A Modelling Ecosystem for Prognostics

Lachlan Astfalck[1], Melinda Hodkiewicz[1], Adrian Keating[1], Edward Cripps[2] and Michael Pecht[3]

[1] *System Health Laboratory, The University of Western Australia, Perth, WA, 6009, Australia*
lachlan.astfalck@uwa.edu.au
melinda.hodkiewicz@uwa.edu.au
adrian.keating@uwa.edu.au

[2] *School of Mathematics and Statistics, The University of Western Australia, Perth, WA, 6009, Australia*
edward.cripps@uwa.edu.au

[3] *Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD, 20742, USA*
pecht@calce.umd.edu

ABSTRACT

This paper evaluates data-driven asset prognostic models from a modelling ecosystem perspective, which includes data description, uncertainty quantification, model selection justification and validation, and application limitations. An easily accessible and comprehensive ecosystem enables efficient reproducibility of previous work to facilitate both the adoption of the models by industry and the development of future scientific methods. The results of this study enable the development of a list of ecosystem elements to accompany the publication of new models. By describing the ecosystem in the communication of new models, researchers can ensure the reproducibility of their models in the wider prognostic community.

1. INTRODUCTION

Prognostics and health management (PHM) is a process that enables the assessment of an asset's reliability under its actual application conditions (Pecht, 2008). An integral element of PHM is the ability to predict the remaining useful life (RUL) of an asset through the use of prognostic models. The ability to develop and use prognostic models is a necessary competency for manufacturers, service providers and asset operators as a means to ensuring assets are reliable, safe, and cost-effective.

Prognostic methods can be broadly classified into two approaches: physics of failure and data-driven. The primary focus of this paper is on data-driven prognostics in which data is collected from laboratory-scale or in-service industrial, commercial, or infrastructure assets. The data required for these data-driven models can include internal covariates (e.g., temperature, vibration) measured by sensors on the asset and only present when the asset is operating and external covariates (e.g., weather data, location), which are present whether or not the asset is operating. This data and an appropriate modelling method(s) generate the RUL estimation. Considerable work has been done since the 1980s by academics and the PHM community to develop prognostic models. However, although asset data is widely used by industry for diagnosis and fault detection, the use of data-driven prognostic models to predict RUL of commercial and infrastructure assets is not part of business-as-usual (Sikorska, Hodkiewicz, & Ma, 2011).

In many sectors, PHM model implementation by industry is in its infancy, in part due to practical difficulties of real-world implementation. There is often a lack of available data to validate techniques, since industry seldom allows their assets to run to failure (Kan, Tan, & Mathew, 2015; Saxena, Goebel, Simon, & Eklund, 2008). Other issues affecting the uptake of PHM by industry include how to assess model performance and uncertainty quantification (Sankararaman, Saxena, & Goebel, 2014). This paper considers these issues alongside a broader movement in the science community towards greater reproducibility.

Reproducibility of scientific findings is currently a topic of interest in many differing fields, including psychology (Open Science Collaboration, 2012), biostatistics (Peng, 2009), epidemiology (Peng, Dominici, & Zeger, 2006), and land ecology (Gutzwiller & Riffell, 2014). These studies have shown concerns for the reproducibility of previous findings, which has brought the validity of some past results into question. The journals *Nature* and *Science* have published editorials on the matter stating that *"Reproducibility, rigour, transparency and independent verifications are cornerstones of the scientific method"* (McNutt, 2014). Some journals have set submission requirements, such as full data and code transparency, to ensure the reproducibility of future findings. Enforcing data and code transparency enables 1) verification of past findings, 2) alternative analysis to address model suitability, 3) raising of concerns with initial findings that may be detrimental to further research, and 4) speeding up the development of ideas among researchers.

Improvements in the ability to reproduce PHM models may be helpful in improving the uptake and use of these models in industry. To standardise PHM reproducibility, the "modelling ecosystem" has been defined as a group of elements prevalent in all prognostic models. This definition is similar to other reproducibility recommendations shown to be common in other sciences. Suggested communication recommendations are given for each of these elements to ensure reproducible and transparent research.

The remainder of this paper is organised as follows: Section 2 presents the background of reproducibility research in other scientific disciplines. Section 3 suggests some standards of reproducible research for PHM. Section 4 states the results and findings of a review of the reproducibility of current PHM research. Section 5 presents discussion on both the reproducibility recommendations of PHM and on the modelling ecosystem. Finally, Section 6 states the conclusions of the research with recommendations for future work.

2. BACKGROUND

Replicability and reproducibility vary in definition across disciplines and authors. This paper uses definitions as given in (Leek & Peng, 2015). The replicability of a study is defined as the confirmation of previous findings by using independent data, research techniques, laboratories, equipment, and investigators. Reproducibility is defined as the ability to recompute results given the data set and knowledge of the data analysis pipeline. The replication of scientific findings is the ultimate standard by which scientific findings are judged, as it acts to address spurious claims and to enforce a disciplined approach to scientific findings (Peng, 2011). However, as studies have increased in complexity they have become harder to independently replicate, which has led to the reproducibility of studies becoming the satisfactory standard for scientific validation. Reports from numerous fields of widespread irreproducible findings have brought reproducibility concerns to the forefront of many high-ranking journals (McNutt, 2014; Nature Editorial, 2014; Casadevall & Fang, 2010; Steckler, 2015).

Attempts to reproduce the findings or assess the reproducibility of studies in many scientific disciplines have produced less than favourable results. In 2015, the Centre for Open Science released findings from a project to reproduce the work of 98 studies from three psychology journals. The project found that only 39 out of 100 cases (two studies were reproduced twice each by independent groups) were able to be reproduced (Open Science Collaboration, 2015). Furthermore, the research found that whilst 97% of the original studies reported significant results (p-value <0.05), whilst only 36% of the reproduced studies found significant results. Preclinical research has shown repeated failings of reproducibility studies. The biotechnology firm Amgen was only able to confirm the scientific findings of 6 out of 53 papers (Begley & Ellis, 2012), and pharmaceutical company Bayer HealthCare reported that only 20%-25% of 67 studies aligned with their in-house findings (Prinz, Schlange, & Asadullah, 2011). Lastly, Kilkenny et al. (2009) found in a survey of 271 studies relating to animal research, that many studies are not seen to be reproducible based upon the amount of information included.

As well as integrity issues, irreproducible research is an economic detriment in many fields. Freedman, Cockburn, and Simcoe (2015) showed that approximately US$28B/year is spent on irreproducible preclinical research, an estimated 50% of research in the field. Furthermore, a series of five articles and an editorial published in *The Lancet* concluded that 85% of biomedical research funding is being wasted on research that is inappropriately analysed and inadequately reported (Chalmers et al., 2014; Ioannidis et al., 2014; Salman et al., 2014; Chan et al., 2014; Glasziou et al., 2014; Macleod et al., 2014).

In response to these findings and others, organisations such as journals, academia, and government bodies have taken measures towards ensuring the reproducibility of future published studies (McNutt, 2014; Nature Editorial, 2014; Nosek et al., 2015; US National Institutes of Health, 2014). For example, *Nature* now mandates that materials, data, code, and associated protocols are to made promptly available to readers without undue qualifications (Nature, 2016). Bissel (2013) has warned that the increase in reproducibility exposure should not encourage scepticism towards other scientists in the field, but encourage that research is conducted in a scientific and ethical manner.

3. DEFINING REPRODUCIBLE RESEARCH FOR PHM

Table 1 defines the reproducibility criteria of a prognostic model, adapted from US National Institutes of Health (2014); Peng et al. (2006); Schwab, Karrenbach, and Claerbout (2000). The reproducibility recommendations can be separated into the three following components: data and code transparency, documentation, and distribution. These criteria

Table 1. Developed criteria for PHM reproducibility: data and code transparency, documentation, and distribution.

Research Component	Reproducibility Requirement
Data and Code Transparency	Cleaned data set made available, with possibility of including preprocessing data set from acquisition stage. Computer code underlying all predictive modelling elements, figure generation, and other principal results is made available with the environment necessary for execution
Documentation	Documentation of the data set, data processing, and computer code made available to enable repeat analysis and to conduct similar analyses. Limitations of model inputs, uncertainties, model selection justification, performance verification, and application limitations all made clear
Distribution	Software, data, and documentation distributed in an appropriate manner with appropriate privacy regulations

are seen to be the minimum standard of reproducibility across all scientific disciplines that were examined.

At a minimum, the processed data that feeds into the model, defined as the prognostic data, should be made available. Due to the nature of PHM, this data is rarely the same as the raw data set from data acquisition and has likely undergone data cleaning, data transformation, and feature vector selection. Whilst not compulsory for reproducibility, it is recommended that the raw data set is included alongside the processed data. Further, given that the simplest way to reproduce a reported outcome is to execute the model, computer code should be made available for independent verification and validation.

Documentation of analyses and findings is recommended across all disciplines, however, the suggested contents of the documentation may differ between disciplines. For PHM research, the suggested documentation should include the following elements: model inputs, uncertainties, model selection justification, performance verification, and application limitations. These elements are herein referred to as the "modelling ecosystem".

Finally, all of the mentioned elements should be distributed in an appropriate manner. Whilst journals are the current standard for research dissemination, they may only accept responsibility for publishing scientific findings. Distribution of data, code, and attached documentation then becomes the responsibility of the author.

4. Study of Current PHM Research

To evaluate the adherence of prognostic model papers to the reproducibility recommendations stated in Table 1, 50 papers from the journals *Mechanical Systems and Signals Processing*, *Reliability Engineering and System Safety*, *IEEE Transactions on Reliability*, and *The International Journal of Prognostics and Health Management* were surveyed between the years 2000 to 2014. Table 2 shows the results of the survey. Papers were reviewed against a set of pre-chosen criteria selected to reflect both the reproducibility of the models and attention given to the components of the modelling ecosystem.

To focus results, this survey only considered data-driven prognostic models. This review will be extended to physics of failure and hybrid models, however, the inclusion of such is outside the scope of this study.

Whilst most papers discuss data acquisition, few provide sufficient descriptors of data cleaning or feature selection processes. Data acquisition, cleaning, and feature selection are all widely accepted as being necessary to prognostics.

Model selection justification, in the context of the input data and the required outputs, occurred in 26% papers. Many papers, whilst providing descriptions of the existing, adapted, or newly developed models, did not include reasoning on why the model type was appropriate for the data.

Sankararaman and Goebel (2015) stated that uncertainty should be included from system-level conception through to operations, and not solely in the later stages of development. Few papers discussed uncertainty and the effects that it may have on model performance. Whilst many models yielded stochastic outputs, uncertainty was rarely discussed previous to the analysis stage. Furthermore, sources of uncertainty and their effects on the model were seldom discussed. Recent research shows a rise in the inclusion of uncertainty discussion. 34% of the papers surveyed across all years address uncertainty, conversely 52% of papers published in years 2013 and 2014 consider prediction and model uncertainty. A good discussion and incorporation of uncertainty was found in Sun, Zuo, Wang, and Pecht (2012), whose results were presented alongside an initial discussion of uncertainties inherent in the model and graphically depicted the prediction uncertainty by use of confidence bounds.

Performance evaluation of models using either basic metrics or prognostic metrics, developed by Saxena, Celaya, Saha, Saha, and Goebel (2010), occurred in 50% papers. Due to the recent development of prognostics metrics with respect to the age of papers in the review, any performance evaluation was accepted by the survey as having included performance evaluation. However, it should be noted that performance evaluation using tailored prognostic metrics should be the future standard for developments.

Eight papers list the availability and location of the data used, whilst only one paper mentions code availability. Six out of the eight datasets available originate from public data stored in the NASA Ames data repository. Authors have not been

Table 2. Results from the examination of PHM data-driven model development papers from: *Mechanical Systems and Signals Processing*, *Reliability Engineering and System Safety*, *IEEE Transactions on Reliability*, and *The International Journal of Prognostics and Health Management*

Criteria	**No. of papers**
Total papers	50
Industrial data used	8
Experimental data used	34
Simulated data used	21
Modelling ecosystem	
Data description	33
Model selection justification	13
Performance validation	25
Uncertainty discussion	17
Application limitations	9
Data reported to be available	8
Prognostic model reported to be available	1

contacted regarding private sharing of their data or code and it is not apparent how many authors would be willing to oblige.

5. DISCUSSION

Reproducible research has been defined by three stages, across all scientific fields, as outlined in Table 1. Discussion on the findings of the review, with reference to each stage, is provided below.

5.1. Data and Code Transparency

Sonnenburg et al. (2007) outlined seven reasons supporting open source software in the machine learning community, the first of which being the reproducibility of scientific results and fair comparison of algorithms. Authors are naturally concerned about dissemination of their data and code, if only due to potential usage and privacy issues. To combat this, a system that allows the partial rights to a data set or computer code may be implemented to restrict public usage but still enable transparency. Peng et al. (2006) defined the following four possible classes of data licences developed out of the Creative Commons project, in order of restrictiveness:

1. **Full access** - Data or code may be used for any purpose.
2. **Attribution** - Data or code may be used for any purpose so long as the authors are cited.
3. **Share alike** - Data or code may be used to produce new findings or results. Any modifications that are used to produce new findings must be made available under the same terms.
4. **Reproduction** - Data or code may be used for the purpose of reproducing results in the associated published article or for commenting on those results via a letter to the editor. No original findings based on the data may be published without explicit permission from the original investigators in a separate agreement.

A common concern is that by making data and code open source researchers may subsequently be limited in their ability to patent the technology or to create closed-source products for industry. Sonnenburg et al. (2007) argues that careful selection of a suitable open source license would satisfy the requirements of most researchers and their employer. Further, the development of closed-sourced products for industry may be aided by the use of external open-source contributions. For further conversation on the economic motivations for open source software the reader is referred to Riehle (2007).

5.2. Documentation of the Modelling Ecosystem

The elements of the modelling ecosystem, as seen in Figure 1, are widely acknowledged as being important factors in any prognostic system. One or two of these elements were often discussed in the 50 publications reviewed. However, documentation of all elements was seldom seen. It is assumed likely that many, if not all, of these criteria were addressed during model development and analysis with only the information deemed important by authors included alongside the findings. Whilst word limitations and avoidance of extraneous information may inhibit inclusions of all of these elements in publication, it is recommended that this information be listed elsewhere. At best, the absence of information surrounding model development represents a lack of clarity. At worst, it may indicate that important assumptions or stages are not addressed.

Reproducibility is directly influenced by transparency. To aid model reproduction to assist in wider industry uptake, it is suggested that the modelling ecosystem should be effectively documented. It is expected that all models should be designed to yield repeatable findings for the data that they were designed and tested for, as well as produce similar findings for data sets within their application limitations.

A summary of the authors' modelling ecosystem inclusion recommendations is presented below. It should be noted that these recommendations have been written as to open up a conversation about requisite descriptors for prognostic models, and not as a critique of any prior work. To promote reproducibility in the prognostics community, documentation of these recommendations should become standardised and publication requirements for future prognostic models agreed upon.

5.2.1. Data Description

In any prognostics model, incoming data is acquired and cleaned, and then features are selected for analysis. Data acquisition, such as asset classifications (e.g., rpm, power, voltage, etc.), sensor placement, and sampling frequency, is of-

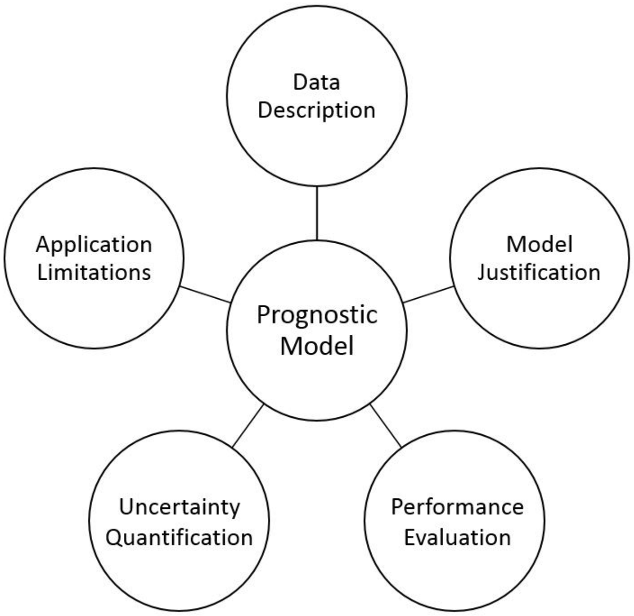


Figure 1. Elements of the prognostic modelling ecosystem: data description, model selection justification, performance evaluation, uncertainty quantification, and the application limitations.

ten described. However, rarely is the data cleaning or feature selection mentioned. It is understandable that much of this work can be highly detailed, and not suitable for publication, should the focus be on prognostic modelling. In this case separate publication or documentation of the data cleaning and feature selection processed are encouraged to be referenced in the modelling work.

A further challenge is measuring and communicating the quality of the data for prognostic modelling, as a standardised method for doing so has not as of yet been developed. Coble (2010) does present an automated method for identifying suitable prognostic parameters based on features such as trendability, monotonicity, and prognosability. Whilst aiding in parameter and feature selection, it does still not provide a metric for evaluating the data as a whole. Evaluating the suitability of a dataset for prognosis ensures that a clear relationship with failure behaviour or a reasonable trend is seen, and that data acquisition is performed correctly.

Appropriate data description should include data acquisition and should mention the requirements for data inclusion, as presented in Hodkiewicz and Montgomery (2014), data cleaning methods, and feature extraction methods. Data acquisition and processing is essential for model reproducibility, so discussion should be comprehensive enough so that a skilled researcher can extract the same features from the same raw data set.

5.2.2. Model Selection Justification

To aid transparency, researchers should discuss the reasons for their model selection to demonstrate model appropriateness, show consideration has been given to other potential models, and guard against 'flavour of the month' models. Many reviews have been released presenting advantages and disadvantages of models as well as selection criteria, such as:

- Aizpurua and Catterson (2015) presented a design decision framework for model selection. Models are separated into data-driven and model-based approaches with selection flowcharts provided based on input data.
- Sikorska et al. (2011) discussed the advantages and disadvantages of prognostic models. Further issues such as the fitness for purpose of the model for a specific business application and the ability of business to support the use of the model are covered.
- Kan et al. (2015) discussed data-driven modelling options for non-stationary and non-linear systems. The resource requirements of each model are discussed along with advantages, disadvantages, application history, and usage requirements. An example of a standardised scoring system is given to rank modelling techniques based on the machine type, available data, and the performance evaluation parameters.
- Lee et al. (2014) suggested the implementation of a quality function as a ranking method for algorithm selection as well as providing an overview of many prognostic models.

From the reviews listed above, a number of points relevant in a model selection discussion include but are not limited to data type and availability, prediction horizon, expert knowledge required to build and maintain the model, computational requirements, and management system support (hardware, software, and competencies required).

5.2.3. Model Performance Evaluation

Classical metrics for performance evaluation have existed in the literature for many years, including: accuracy (bias), precision (spread), timeliness, mean square error, and mean absolute percentage error (Vachtsevanos et al., 2006). Furthermore, off-line evaluation metrics have been developed with specific reference to prognostics, including providing a prognostic horizon, alpha-beta performance, relative accuracy, and convergence (Saxena et al., 2010).

For models to be used for maintenance planning, there is a work planning period. This period is necessary to ensure the right resources are available (Kelly, 2006). For the model to support planned maintenance activities, it must be able to predict RUL further out than the planning window. For example, if the planning period is 7 days, the model should be able to infer that failure might occur in more than 7 days' time with an appropriate level of confidence. Some discussion about these types of practical issues would assist potential model users.

5.2.4. Uncertainty

The nature of prognostics results in uncertainties from all stages of model design and implementation, such as data noise, sensor error, model uncertainties, processing faults and prediction uncertainties. The need to discuss uncertainty from representation, quantification, and management points of view has been widely discussed (Sankararaman & Goebel, 2015; Sankararaman, Daigle, & Goebel, 2014). Uncertainty should be included from initial system-level conception through to operations and not just in the latter stages after a prediction has been established. An in-depth review of the impacts, interpretations, sources, types, and challenges of uncertainty in prognostic models is presented in (Sankararaman, Saxena, & Goebel, 2014).

Uncertainty quantification is a complex issue, and still the subject of research in prognostics. Extensive discussion or quantification of uncertainty may not be feasible or logical for inclusion of all prognostic papers. This may especially be the case where follow-up papers would allow for better communication and research into model uncertainty. However in it's nature, all prognostic models will include multiple sources of uncertainty. At a minimum this should be acknowledged in the publication of prognostic models, regardless of depth of subsequent uncertainty analysis and quantification.

5.2.5. Application Limitations

No one-size-fits-all model has been developed, implying the existence of limitations in all prognostic systems. Industry's capability to perform prognostic modelling and distribute outputs to users is dependant on the availability of required data, skilled personnel, and computing infrastructure. Clear descriptions of model inputs, coupled with application limitations, are necessary to allow for transparency with industry so that tested results can be reproduced in practise (Sikorska et al., 2011). Whilst selection criteria that outline the advantages and disadvantages for generic models exist, newly developed or adapted models do not come with this understanding. Examples of application limitations recommended for inclusion are data processing limitations, robustness to noise and uncertainties, training requirements, prediction horizon, and computation limitations. Many of these items can be covered under model selection justification and performance evaluation. However, model limitations should be stated concisely and challenges for their practical implementation discussed.

5.3. Distribution

Journal articles are the most widely used avenue for the presentation of scientific results, but generally, journals bear the responsibility of only reporting scientific findings. For a study to be classified as reproducible, as defined within this study, the data and code ought to be submitted along with the documentation of the study. Three on-line avenues for the dissemination of this information are as follows: 1) hosted by the publishing journal, 2) uploaded to a secure centralised or private database, and 3) available on request. The information location should be included within the publication.

Many high-impact journals now require articles to submit data, code, and documentation along with reports. The Transparency and Openness Promotion, consisting of signatories from 538 journals and 58 organisations (full list available at https://cos.io/top/), has produced guidelines for journals to translate scientific openness values into concrete actions (Nosek et al., 2015). The guidelines consist of eight standards with four levels of journal commitment. These guidelines, developed primarily by researchers from social and behavioural sciences, aim to be generic across all disciplines. Their relevance to the work of PHM researchers depends on groups such as the PHM Society and relevant journals.

Should publishing journals not support the hosting of data, code, and further documentation, on-line databases to promote open research have been developed, such as The Open Science Frame Work, a free cloud-based server initiated by the Center for Open Science (Center for Open Research, 2016). Further, The University of Western Australia has produced a prognostics-specific data management system to pool data sets from academia and industry (Sikorska et al., 2016). Authors also may upload information onto private or institution-based servers.

Should authors not want to pursue any of the aforementioned routes, all information should be made available upon request. Although this is the least transparent and most prohibitive option due to increased communication boundaries, it still satisfies our reproducibility recommendations.

5.4. Towards Reproducible PHM Research

Of the 50 data-driven prognostic papers reviewed, only 8 were based on industrial data. Most prognostic work is developed in a controlled experimental setting or with data derived from simulation. Note that when industrial data sets are available, they are often used by multiple researchers, which has the benefit of allowing models to be compared. An example is the condition and event data from centrifugal pumps at Irving Pulp and Paper originally published by Sundin, Eng, Montgomery, and Jardine (2007) and subsequently used in Heng et al. (2009) and Tian, Wong, and Safaei (2010).

The intention of this study is to encourage conversation amongst the PHM community with regards to model reproducibility. In order to support developments, Figure 2 presents a flowchart of all the identified criteria for reproducible PHM research. Moving forward, the PHM community may learn from other disciplines addressing reproducibility by:

- *Encouraging reproduction studies.* This may add validation toward successful and well-researched studies and will add an extra layer of scrutiny to the research.
- *Changing journal requirements.* Journals may state as a requirement of publication that data, code, and associated documentation be provided alongside the findings.

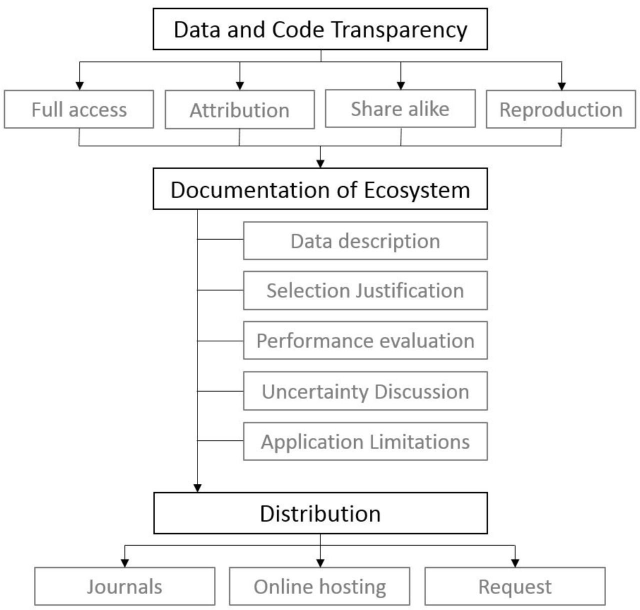


Figure 2. Flowchart of three identified reliability criteria with requisite steps for each stage.

6. CONCLUSION AND FUTURE WORK

This paper presents a general outline of reproducibility recommendations for PHM research. Suggestions for future practice have been adapted from the ongoing research of reproducibility across multiple scientific disciplines, with three main recommendations: data and code transparency, documentation, and distribution. To create a methodology that incorporates prognostic-specific elements, this paper defined the "modelling ecosystem", consisting of data description, model justification, performance evaluation, uncertainty quantification, and application limitations. Description of the modelling ecosystem inside of the documentation stage aims to further enable the reproducibility of prognostic models. Finally, suggestions for further promoting reproducibility and independent evaluation amongst the PHM community were made.

Avenues of future work for this project include:

- Widen the review of papers to include data-driven, model-based and hybrid models from other significant journals.
- Create a comprehensive check-list for each ecosystem element so that model developers may efficiently include ecosystem discussion.
- Present an original case study, with examples of good practice of reproducibility.
- Contact reviewed authors to assess their readiness to share data and code.
- Reproduce past models with the inclusion of ecosystem element documentation.

ACKNOWLEDGMENTS

The authors wish to acknowledge funding from the Australian ARC Industrial Transformation Research Hub for Offshore Floating Facilities (IH140100012) and the over 150 members of the CALCE Consortium for their long-term funding of research into systems reliability and sustainment.

REFERENCES

Aizpurua, J. I., & Catterson, V. M. (2015). Towards a methodology for design of prognostic systems. In *The Annual Conference of the PHM Society.* Coronado, CA, USA.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533.

Bissel, M. (2013). Reproducibility: The risks of the replication drive. *Nature Comment*, *503*(7476), 333–334.

Casadevall, A., & Fang, F. C. (2010). Reproducible science. *Infection and immunity*, *78*(12), 4972–4975.

Center for Open Research. (2016). *Open science framework.* Retrieved 2016-06-03, from `https://osf.io/`

Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A. M., ... Oliver, S. (2014). How to increase value and reduce waste when research priorities are set. *The Lancet*, *383*(9912), 156–165.

Chan, A.-W., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gøtzsche, P. C., ... Van Der Worp, H. B. (2014). Increasing value and reducing waste: addressing inaccessible research. *The Lancet*, *383*(9913), 257–266.

Coble, J. B. (2010). Merging data sources to predict remaining useful life–an automated method to identify prognostic parameters.

Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biol*, *13*(6), e1002165.

Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., ... Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, *383*(9913), 267–276.

Gutzwiller, K. J., & Riffell, S. K. (2014). Rigor and transparency in statistical analyses can help to ensure valid research. *Landscape Ecology*, *29*(7), 1115–1122.

Heng, A., Tan, A. C., Mathew, J., Montgomery, N., Banjevic, D., & Jardine, A. K. (2009). Intelligent condition-based prediction of machinery reliability. *Mechanical Systems and Signal Processing*, *23*(5), 1600–1614.

Hodkiewicz, M., & Montgomery, N. (2014). Data fitness for purpose: assessing the quality of industrial data for use in mathematical models.

Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., ... Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*, *383*(9912), 166–175.

Kan, M. S., Tan, A. C., & Mathew, J. (2015). A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, *62*, 1–20.

Kelly, A. (2006). *Maintenance systems and documentation*. Butterworth-Heinemann.

Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F., Cuthill, I. C., Fry, D., ... Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PloS One*, *4*(11), e7824.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systemsreviews, methodology and applications. *Mechanical Systems and Signal Processing*, *42*(1), 314–334.

Leek, J. T., & Peng, R. D. (2015). Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences*, *112*(6), 1645–1646.

Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J., ... Glasziou, P. (2014). Biomedical research: increasing value, reducing waste. *The Lancet*, *383*(9912), 101–104.

McNutt, M. (2014). Reproducibility [editorial]. *Science*, *343*(6168), 229–229.

Nature. (2016). *Availability of data, material and methods.* Retrieved 2016-06-02, from `http://www.nature.com/authors/policies/availability.html`

Nature Editorial. (2014). Journals unite for reproducibility [editorial]. *Nature*, *515*(6210), 7.

Nosek, B., Alter, G., Banks, G., Borsboom, D., Bowman, S., Breckler, S., ... Christensen, G. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, *348*(6242), 1422–1425.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657–660.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Pecht, M. (2008). *Prognostics and health management of electronics*. Wiley.

Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics*, *10*(3), 405–408.

Peng, R. D. (2011). Reproducible research in computational science. *Science*, *334*(6060), 1226–1227.

Peng, R. D., Dominici, F., & Zeger, S. L. (2006). Reproducible epidemiologic research. *American Journal of Epidemiology*, *163*(9), 783–789.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews*, *10*(9), 712–712.

Riehle, D. (2007). The economic motivation of open source software: Stakeholder perspectives. *Computer*, *40*(4), 25–32.

Salman, R. A.-S., Beller, E., Kagan, J., Hemminki, E., Phillips, R. S., Savulescu, J., ... Chalmers, I. (2014). Increasing value and reducing waste in biomedical research regulation and management. *The Lancet*, *383*(9912), 176–185.

Sankararaman, S., Daigle, M. J., & Goebel, K. (2014). Uncertainty quantification in remaining useful life prediction using first-order reliability methods. *IEEE Transactions on Reliability*, *63*(2), 603-619.

Sankararaman, S., & Goebel, K. (2015). Uncertainty in prognostics and systems health management. *International Journal of Prognostics and Health Management*, *6*(10).

Sankararaman, S., Saxena, A., & Goebel, K. (2014). Are current prognostic performance evaluation practices sufficient and meaningful? In *The Annual Conference of the PHM Society.* Fort Worth, TX, USA.

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, *1*(1), 4–23.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *The International Conference on Prognostics and Health Management.* Denver, CO, USA.

Schwab, M., Karrenbach, M., & Claerbout, J. (2000). Making scientific computations reproducible. *Computing in Science & Engineering*, *2*(6), 61–67.

Sikorska, J., Hodkiewicz, M., DCruz, A., Astfalck, L., & Keating, A. (2016). A collaborative data library for testing prognostic models. In *The Third European Conference of the PHM Society.* Bilbao, Spain.

Sikorska, J., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, *25*(5), 1803–1836.

Sonnenburg, S., Braun, M. L., Ong, C. S., Bengio, S., Bottou, L., Holmes, G., … Williamson, R. (2007, 10). The need for open source software in machine learning. *Journal of Machine Learning Research*, *8*, 2443-2466.

Steckler, T. (2015). Preclinical data reproducibility for r&dthe challenge for neuroscience [editorial]. *Psychopharmacology*, *232*(2), 317–320.

Sun, J., Zuo, H., Wang, W., & Pecht, M. G. (2012). Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. *Mechanical Systems and Signal Processing*, *28*, 585–596.

Sundin, P., Eng, P., Montgomery, N., & Jardine, A. (2007). Pulp mill on-site implementation of cbm decision support software. In *ICOMS Asset Management Conference.* Melbourne, Australia.

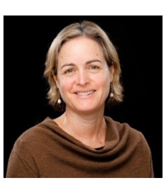
Tian, Z., Wong, L., & Safaei, N. (2010). A neural network approach for remaining useful life prediction utilizing both failure and suspension histories. *Mechanical Systems and Signal Processing*, *24*(5), 1542–1555.

US National Institutes of Health. (2014). *Principles and guidelines for reporting preclinical research.* Retrieved 2016-06-02, from `https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research`

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems*.

BIOGRAPHIES

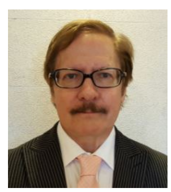
Lachlan Astfalck is a PhD candidate from the University of Western Australia working in the System Health Laboratory. He recieved his B.Eng (1st class Hons) degree from The University of Western Australia in 2014 and graduated as valedictorian of his class. In 2014 he was awarded the Chevron Chair prize in Gas Process Engineering, and in 2012 the Convocation of UWA Graduates Undergraduate Prize. His current research focuses on the industrial implementation of prognostic systems, with particular focus on remote systems.



Melinda R. Hodkiewicz is the BHP Billiton Fellow for Engineering for Remote Operations at the University of Western Australia (UWA). She has a BA (Hons) in Metallurgy and Science of Materials from Oxford University in 1985, and a Ph.D. in Mechanical Engineering from the University of Western Australia in 2004. Prior to her PhD, she worked in industry in operations and maintenance roles. She now leads the System Health Laboratory at UWA and works in the areas of asset health, maintenance and safety. She is a Chartered Engineer, a Member of the Institute of Materials, Minerals and Mining (IOM3) and the Asset Management Council.

Adrian Keating (M90-SM07) received his B.E. (Hon) and Ph.D. (Photonics) degrees in electrical and electronic engineering from the University of Melbourne, Australia in 1990 and 1995, respectively. Since 1996 he has worked at NTT Research Labs (Musashino-shi, Japan), the University of California, Santa Barbara, and at Calient Networks as the Fiber Optics Technology Manager. He joined the School of Electrical, Electronic and Computer Engineering at the University of Western Australia (UWA) in 2004 and later the School of Mechanical Engineering where he is currently an Associate Professor. His current research activities are in infrared optics sensors, sensing systems, optical microelectro-mechanical systems (MEMS), and porous silicon-based sensor technologies.

Edward Cripps is a Senior Lecturer at the School of Mathematics and Statistics, University of Western Australia (UWA). He obtained a BSc (Hons) and a PhD in Statistics, from the University of New South Wales, Australia, in 2000 and 2005, respectively. He joined UWA in 2009. His current research activities focus on Bayesian mixture models, predictive analytics, and asset management.



Michael G. Pecht holds an MS in Electrical Engineering, and an MS and PhD in Engineering Mechanics from the University of Wisconsin at Madison. He is a Professional Engineer, an IEEE Fellow, an ASME Fellow, an SAE Fellow and an IMAPS Fellow. He has previously received the European Micro and Nano-Reliability Award for outstanding contributions to reliability research, 3M Research Award for electronics packaging, and the IMAPS William D. Ashman Memorial Achievement Award for his contributions in electronics reliability analysis. He served as chief editor of the IEEE Transactions on Reliability for eight years and on the advisory board of IEEE Spectrum. He is chief editor for Microelectronics Reliability and an associate editor for the IEEE Transactions on Components and Packaging Technology. Professor Michael G. Pecht is the founder and Director of CALCE (Center for Advanced Life Cycle Engineering) at the University of Maryland, which is funded by over 150 of the worlds leading electronics companies at more than US$ 6M/year. He is also a Chair Professor in Mechanical Engineering and a Professor in Applied Mathematics at the University of Maryland. He has written more than 20 books on product reliability, development, use and supply chain management and over 400 technical articles. In 2008, he was awarded the highest reliability honor, the IEEE Reliability Society's Lifetime Achievement Award. In 2010, he received the IEEE Exceptional Technical Achievement Award for his reliability contributions in the area of prognostics and systems health management