows. Besides, the introduction in section 1, section 2 describes PrognosEase and its main features. Section 3 is devoted to some ML experiments conducted using data generated by PrognosEase. Section 4 is devoted to the study of the impact of PrognosEase in the field of PHM. Section 5 presents limitations and potential improvements. Finally, Section 5 concludes this work.

2. Software Description

This section is dedicated to introducing different aforementioned types of measurements in life cycles generated by PrognosEase in two different subsections, both sensor measurements generation and also types of HI trends.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2023 doi:10.20944/preprints202301.0473.v1

2.1. Generating Sensors Measurements

As mentioned earlier, the measurement trends are generated according to some specific variations inspired by some well-known works in the literature, including linear, exponential and exponentially growing sinusoidal, cyclic with linear and exponential trends. Additional features such as noise and distortions have been added to emulate real-world conditions affecting the system.

Linear: The sensors measurement describing a linear trend L are generated according to a linearly growing function with the slop a and initial value b for an input time unites x as in one (1). A random noise μ is generated from a specific type of probability distribution $P(x)$ defined according to user experience as addressed by (2). After that, the noise μ and the degradation trend L are summed up while the noise is penalized with some specific noise rate ϑ to construct the signal S_N^* as in (3). Next, S_N^* will be scaled in range $[0,1]$ using min-max normalization as in (4) to obtain S_L^{**}. Finally, S_L^{**} will subject to some random distortion ρ with random number of pulses $n\rho$ controlled according specific amplitude w, and normalization factors $\{\alpha_\rho, \beta_\rho\}$, to generate the final measurements S_L as in (6) while, the pulses are periodically generated according to the discreet function as in (5).

$$L = ax + b \tag{1}$$

$$\mu = P(x) \tag{2}$$

$$S_N^* = (\vartheta L)\mu + L \tag{3}$$

$$S_N^{**} = \frac{S_N^* - min\,(S_N^*)}{max(S_N^*) - min\,(S_N^*)} \tag{4}$$

$$\rho = \{ we^{\pm ln(\alpha_\rho)x(i)/\beta_\rho} \}_{i=1}^{n\rho} \tag{5}$$

$$S_L = S_N^{**} + \rho^L \tag{6}$$

Figure 1 showcases an example constructing a linear sensors measurement trend with $P(x)$ is a Gaussian noise, $\vartheta = 0.9,\ n\rho = 2,\ w = 0.1$.

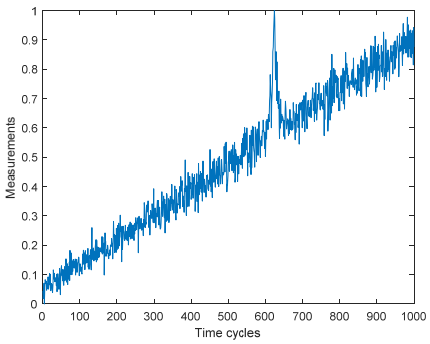

Figure 1. An example of a sensors measurements generated according a linear trend.

Exponential: Sensors measurements describing an exponential trend E are generated according to Formula (7) with the base of exponentiation e and exponentiation parameter α_E. The measurements of E follows similar steps of corruptions by adding the noise and distortions pulses while, scaling is also necessary (i.e. Equations (3)–(5)) to finally attend the S_E as in (8).

$$E = e^{(\alpha_E x)} \tag{7}$$

$$S_E = S_N^{**} + \rho \tag{8}$$

Figure 2 is an example constructing a sensor measurements according to an exponential trend with $P(x)$ is a Gaussian noise, $\vartheta = 0.03,\ n\rho = 2,\ w = 0.1.$ and $\alpha_E = 0.01$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2023 doi:10.20944/preprints202301.0473.v1

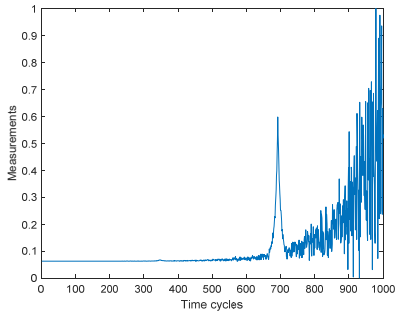

Figure 2. An example of a sensors measurements generated according an exponential degradation trend.

Sinusoidal with exponential growth: This type of measurements of sinusoidal with exponential growth SE are generated according to Formula (9) with an exponentiation parameter α_{SE} and angular frequency ω while being subject to distortion by noise and randomly injected pulses until the final shape S_{SE} reached using Formula (10).

$$SE = e^{(\alpha_{SE}x)}cos(\omega x) \tag{9}$$

$$S_{SE} = S_N^{**} + \rho \tag{10}$$

Figure 3 is an example constructing a sensor measurements according to a sinusoidal with exponential growth trend with $P(x)$ is a Gaussian noise, $\vartheta = 0.1,\ n\rho = 2,\ w = 0.1.\ \alpha_{SE} = 0.02$, and $\omega = 0.2$.

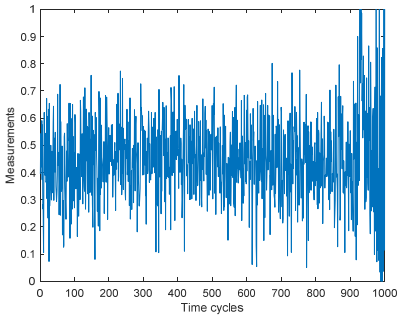

Figure 3. An example of a sensors measurements generated according a sinusoidal with exponential growth trend.

Cyclic degradation: The cyclic degradation S_C comes up with two types (i.e. linear and exponential trends). The pulses are similarly generated according to Equation (5) but periodically with a distance between pulses D dynamically changes at each cycle with a specific user ratio ϑ_C as in Formula (11) while (12) is describing the final output signal.

$$\rho = \{ we^{\frac{\pm ln(\alpha_\rho)x(i)}{\beta_\rho}} + \vartheta_C D\}_{i=1}^{n\rho} \tag{11}$$

$$S_C = S_N^{**} + \rho \tag{12}$$

Figure 4 is an example that showcases both types of cyclic degradation that PrognosEase programed to do.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2023 doi:10.20944/preprints202301.0473.v1

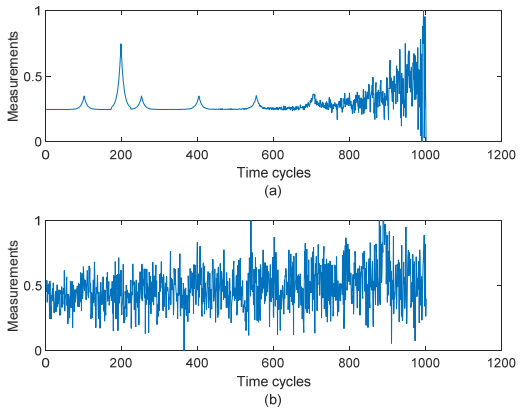

Figure 4. An example of a sensors measurements generated according a cyclic degradation trend: (a) An exponential cyclic degradation; (b) A linear cyclic degradation.

2.2. Generating RUL Measurements

In PrognosEase, RUL degradation function is actually a HI generated according to two main trends either exponential or linear following Formulas (1) and (7) respectively with different parameters values from the sensors measurements. These parameters are user-defined ones depends on accuracy of predictions and also user experience. The example in Figure 5 illustrates both types of HIs.

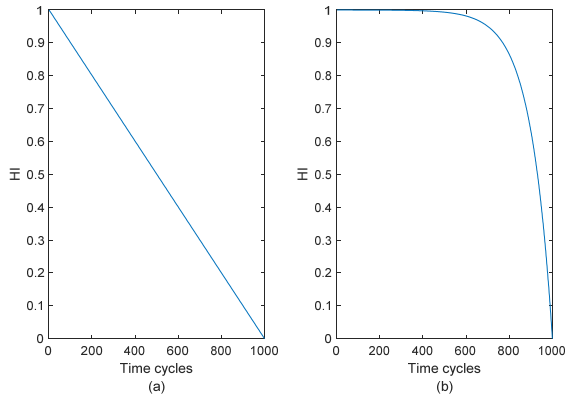

Figure 5. An example of HI degradation functions: (a) A linear degradation; (b) An exponential degradation.

3. Illustrative Examples

In this study, PrognosEase is used to generate a dataset that combines a mixture of sensor measurements types with a higher level of complexity, similar to the work done in the literature in terms of data generation. 10 features, 5 lifecycles for training and 2 cycles for testing are generated with cycles that are 1000 samples long. The training and testing lifecycles are showcased in Figure 6. The data visualization addresses that the constructed features space includes different types of complex samples with higher-level non-stationarity. In fact, these samples are supposed to be flown from the same devices under different conditions. This means that all data generation parameters are fixed when reconstructing the data set while retaining only the noise generation and distortion that changes the conditions. This is done with the aim of reaching a certain level by mimicking real-world conditions. HI is set in this case by default to the exponential deterioration function due to the existence of an exponential deterioration phenomenon in the collected measurements.

An ML model, namely a long-term memory neural network, is tested for its potential capabilities using a grid search mechanism to adjust its parameters for a better approximation. The curve fitting

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2023 doi:10.20944/preprints202301.0473.v1

results for training and testing are shown in Figure 7. The learning model shows that it tries to mimic the shape of the decay trend in the training process as it is logically difficult to do for new unseen samples to the model. Thus, improving ML models will consist of improving the curve fitting as much as possible in the testing phase while taking into account real-world conditions and training constraints such as centralized, decentralized and federated learning, etc., with all types of learning paradigms such as online, offline and reinforcement learning etc.

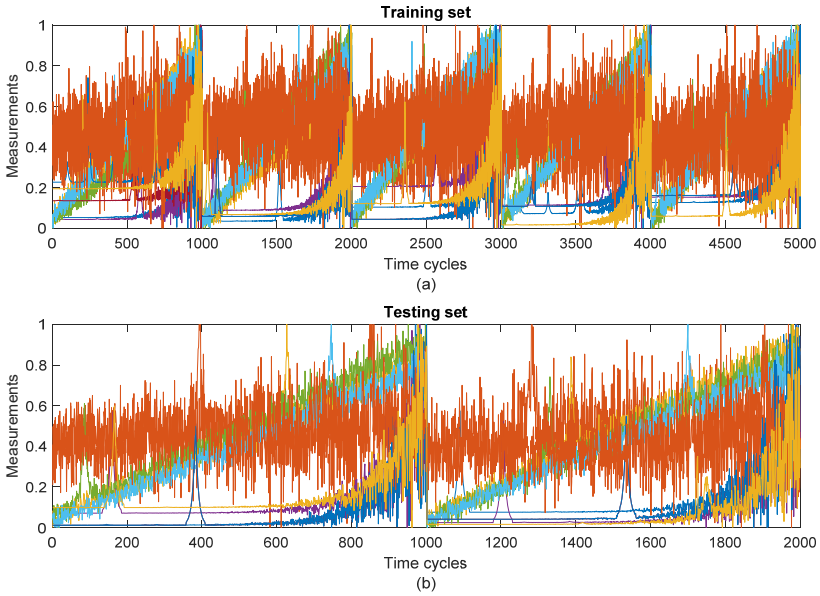

Figure 6. Visualizing feature spaces of a dataset generated by Prognosis: (a) Training set; (b) Testing set.

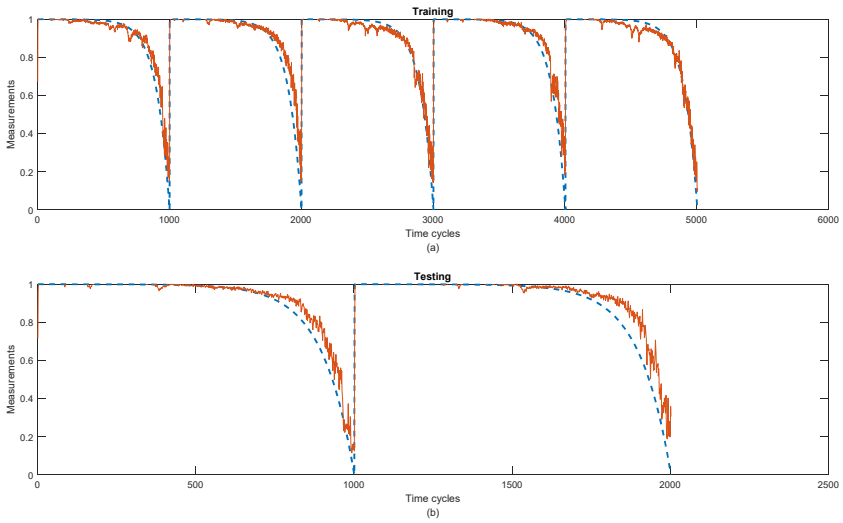

Figure 7. Visualizing Curve fit results with a deep learning network: (a) Training set; (b) Testing set.

4. Impact

PrognosEase actually introduced an effective and a simple way to study ML models for deterioration analysis. Its simplicity remains in following items:

- Easily generating many types of degradation samples with any preferred number of training and testing cycles making useful for many types of many applications.
- PrognosEase simplifies centralized learning for both offline and online learning.

- PrognosEase also allows experiments to be done according decentralized and federated learning [12].
- PrognosEase provides a feature space accessing times series experiments for both cyclic and non-cyclic degradation.
- PrognosEase provides features achieving great experiments on direct HI predictions experiments.
- PrognosEase doesn't follow supervised direct HI and time series predictions; in fact, it can be used for unsupervised HS assessment with clustering methods also.
- PrognosEase allows experiments not only be limited to improving ML methods. In fact, it can also be used for studying data quality and preprocessing tools also.

5. Limits and Potential Improvements

PrognosEase has limits in addressing reality in context of following items:

- Similar to simulation and accelerated aging, PrognosEase doesn't have the capability of estimating actual RUL timing. Rather than that, it provides His instead.
- PrognosEase comes up with Gaussian noise only, so potential improvement will consist of adding more types of noises to approach reality when emulating real conditions.
- Only single type of pulses generation when distorting generated measurements. Accordingly, more efforts can be made on generating other pulses types mimicking distortions in real applications.
- For cyclic degradations better formulas could be released to better control cycles dynamic changes as in real systems such as for batteries for instance.

6. Conclusions

This paper presented PrognosEase; a software for generating run-to-failure data for Data-driven prognosis studies. PrognosEase mimics real conditions of different systems by generating similar measurements to their real degradation trends and working conditions. Data visualization for feature spaces generated by PrognosEase and Some ML experiments shows that the software is able to address prediction complexity for unseen samples as in real samples. This paper also presented impact of PrognosEase in PHM filed besides to some future prospects on its potential improvements.

References

1. T. Berghout and M. Benbouzid, "A Systematic Guide for Predicting Remaining Useful Life with Machine Learning," *Electronics*, vol. 11, no. 7, p. 1125, Apr. 2022, doi: 10.3390/electronics11071125.
2. Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018, doi: 10.1016/j.ymssp.2017.11.016.
3. A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, Oct. 2008, pp. 1–9, doi: 10.1109/PHM.2008.4711414.
4. P. Nectoux *et al.*, "PRONOSTIA : An experimental platform for bearings accelerated degradation tests.," in *IEEE International Conference on Prognostics and Health Management, PHM'12*, 2012, pp. 1–8, [Online]. Available: http://hal-obspm.ccsd.cnrs.fr/UNIV-BM/hal-00719503.
5. Prognostics Center of Excellence, "PRONOSTIA-FEMTO Bearing Data Set," 2012. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan.
6. T. Berghout, L.-H. Mouss, T. Bentrcia, and M. Benbouzid, "A Semi-Supervised Deep Transfer Learning Approach for Rolling-Element Bearing Remaining Useful Life Prediction," *IEEE Trans. Energy Convers.*, vol. 37, no. 2, pp. 1200–1210, Jun. 2022, doi: 10.1109/TEC.2021.3116423.
7. T. Berghout, M. Benbouzid, T. Bentrcia, Y. Amirat, and L. Mouss, "Exposing Deep Representations to a Recurrent Expansion with Multiple Repeats for Fuel Cells Time Series Prognosis," *Entropy*, vol. 24, no. 7, p. 1009, Jul. 2022, doi: 10.3390/e24071009.
8. T. Berghout, L. H. Mouss, O. Kadri, L. Saïdi, and M. Benbouzid, "Aircraft engines Remaining Useful Life prediction with an adaptive denoising online sequential Extreme Learning Machine," *Eng. Appl. Artif. Intell.*, vol. 96, p. 103936, Nov. 2020, doi: 10.1016/j.engappai.2020.103936.

9. T. Berghout, L. Mouss, O. Kadri, L. Saïdi, and M. Benbouzid, "Aircraft Engines Remaining Useful Life Prediction with an Improved Online Sequential Extreme Learning Machine," *Appl. Sci.*, vol. 10, no. 3, p. 1062, Feb. 2020, doi: 10.3390/app10031062.
10. T. Berghout, M. Benbouzid, and L. H. Mouss, "Leveraging label information in a knowledge-driven approach for rolling-element bearings remaining useful life prediction," *Energies*, vol. 14, no. 8, p. 2163, Apr. 2021, doi: 10.3390/en14082163.
11. B. Saha and K. Goebel, "Battery data set," *NASAAMES Progn. Data Repos. 2007*.
12. T. Berghout, T. Bentrcia, M. A. Ferrag, and M. Benbouzid, "A Heterogeneous Federated Transfer Learning Approach with Extreme Aggregation and Speed," *Mathematics*, vol. 10, no. 19, p. 3528, Sep. 2022, doi: 10.3390/math10193528.