Proceedings of the 7th European Conference of the Prognostics and Health Management Society 2022 - ISBN – 978-1-936263-36-3

On the Integration of Fundamental Knowledge about Degradation Processes into Data-Driven Diagnostics and Prognostics Using Theory-Guided Data Science

Simon Hagmeyer[1], Peter Zeiler[1], and Marco F. Huber[2,3]

[1] *Esslingen University of Applied Sciences, Goeppingen, 73037, Germany*
simon.hagmeyer@hs-esslingen.de
peter.zeiler@hs-esslingen.de

[2] *Institute of Industrial Manufacturing and Management IFF, University of Stuttgart, 70569 Stuttgart, Germany*

[3] *Fraunhofer Institute of Manufacturing Engineering and Automation IPA, 70569 Stuttgart, Germany*
marco.huber@ieee.org

Abstract

In Prognostics and Health Management, there are three main approaches for implementing diagnostic and prognostic applications. These approaches are data-driven methods, physical model-based methods, and combinations of them, in the form of hybrid methods. Each of them has specific advantages but also limitations for their purposeful implementation. In the case of data-driven methods, one of the main limitations is the availability of sufficient training data that adequately cover the relevant state space. For model-based methods, on the other hand, it is often the case that the degradation process of the considered technical system is of significant complexity. In such a scenario physics-based modeling requires great effort or is not possible at all. Combinations of data-driven and model-based approaches in form of hybrid approaches offer the possibility to partially mitigate the shortcomings of the other two approaches, however, require a sufficiently detailed data-driven and physics-based model.

This paper addresses the transitional field between data-driven and hybrid approaches. Despite the issues of formulating a physics-based model that provides a representation of the degradation process, basic knowledge of the considered system and of the laws governing its degradation process is usually available. Integration of such knowledge into a machine learning process is part of a research field that is either called theory-guided data science, (physics) informed machine learning, physics-based learning or physics guided machine learning. First, the state of research in Prognostics and Health Management on methods of this field is presented and existing research gaps are outlined. Then, a concept is introduced for incorporating fundamental knowledge, such as monotonicity constraints, into data-driven diagnostic and prognostic applications using approaches from theory-guided data science. A special aspect of this concept is its cross-application usability through the consideration of knowledge that repeatedly occurs in diagnostics and prognostics. This is, for example, knowledge about physically justified boundaries whose compliance makes a prediction of the data-driven model plausible in the first place.

1. Introduction

The choice between a model-based or a data-driven approach is a crucial element of any Prognostics and Health Management (PHM) application. Whether, for example, in the case of condition diagnosis or subsequent prediction of remaining useful life (RUL), the suitability of the respective approach depends on the properties of the particular application. The central prerequisite for a model-based approach is that knowledge on causal relationships of the technical system and its degradation process is available for the formation of a physics-based model. The model-based approach is often characterized by a rather high predictive accuracy and a comparably small amount of required data. However, the utilization of this approach is severely limited by the fact that the degradation processes of many technical systems are of such high complexity that a detailed, purely physics-based modeling is hardly possible (Eker et al., 2016). In addition, such physics-based models are also highly application-specific and therefore have restricted transferability (Byington et al., 2002).

The counterpart to the model-based approaches are the data-driven ones. These originate primarily from the domains of statistics and machine learning. Their implementation requires comparatively small effort; and at least the fundamental learning algorithm has a wide range of applicability (Eker et al., 2016). The methods are based on inductive inference, which underlies the statistical modeling of the training data provided (Huellermeier & Waegeman, 2021). The causal relationships, however, that yield the values of the training data are not learned. Since the data are the only source of information for these methods, they are not suitable for extrapolation into areas with sparse and, in particular, no training data. Accordingly, their purposeful use requires sufficient coverage of the state space by data (Coveney et al., 2016). Furthermore, the lack of comprehension of the causal relationships means that the predictions can take on implausible values that violate fundamental constraints (von Rueden, Mayer, et al., 2021). A behavior that intensifies in areas without training data.

Even though data-driven methods are currently predominant in research on PHM, the lack of data is a major limitation to their widespread industrial application. This affects diagnosis as well as prognosis, which can be subdivided in accordance with Jia et al. (2018) into four tasks:

- **Fault detection:** Detect a fault state/anomaly of a technical system without knowing the root cause. This results in a binary classification problem with the states fault or no fault.
- **Diagnosis:** Assign one or more causes to a detected fault state.
- **Health assessment:** Assess the state of health or the current risk of failure of a system based on its current condition.
- **Prognosis:** Predict the future state of health or RUL.

Each of these tasks involves its own estimation process and is individually affected by the lack of data.

Having a representative data set containing several run-to-failure data sets for each fault mode in each system configuration typically corresponds to a practically impossible amount of effort due to the typical lifetime and variant diversity of many systems. Even the recording of one run-to-failure cycle can take several months or years (Hagmeyer et al., 2021; Hemmer et al., 2019; Pillai et al., 2016). Therefore, Chao et al. (2022) even state that the two aforementioned problems of incomplete physical models and the lack of representative data sets are among the main problems in RUL prediction.

The combination of data-driven and physics-based models is usually referred to as hybrid in PHM. In this context, the term hybrid has a wide range of definitions depending on the literature, as among others Javed et al. (2017), N.-H. Kim et al. (2016), and Liao & Koettig (2014) demonstrate. In this paper, only the combination of entire data-driven and physics-based models is referred to as a hybrid approach. These offer the possibility to mitigate the limitations of the two approaches described above, but require sufficiently detailed models of both types. In addition, in order to restrict the scope of the following investigations, only approaches in which the data-driven models and incorporated knowledge or physics-based models relate to the same PHM task will be considered. The wide range of approaches to joining models in which they complement each other, for example, by one model doing fault detection and the other doing cause assignment based on it, or by one model doing health estimation and the other describing the degradation progression, is out of the scope of the paper.

Fundamentally, the integration of knowledge into machine learning is a whole research area that has been experiencing a great growth especially in the last five years. Depending on the literature, this research field is referred to as

- *theory-guided data science* see (Karpatne et al., 2017),
- *(physics) informed machine learning* see (von Rueden, Mayer, et al., 2021), (Yucesan & Viana, 2020b),
- *physics-based learning* see (Liu & Goebel, 2018) or
- *physics guided machine learning* see (Rai & Sahu, 2020).

In the following, the term theory-guided data science (TGDS) is used, as Karpatne et al. (2017) were the first to introduce such a designation of the research field. The research area TGDS does not only address the integration of entire physics-based models to increase predictive accuracy in machine learning, but already starts with the integration of knowledge about single principles of the process to be modeled. Here, the term knowledge is used in accordance with von Rueden, Mayer, et al. (2021), in that knowledge is seen as "validated information about relations between entities in certain contexts" (von Rueden et al, 2021).

The topic of this paper, integrating basic knowledge that is not sufficient for holistic modeling, lies in the transition area between data-driven and hybrid. Such basic knowledge already begins with the fact that most technical systems are not capable of self-healing and consequently a predicted degradation curve has to show a monotone increase. However, integration of such knowledge is an aspect that has received comparatively little attention in PHM so far. Although individual approaches have been used in case studies, any overall consideration of their use in PHM is missing. Studies in general, as well as those related to PHM described in the next section, nevertheless already demonstrate the potential of combining data and knowledge for increasing predictive accuracy. Thus, for example, insufficient amounts of data could be compensated. The term predictive accuracy is dependent on the respective PHM task and is evaluated by different metrics, some of which are subject-specific. Typical examples of these metrics are for fault detection *fault detection rate*, for diagnosis *isolation classification rate*, for health assessment *root mean* *squared error*, and for prognosis *prognostic horizon* (Saxena et al., 2010; Feldman et al., 2010; Gao et al., 2019).

The purpose of this paper is to take a first step towards a general examination of the use of TGDS in PHM as well as to initiate new research. Therefore, in Section 2 an overview of relevant TGDS approaches that do not require complete physics-based modeling and their employment in PHM is given. Then, in Section 3, concepts of assigning knowledge that occurs across diagnostic and prognostic applications to suitable TGDS methods are introduced. In the last section a conclusion and outlook on future work is given. The overall relevance of the paper's topic for PHM research stems firstly from the fact that the scenario of insufficient training data combined with incomplete physics-based modeling is common for industrial applications of PHM and secondly, that studies already show the potential of TGDS in such scenarios.

2. Overview of Approaches for Integrating Knowledge into Machine Learning

Already in regular machine learning, knowledge is partially integrated at several places of the learning pipeline. This includes, for example, feature engineering or the selection of the hypothesis set by defining hyperparameter values. TGDS extends the usual building of data-driven models by considering knowledge as a second source of information besides the data (von Rueden, Mayer, et al., 2021). In this paper, methods that do not require entire physics-based models for fusing knowledge and machine learning are considered. This involves knowledge about partial facets of the learning task, such as a subdivision of a problem into subproblems based on physics or knowledge about regularities such as valid bounds of variables, monotonicity conditions, correlations or curve shapes of intermediate and target variables. This form of knowledge integration is characterized by utilizing knowledge about intermediate variables or about valid properties of the target variables. The feature that distinguishes an incorporation of physics-based models from this is that a complete model provides a sufficiently precise estimate of the concrete value of the target variable(s). Thus, the data-driven and the physics-based models provide basically the same kind of information about the target variable, such as the health index (HI) or the RUL information. It is only through this uniformity that the method spectrum of hybrid model ensembles becomes possible.

Literature reviews of TGDS methods already exist, but these are independent of PHM and do not distinguish whether the formation of an entire physics-based model is required, which is highly relevant for PHM due to the complexity of many degradation processes. These PHM-independent works perform a mutually differing distinction of TGDS methods, as shown for example by von Rueden, Mayer, et al. (2021), Aykol et al. (2021), Willard et al. (2020), Karpatne et al. (2017), and Rai & Sahu (2020). In the following, six approaches are presented that allow the integration of knowledge that does not allow complete modeling. Furthermore, references to the already existing implementations of these approaches in PHM are given.

2.1. Physics-Based Generation of Synthetic Training Data

This method is the most intuitive form of knowledge integration. Here, the available amount of training data is extended by synthetic data points generated on the basis of knowledge. However, the labeling of such data points requires concrete values of the target variable and thus actually a process model. This issue is solved by drawing random samples from the entire range of values of the target variable that are considered valid based on knowledge. This could be data in which the values of the target variable comply with a given set of curves. Even more than when using a physics-based model for labeling, the deviation of the synthetic training data from the correct value is expected to have not only a high variance but also a high bias. Therefore, to improve the accuracy with the data, it is used in a pretraining for a physics-guided initialization instead of being mixed with the regular training data. In the pretraining, the model is trained on a rather simple problem. The actual training based on this, especially with small data sets, serves the subsequent fine-tuning of the machine learning model (Jia et al., 2019).

Several examples for the use of physics-based models to generate synthetic training data exist in PHM, such as Yu et al. (2018) and Sankararaman et al. (2011). However, most of these aim not to improve accuracy but to save computation time in the application phase by replacing the physics-based model with the data-driven one. The enrichment of the training data by knowledge that does not provide a complete modeling has hardly been investigated so far. The authors are so far only aware of Yucesan & Viana (2020a), Yucesan & Viana (2020b), and Dourado & Viana (2019) which apply such pretraining. Based on known correlations of input and target variables, these variables are brought in connection by a hyperplane. Since such linear equations do not correspond to the true hypersurface and in particular since weights are unknown in the equation, random initializations of the weights and thus of the plane are used for the generation of synthetic training data. These paper include just the application of physics-based generation of synthetic training data, but without any investigation on the effect of the pretraining.

2.2. Physics-Based Regularization

The training of a machine learning model is basically an optimization problem. The so-called loss function forms the objective function of the optimization, which evaluates the quality of a hypothesis. The goal of the training is to find a hypothesis that minimizes the loss function. This optimization problem can be supplemented by physically based constraints in order to obtain a physically consistent hypothesis as training outcome. The main approach to this is the addition of a special regularization term to the loss function. In regular machine learning, the loss function $L(f)$ mostly consists of a component $\text{loss}(\hat{Y}, Y)$ that captures the agreement of the model output $\hat{Y}$ and the real measured values Y, and a regularization component for constraining the model complexity $R(f)$

$$L(f) = \text{loss}\big(\hat{Y}, Y\big) + \lambda \cdot R\big(f\big). \tag{1}$$

In physics-based regularization, the loss function is extended by the term $\text{loss}_{phys}(\hat{Y})$, which evaluates whether it satisfies governing physics laws

$$L(f) = \text{loss}\big(\hat{Y}, Y\big) + \lambda \cdot R\big(f\big) + \gamma \cdot \text{loss}_{phys}\big(\hat{Y}\big). \tag{2}$$

Noncompliance with laws is penalized by an increased loss value, which is why, depending on the weighting γ, physically consistent solutions are favored by the training. Since the $\text{loss}_{phys}(\hat{Y})$ is independent of actual measured values, the evaluation of physical conformance is not bound to areas present in the collected data (Muralidhar et al., 2018), (Y. Zhu et al., 2019), (von Rueden, Mayer, et al., 2021). For instance, Muralidhar et al. (2018) introduce equations to embed valid ranges of values by means of rectified linear functions and monotonicity constraints by means of logic operations into the loss function as regularization.

Although this is a relevant approach, a work on the implementation of physics-based regularization in PHM is not known to the authors. However, there is an approach in diagnostic and prognostic applications that can be argued in a wider perspective also as an integration of knowledge and fundamentally shares the same concept. Instead of physics-based knowledge about the degradation process, operational knowledge is incorporated into the loss function. For this purpose, in the case of a regression task instead of a symmetric function such as the squared error an asymmetric function is used for $\text{loss}(\hat{Y}, Y)$. Applied to a RUL prediction, the asymmetric function represents the different costs that arise due to excessive maintenance in the case of RUL underestimation and due to unplanned outages in the case of RUL overestimation. Depending on the application, such a model can be trimmed more towards RUL underestimation or overestimation. The use of such asymmetric loss functions is discussed in the evaluation of several data challenges of the PHM Society as well as by Hoenig et al. (2019), Li et al. (2018), and Saxena et al. (2008).

2.3. Final Hypothesis Set Evaluation

A sufficient generalization of a machine learning model cannot be automatically guaranteed after training. In order to validate training results, extensive test data is usually required, which is specifically retained from the training. When generating several different models through training, this set of final solution hypotheses cannot only be evaluated using test data, but can also be compared to existing knowledge (von Rueden, Wirtz, et al., 2021). For the evaluation and selection of trained models, both the compliance with individual physics laws and the compliance with phyiscal models can be considered. Even though there is no direct integration of knowledge into the learning process in the final hypothesis set evaluation, the method is still included in the list here because the selection process can lead to better model accuracy in the application phase.

The final hypothesis set evaluation is an intuitive approach that is certainly used regularly in PHM in a basic form. In addition, there are comparable approaches and objectives in explainable machine learning. Based on knowledge, the trained models are analyzed in the so-called post-hoc explanation and assessed with respect to their validity (Burkart & Huber, 2021). One application of the approach is presented by Grezmak et al. (2019). They show that in a learned model for gearbox diagnosis, the damage frequencies which are most relevant for classification are consistent with knowledge of sideband frequencies.

2.4. Intermediate Physical Variables

The basic idea of this approach is to adapt the hypothesis space by dividing the problem of modeling the relationship between input and target variables into modules based on process knowledge. The inputs and outputs of the modules are thus assigned a physical meaning and, as far as possible, they are related to each other on the basis of knowledge. Thereby on the one hand the problem structure can be considered within the architecture of a single data-driven model, e.g. by adapting the architecture of a neural network and assigning meanings to neurons. On the other hand, for each defined modul an indiudual data-driven model can also be used (Karpatne et al., 2017), (Willard et al., 2020). If at least a modul can be modeled physics-based in sufficient detail, it is also possible to substitute the respective data-driven model by it. Besides intermediate physical variables, this approach can also be designated for instance as physics-guided architecture or as theory-guided design of model architecture.

This physics-based problem subdivision thus also bridges the gap to knowledge-based feature engineering by in both cases providing information on individual intermediate variables related to the target variable. The goal of this approach is the physics-based subdivision of a problem. However, the ability to incorporate a physics-based model of a subproblem also bridges another gap. This is to the, in the first section excluded hybrid approaches where data-driven and physics-based models are used for different PHM tasks. One such example is the state estimation and the prediction of further degradation using respectively one of the model types.

In PHM, a physics-based problem subdivision is applied several times, particularly noteworthy here are the same papers as mentioned in physics-based generation of synthetic training data. The idea of incorporating knowledge about the structure of a problem, which is not sufficient for complete modeling, into a data-driven model is applied by Yucesan & Viana (2020a), Yucesan & Viana (2020b), and Dourado & Viana (2019) to the examples of bearing damage in wind turbines and corrosion-influenced material fatigue of aircraft components using recurrent neural networks.

2.5. Auxiliary Task in Multi-Task Learning

Another possibility for the integration of knowledge mentioned by Willard et al. (2020) is the use of multi-task learning. In addition to the actual prediction task, auxiliary tasks are used to estimate related physical variables. These auxilary tasks are defined based on knowledge of the process and admissible properties of these variables. The unification of both tasks by multi-task learning is intended to leverage their synergy for a more precise as well as physically consistent prediction. It should also be emphasized that the physics-based regularization and auxiliary task in multi-task learning approaches have considerable commonalities. Both shift the position of the optimum, which is searched for during the training process, towards models, which comply with given knowledge. Nevertheless, there is also an affiliation of this approach to intermediate physical variables. The hypothesis set is adjusted by linking related physical variables to the target variable on the basis of knowledge.

In PHM, especially Ozdagli & Koutsoukos (2021) address the use of knowledge about related variables in the context of multi-task learning. The method of employing knowledge-based auxiliary tasks is applied to damage detection in structural health monitoring using neural networks. The labels for the auxillary tasks are provided in this case by a physics-based model, which, nevertheless, is not fundamentally required for the approach. Compared to the baseline of a purely data-driven neural network, a significant improvement of the classification accuracy is shown. In addition to incorporating knowledge, another advantage of the multi-task approach is the possibility to use labeled data of the additional target variables for training in order to obtain enhanced learning results also for the actual target variable (Caruana, 1997). Examples of such work in PHM include T. S. Kim & Sohn (2020), Chen et al. (2019), and Hinchi & Tkiouat (2018). One aspect that is entirely absent in these studies is having knowledge about admissible properties for the related variables and the incorporation of this knowledge into the learning process.

2.6. Knowledge Integration into Probabilistic Graphical Models

The probabilistic graphical models are particularly suitable for the integration of knowledge due to their inherent interpretability. Based on knowledge, nodes and edges can be parameterized, e.g. by specifying an adjacency matrix. As with the multi-task approach, probabilistic graphical models are considered here as a separate case, wherein the integration of knowledge in probabilistic graphical models has already received extensive consideration both in general and in PHM in particular. Depending on the further learning process, this can be seen as an architectural constraint adjusting the hypothesis space in the sense of intermediate variables. If the parameterization of edges represents a priori information that is adapted during training, the learning process is rather guided in one direction in the sense of a regularization. Such ambiguity is also reflected in the different treatment of this approach in the review papers on TGDS mentioned at the beginning of the second section.

The ability to perform knowledge integration of these models is also reflected in the extensive work being done on this at PHM. Liu & Goebel (2018) present a research and development project of the US federal agency National Aeronautics and Space Administration. The goal here is to develop a predictive system that not only assesses the safety status of aircrafts, but of the entire airspace. As a central element of the information fusion, a Bayesian network is used. Juesas et al. (2016) in turn present the integration of imprecise state knowledge into an autoregressive hidden Markov model (ARHMM) using the CMAPSS dataset as a benchmark. The possibility to represent imprecise knowledge allows chosing a compromise between belief and evidence in model generation. Palazuelos et al. (2020) and González et al. (2019) present a graph network where nodes represent the state of system components. An adjacency matrix is used to define connections between nodes of physically related components. The matrix can be learned from data but also created or adapted based on knowledge.

3. Conception of a PHM Related Use of TGDS Methods

Despite the outlined potential of TGDS to improve data-driven diagnostic and prognostic applications, it is also apparent that there are still significant research gaps in this regard. As a first step towards a holistic treatment of the topic, the following sub-sections introduce concepts of assigning knowledge that occurs across diagnostic and prognostic applications to suitable TGDS methods presented in the previous section. The selection of cross-application knowledge is based on the authors' assessment and focuses on knowledge of the degradation process. In PHM, there are also other sources of recurring knowledge related to the degradation process, which are not considered here. Examples of this include knowledge due to a previous risk assessment such as an FMEA or knowledge about operating conditions.

The basic assumption is that a larger amount of integrated information, whether in the form of knowledge or data, is generally associated with an improvement in the predictive accuracy of a diagnostic or prognostic application. Another assumption is that, although limited in volume, labeled data of the examined degradation process for supervised learning are available in the first place.

In supervised learning, models are trained to reflect the relationship between input and target variables. The structure of the learned model or its information processing to form the estimate of the target variable is not bound to the cause-effect relationships of the modeled process. Consequently, from the authors' point of view, an essential characteristic of knowledge of the modeled process is whether it relates to the target variable that is always present or only to an intermediate variable associated with the target variable that is not inherently included in the model. Hence the following subdivision is provided:

- Concepts for the integration of cross-application knowledge on target variables
- Concepts for the integration of cross-application knowledge on intermediate variables

3.1. Concepts for the Integration of Cross-Application Knowledge on Target Variables

The three TGDS methods that specifically require and incorporate knowledge of the target variable are physics-based generation of synthetic training data (Section 2.1), physics-based regularization (Section 2.2), and final hypothesis set evaluation (Section 2.3).

Knowledge on the curve shape of the degradation process: If the fault mechanism is the same, the shape of the health progression is often identical across applications and therefore known. For example, the fault mechanism determines whether a system is capable of self-healing and thus whether a positive HI gradient is admissible. If this is not the case, a monotone damage progression must be observed. Further examples are the typical convex curves of crack propagation under cyclic loading (Castillo et al., 2010) and in the case of filter clogging the differential pressure increase (Thomas et al., 2001). The latter additionally becoming a linear increase when depth filtration transitions to cake filtration. Even though the level of degradation over time can only be described very imprecisely, there is nevertheless knowledge of shape constraints that should be fundamentally fulfilled in a prediction. Mathematical shape constraints can be well expressed by formulas, which is why physics-based regularization is particularly suitable. Physics-based regularization provides the ability to guide the training into such a direction that the constraints are met, also for high-dimensional problems. By relying on formulas to express shape constraints, one can also use them to evaluate the final hypothesis set. Although compliance is not enforced in training, it does have an advantage of general applicability especially when considering different types of machine learning methods that involve different loss and training functions.

Knowledge of correlations: If there is no knowledge on strict shape constraints but only on correlations between input variables and target variables, which are not universally met, the physics-based generation of synthetic training data approach is most suitable. The reason for this is that local areas where the synthetic data show a significant bias compared to the actual data can still be adjusted following the pretraining. Other approaches instead would likely result in rather soft constraints or a flawed model as training result with such inaccurate knowledge. As Yucesan & Viana (2020a), Yucesan & Viana (2020b), and Dourado & Viana (2019) demonstrate, engineering estimates on correlations are sufficient to obtain a reasonable initialization of an iteratively trained model by means of a pretraining. Furthermore, for example by Lauer & Bloch (2008) different approaches are presented for also incorporating synthetic training data with different quality than the actual training data in support vector machines with their convex training tasks.

Non-formalized expert knowledge (tacit knowledge): Often, knowledge about degradation processes is available in the form of expert knowledge that is difficult to express in mathematically precise terms. Especially in such cases final hypothesis set evaluation is well suited, since no formulation is required and basic physical correctness of the learned model can be ensured. However, the knowledge-based analysis of trained models is closely related to the topic and problems of interpretable machine learning. The main issue here is the mostly abstract, high-dimensional representation of learned results that are beyond human cognitive comprehension. So, the application of the final hypothesis set evaluation approach requires that the models to be evaluated are intrinsically interpretable or that post-hoc explanations can be applied. Post-hoc explanations require that a low-dimensional and to some extent local representation of the learned behavior can be created without too much loss of information, which can for example be visually perceived. Thereby, significant research gaps in PHM on interpretable machine learning especially on models of time series analysis and prediction applications in general exist (Vollert et al., 2021).

Boundaries of target variables: If, the boundaries of the target variable's permissible range are known, several methods are suitable for ensuring compliance with these boundaries and thus, the basic validity of an estimate. In addition to physically induced boundaries, external requirements, such as a maximum service life of a component, can also yield

Table 1. Summary of recommended methods for integrating knowledge that occurs across applications.

Type of knowledge	Proposed approach for knowledge integration
Knowledge on the curve shape of the degradation process	Physics-Based Regularization (2.2) Final Hypothesis Set Evaluation (2.3)
Knowledge of correlations	Physics-Based Generation of Synthetic Training Data (2.1)
Non-formalized expert knowledge	Final Hypothesis Set Evaluation (2.3)
Boundaries of target variables	Physics-Based Regularization (2.2) Final Hypothesis Set Evaluation (2.3) Probabilistic Graphical Model (2.6) Intermediate Physical Variables (2.4)
Knowledge of the problem structure	Intermediate Physical Variables (2.4) Probabilistic Graphical Model (2.6)
Knowledge and extensive data of intermediate variables	Auxiliary Task in Multi-Task Learning (2.5) Intermediate Physical Variables (2.4) Probabilistic Graphical Model (2.6)

such boundaries. Besides physics-based regularization and final hypothesis set evaluation, the approaches of using graphical models and intermediate physical variables can also be utilized to enforce compliance with such boundaries. With graphical models, edges can be parameterized accordingly. In the case of the intermediate physical variables approach, a model structure that fundamentally ensures the compliance can be specified, in simple cases of constant boundaries already by the choice of an output function.

3.2. Concepts for the Integration of Cross-Application Knowledge on Intermediate Variables

The three methods that can be used also in case of knowledge of the problem structure and intermediate variables are intermediate physical variables (Section 2.4), auxiliary task in multi-task learning (Section 2.5), and probabilistic graphical models (Section 2.6). In the following, a distinction is made between only two cases.

Knowledge of the problem structure: If knowledge about the structure of a problem and about relevant intermediate variables is available, a physically based subdivision according to the approach intermediate physical variables can usually be applied. The same holds for graphical models whose nodes can be assigned a meaning and also edges can be specified accordingly between nodes. Especially if the data-driven estimation of intermediate variables is considered as a learning task on its own, the concepts described above for the integration of knowledge about target variables can be applied to this subproblem. That knowledge about a problem's structure and intermediate variables is often available is shown by the extensive work on hybrid methods where different model types take over individual subtasks (Eker et al., 2019). A further evidence is the physics-based feature engineering already mentioned in Section 2, where also extensive work is done, especially on rotating systems like rolling bearings or gears (J. Zhu et al., 2014).

Knowledge and extensive data of intermediate variables: Multi-task learning addresses among others the case when, in addition to knowledge of intermediate variables, extensive labeled data on these variables is also available. Instead of learning to assign known intermediate variables as a submodule or using probabilistic graphical models, multi-task learning can alternatively use them as additional target variables. Although there is still a considerable need for research on the integration of knowledge in multi-task learning, the approach of using additional labeled data that do not include the actual target variable already offers als great potential. In accordance with T. S. Kim & Sohn (2020), the estimation of the current HI can form an auxillary task in a prediction application. From the authors' point of view, this approach is of high relevance, since it allows data to be used for learning a prognosis model, which do not contain any health change and thus are of minor use in a regular prognosis application. In many applications with long test durations, such as ball bearings, tests with predamaged components on fixed fault conditions und therefore health are common practice (Chen et al., 2018). With multi-task learning and knowledge of such related variables, this kind of test data can also be used for prognosis development.

A summary of the concepts for assigning knowledge types and TGDS methods is given in Table 1.

4. Conclusions and Outlook

There are three approaches to realizing diagnostic and prognostic tasks. In this paper, at first, these approaches are characterized. Thereby, connections between hybrid methods and the research field of TGDS can be identified. Subsequently, main aspects of TGDS are introduced and the potential of TGDS in PHM is outlined. The focus here is on methods for the integration of knowledge in machine learning, which do not require complete physics-based models, but rather knowledge of individual properties of the degradation process. For this purpose, a definition for the designation model is given. The presented overview of the relevant TGDS approaches illustrates them in detail and also points out studies on PHM that already employ them. In doing so, several research gaps can be identified. Based on the overview, cross-application knowledge occurring in diagnostics and prognostics is stated and concepts for integrating it into the learning process are proposed. The description of suitable methods contained therein is based primarily on theoretical considerations and, where available, on transferable findings from other work.

Overall, the paper makes an initial contribution to a holistic investigation of the incorporation of knowledge into machine learning in diagnostics and prognostics. There is significant potential for the use of TGDS in PHM, but also a great need for further research. Concerning the latter, on the one hand, there is much more knowledge for which a procedure for the integration is of cross-application benefit. On the other hand, the given theoretical concepts have to be investigated more thoroughly, supplemented by PHM-specific aspects such as uncertainty considerations, and verified by empirical studies. In addition, overlapping research fields such as transfer learning and fuzzy machine learning have to be considered, where the integration of knowledge is also a partial aspect.

References

Aykol, M., Gopal, C. B., Anapolsky, A., Herring, P. K., van Vlijmen, B., Berliner, M. D., ... Storey, B. D. (2021). Perspective—combining physics and machine learning to predict battery lifetime. *Journal of The Electrochemical Society*, *168*(3). doi: 10.1149/1945-7111/abec55

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, *70*, 245–317. doi: 10.1613/jair.1.12228

Byington, C. S., Roemer, M. J., & Galie, T. (2002). Prognostic enhancements to diagnostic systems for improved condition - based maintenance. In *Proceedings of the IEEE aerospace conference.* Big Sky, MT, USA. doi: 10.1109/AERO.2002.1036120

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*(1), 41–75. doi: 10.1023/a:1007379606734

Castillo, C., Fernández-Canteli, A., Castillo, E., & Pinto, H. (2010). Building models for crack propagation under fatigue loads: application to macrocrack growth. *Fatigue & Fracture of Engineering Materials & Structures*, *33*(10), 619–632. doi: 10.1111/j.1460-2695.2010.01475.x

Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2022). Fusing physics-based and deep learning models for prognostics. *Reliability Engineering & System Safety*, *217*. doi: 10.1016/j.ress.2021.107961

Chen, Y., Peng, G., Xie, C., Zhang, W., Li, C., & Liu, S. (2018). ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. *Neurocomputing*, *294*, 61–71. doi: 10.1016/j.neucom.2018.03.014

Chen, Y., Zhang, C., Zhang, N., Chen, Y., & Wang, H. (2019). Multi-task learning and attention mechanism based long short-term memory for temperature prediction of EMU bearing. In *Proceedings of the prognostics and system health management conference (PHM-qingdao).* Qingdao, China: IEEE. doi: 10.1109/phm-qingdao46334.2019.8942914

Coveney, P. V., Dougherty, E. R., & Highfield, R. R. (2016). Big data need big theory too. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2080), 20160153. doi: 10.1098/rsta.2016.0153

Dourado, A., & Viana, F. A. C. (2019). Physics-informed neural networks for corrosion-fatigue prognosis. In *Proceedings of the annual conference of the phm society* (Vol. 11). Scottsdale, Arizona, USA.

Eker, O. F., Camci, F., & Jennions, I. K. (2016). Physics-based prognostic modelling of filter clogging phenomena. *Mechanical Systems and Signal Processing*, *75*, 395–412. doi: 10.1016/j.ymssp.2015.12.011

Eker, O. F., Camci, F., & Jennions, I. K. (2019). A new hybrid prognostic methodology. *International Journal of Prognostics and Health Management*, *11*(2).

Feldman, A., Kurtoglu, T., Narasimhan, S., Poll, S., Garcia, D., Kleer, J. D., ... Gemund, A. V. (2010). Empirical evaluation of diagnostic algorithm performance using a generic framework. *International Journal of Prognostics and Health Management*, *1*. doi: 10.36001/ijphm.2010.v1i1.1344

Gao, T., Lu, C., & Hao, M. (2019). Design requirements of PHM system fault diagnosis capability. In *Proceedings of the 2019 chinese automation congress (CAC).* Hangzhou, China: IEEE. doi: 10.1109/cac48633.2019.8996303

González, I., Cáceres, J., Droguett, E. L., & López-Campos, M. (2019). Graph convolutional networks for health state diagnostics. In *Proceedings of the 29th european safety and reliability conference* (p. 1208-1213). Hannover, Germany.

Grezmak, J., Wang, P., Sun, C., & Gao, R. X. (2019). Explainable convolutional neural network for gearbox fault diagnosis. *Procedia CIRP*, *80*, 476–481. doi: 10.1016/j.procir.2018.12.008

Hagmeyer, S., Mauthe, F., & Zeiler, P. (2021). Creation of publicly available data sets for prognostics and diagnostics addressing data scenarios relevant to industrial applications. *International Journal* *of Prognostics and Health Management*, *12*(2). doi: https://doi.org/10.36001/ijphm.2021.v12i2.3087

Hemmer, M., Klausen, A., van Khang, H., Robbersmyr, K. G., & Waag, T. I. (2019). Simulation-driven deep classification of bearing faults from rawvibration data. *International Journal of Prognostics and Health Management*.

Hinchi, A. Z., & Tkiouat, M. (2018). A multi-task deep learning model for rolling element-bearing diagnostics. In *Proceedings of the european conference of the phm society* (Vol. 4). Philadelphia, Pennsylvania, USA.

Hoenig, M., Hagmeyer, S., & Zeiler, P. (2019). Enhancing remaining useful lifetime prediction by an advanced ensemble method adapted to the specific characteristics of prognostics and health management. In *Proceedings of the 29th european safety and reliability conference* (pp. 1155–1162). Hannover, Germany.

Huellermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. doi: 10.1007/s10994-021-05946-3

Javed, K., Gouriveau, R., & Zerhouni, N. (2017). State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mechanical Systems and Signal Processing*, *94*, 214–236. doi: 10.1016/j.ymssp.2017.01.050

Jia, X., Huang, B., Feng, J., Cai, H., & Lee, J. (2018). Review of phm data competitions from 2008 to 2017: Methodologies and analytics. In *Proceedings of the annual conference of prognostics and health management society* (Vol. 10). Philadelphia, Pennsylvania, USA.

Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2019). Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. In *Proceedings of the 2019 SIAM international conference on data mining* (pp. 558–566). Calgary, Kanada. doi: 10.1137/1.9781611975673.63

Juesas, P., Ramasso, E., Drujont, S., & Placet, V. (2016). On partially supervised learning and inference in dynamic bayesiannetworks for prognostics with uncertain factual evidence:illustration with markov switching models. In *Third european conference of the phm society.* Bilbao, Spanien.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., … Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. doi: 10.1109/tkde.2017.2720168

Kim, N.-H., An, D., & Choi, J.-H. (2016). *Prognostics and health management of engineering systems: An introduction*. Springer International Publishing.

Kim, T. S., & Sohn, S. Y. (2020). Multitask learning for health condition identification and remaining useful life prediction: deep convolutional neural network approach. *Journal of Intelligent Manufacturing*, *32*(8), 2169–2179. doi: 10.1007/s10845-020-01630-w

Lauer, F., & Bloch, G. (2008). Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing*, *71*(7-9), 1578–1594. doi: 10.1016/j.neucom.2007.04.010

Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, *172*, 1–11. doi: 10.1016/j.ress.2017.11.021

Liao, L., & Koettig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability*, *63*(1), 191–207. doi: 10.1109/TR.2014.2299152

Liu, Y., & Goebel, K. (2018). Information fusion for national airspace system prognostics: A nasa uli project. In *Proceedings of the annual conference of the phm society* (Vol. 10). Philadelphia, Pennsylvania, USA.

Muralidhar, N., Islam, M. R., Marwah, M., Karpatne, A., & Ramakrishnan, N. (2018). Incorporating prior domain knowledge into deep neural networks. In *Proceedings of the IEEE international conference on big data (big data).* Seattle, WA, USA. doi: 10.1109/bigdata.2018.8621955

Ozdagli, A. I., & Koutsoukos, X. (2021). Model-based damage detection through physics guided learning. In *Proceedings of the proceedings of the annual conference of the PHM society* (Vol. 13). Online: PHM Society. doi: 10.36001/phmconf.2021.v13i1.3012

Palazuelos, A. R.-T., Droguett, E. L., & Groth, K. M. (2020). A system-level prognostics and health management framework based on graph convolutional neural networks. In *Proceedings of the 30th european safety and reliability conference and 15th probabilistic safety assessment and management conference.* Venice, Italy, (online). doi: 10.3850/978-981-14-8593-0_4165-cd

Pillai, P., Kaushik, A., Bhavikatti, S., Roy, A., & Kumar, V. (2016). A hybrid approach for fusing physics and data for failure prediction. *International Journal of Prognostics and Health Management*, *7*(4).

Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus. *IEEE Access*, *8*, 71050–71073. doi: 10.1109/access.2020.2987324

Sankararaman, S., Ling, Y., Shantz, C., & Mahadevan, S. (2011). Uncertainty quantification in fatigue crack growth prognosis. *International Journal of Prognostics and Health Management*, *2*(1).

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., & Schwabacher, M. (2008). Metrics for evaluating performance of prognostic techniques. In *Proceedings of the international conference on prognostics and health management.* Denver, CO, USA: IEEE. doi: 10.1109/phm.2008.4711436

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*. doi: 10.36001/ijphm.2010.v1i1.1336

Thomas, D., Penicot, P., Contal, P., Leclerc, D., & Vendel, J. (2001). Clogging of fibrous filters by solid aerosol particles experimental and modelling study. *Chemical Engineering Science*, *56*(11), 3549–3561. doi: 10.1016/s0009-2509(01)00041-0

Vollert, S., Atzmueller, M., & Theissler, A. (2021). Interpretable machine learning: A brief survey from the predictive maintenance perspective. In *Proceedings of the IEEE international conference on emerging technologies and factory automation (ETFA).* Vasteras, Sweden: IEEE. doi: 10.1109/etfa45728.2021.9613467

von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., … Schuecker, J. (2021). Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*. doi: 10.1109/tkde.2021.3079836

von Rueden, L., Wirtz, T., Hueger, F., Schneider, J. D., Piatkowski, N., & Bauckhage, C. (2021). Street-map based validation of semantic segmentation in autonomous driving. In *Proceedings of the international conference on pattern recognition (ICPR).* Milan, Italy: IEEE. doi: 10.1109/icpr48806.2021.9413292

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating scientific knowledge with machine learning for engineering and environmental systems. *arXiv: 2003.04919v6*.

Yu, Y., Yao, H., & Liu, Y. (2018). Physics-based learning for aircraft dynamics simulation. In *Proceedings of the annual conference of the phm society* (Vol. 10). Philadelphia, Pennsylvania, USA. doi: 10.36001/phmconf.2018.v10i1.513

Yucesan, Y. A., & Viana, F. (2020a). Hybrid model for wind turbine main bearing fatigue with uncertainty in grease observations. In *Proceedings of the annual conference of the PHM society* (Vol. 12). online. doi: 10.36001/phmconf.2020.v12i1.1139

Yucesan, Y. A., & Viana, F. A. C. (2020b). A physics-informed neural network for wind turbine mainbearing fatigue. *International Journal of Prognostics and Health Management*, *11*. doi: 10.36001/ijphm.2020.v11i1.2594

Zhu, J., Nostrand, T., Spiegel, C., & Morton, B. (2014). Survey of condition indicators for condition monitoring systems. In *Proceedings of the annual conference of prognostics and health management society* (Vol. 6). Fort Worth, Texas, USA.

Zhu, Y., Zabaras, N., Koutsourelakis, P.-S., & Perdikaris, P. (2019). Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *Journal of Computational Physics*, *394*, 56–81. doi: 10.1016/j.jcp.2019.05.024