

Received 9 March 2023, accepted 3 April 2023, date of publication 10 April 2023, date of current version 13 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265722

RESEARCH ARTICLE

A Comparative Study on Methods for Fusing Data-Driven and Physics-Based Models for Hybrid Remaining Useful Life Prediction of Air Filters

SIMON HAGMEYER¹ AND **PETER ZEILER**

Research Group for Reliability Engineering and Prognostics and Health Management, Faculty of Mechanical and Systems Engineering, Esslingen University of Applied Sciences, 73728 Esslingen, Germany

Corresponding author: Simon Hagemeyer (simon.hagemeyer@hs-esslingen.de)

This work was supported in part by the German Federal Ministry of Education and Research through the Program “Professorales Personal an Fachhochschulen” (Professorial Staff at Universities of Applied Sciences) under Grant 03FHP115; in part by the Baden-Württemberg Ministry of Science, Research and Culture, Germany; and in part by the Esslingen University of Applied Sciences in the Funding Program Open Access Publishing.

ABSTRACT Approaches for diagnosis and prognosis of the health of engineering systems are divided into data-driven, model-based, and hybrid methods. Data-driven methods depend on the availability of data. Model-based methods require knowledge of the degradation process. A great effort of data generation along with the high complexity of degradation processes often limits both approaches. To mitigate these limitations, the combination of data and knowledge through hybrid methods is examined in this paper. This approach is compared to the alternative approach of reducing the effort of generating training data, as both are gaining importance in diagnostics and prognostics. A new categorization of hybrid prognostic methods for combining data-driven and physics-based models is presented, along with references to existing realizations of these methods. Based on the categorization, a case study on the hybrid remaining useful life prediction of a filtration process is conducted. Several hybrid methods are implemented and tested in this study. Through the combination of models, an improvement in predictive accuracy is achieved. In addition, the paper examines systematic attributes of the individual hybrid methods. Statements on the influence of data scarcity on the predictive accuracy, data-driven models with high variance, and the computational efficiency of the hybrid methods are made. It is shown that these statements are supported by the case study’s results.

INDEX TERMS Data-driven methods, filtration, hybrid methods, model-based methods, physics-informed machine learning, prognostics and health management, PHM, remaining useful life prediction.

I. INTRODUCTION

A fundamental feature that characterizes the engineering discipline of prognostics and health management (PHM) is the assessment of degradation or health of the individual engineering system in use. By evaluating data, such as the system’s sensor readings or operating data, inferences are made about its condition. These inferences can be divided into the following four main tasks: fault detection, diagnosis of the fault cause, health assessment, and prediction of the

future health or remaining useful life (RUL) [1]. The purpose of these tasks is to provide condition information on the engineering system for its health management.

In general, approaches to these diagnostic and prognostic tasks are categorized as data-driven, model-based, or hybrid. In the literature on PHM, data-driven methods are currently the predominant methods addressed. They stem from the fundamental research areas of statistics and machine learning. Examples of data-driven methods that are commonly employed for diagnostics and prognostics include artificial neural networks, Gaussian processes, and Wiener processes [2]. Major advantages of data-driven methods

The associate editor coordinating the review of this manuscript and approving it for publication was Yongquan Sun¹.

compared to model-based and hybrid methods are their comparatively small implementation effort, the reduced amount of knowledge required about the system as well as its degradation process, and their rather wide range of applicability [3].

Data-driven methods are based on inductive inference, which underlies the statistical modeling of the training data provided [4]. This results in the causal relationships that constitute the characteristics of the training data not being learned. The general training objective of a data-driven method is to achieve good generalization, especially in areas of sufficient data density. However, since they lack understanding of the cause-effect relationships, they are not suitable for making predictions in areas of the state space with little or no training data. Accordingly, their purposeful use requires sufficient coverage of the relevant state space [5]. A fundamental problem in industrial applications of diagnostics and prognostics, as well as in the overarching field of reliability engineering in general, is the availability of a sufficient amount of data. A long service life as well as high investment and operating costs are reasons why the availability of a sufficient amount of run-to-failure data sets for each fault mode often represents a high or unattainable challenge.

If there are not enough training data available, a likely consequence is that data-driven methods do not achieve the required predictive accuracy of the respective modeling task, despite intensive optimization attempts. Thereby, the predictive accuracy can be evaluated according to different metrics depending on the diagnostic or prognostic task, such as for diagnosis by the *fault isolation rate* and for RUL prediction by the *mean squared error* or *prognostic horizon* [6], [7]. In the case of insufficient predictive accuracy, two solutions are available. On the one hand, according to the general machine learning principle “More data beats a cleverer algorithm” [8], a solution may be to directly address their deficiency by generating additional training data. Yet this is bound by the often impossibly high effort described previously. On the other hand, accuracy can also be increased by including knowledge about the causal relationships of the system and its degradation process. This knowledge enriches the data-driven model (DM), providing an additional source of information besides the data themselves. In addition to the poor data availability, however, an equally relevant challenge for diagnostics and prognostics in industrial applications is that the degradation processes of many engineering systems are exceptionally complex. This is the reason why a highly precise physics-based model (PM) of the cause-effect relationships is also hardly attainable [9]. Nevertheless, basic knowledge about the system under consideration and its degradation process is usually available, which allows the specification of individual boundary conditions or a coarse physics-based modeling.

This paper addresses the case where both approaches, data-driven and model-based, are insufficient on their own due to the limitations discussed previously. The aim, therefore, is to

contribute to the research on hybrid methods for combining both types of models, with a focus on the task of RUL prediction. In doing so, the paper targets the following innovations compared to the state of the art:

- A structuring of hybrid methods for combining DMs and PMs sharing the same target variable.
- Conducting a case study on RUL prediction of air filters that assesses four hybrid methods.
- Conclusions about the attributes of specific hybrid methods are drawn and empirically underlined on the basis of the case study.

The structure of this paper is as follows: Section II introduces approaches for improving the predictive accuracy of data-driven methods, which include approaches that reduce the effort required to generate additional training data and approaches that integrate knowledge. As a specific approach for combining data-driven methods and knowledge, the fusion of DMs and PMs is examined in Section III. Different types of hybrid methods are described in detail and references are given to their use in diagnostics and prognostics. This is followed by a detailed case study on RUL prediction using hybrid methods. For this, Section IV first introduces the filter loading process on which the case study is based. Then, the applied PM, DM, and hybrid methods are described. At the end of Section IV, the results of the case study are analyzed. Building on this, Section V evaluates the individual hybrid methods with respect to systematic attributes that are relevant for prognostics. In Section VI, conclusions on this paper are drawn.

II. APPROACHES FOR IMPROVING DATA-DRIVEN PREDICTIONS

As argued above, the scarcity of training data is a major challenge in diagnostics and prognostics. For this reason, research on reducing the effort of generating training data as well as on fusing machine learning with knowledge is gaining significant attention. Several approaches for effort reduction are described in Section II-A. Since these are particularly induced by deep learning (DL), it is described at the beginning of this section. Section II-B introduces the emerging research segment of machine learning, which focuses on knowledge as a second source of information in addition to data.

A. REDUCING THE EFFORT OF GENERATING TRAINING DATA

In the field of machine learning, DL has proven to be a powerful instrument. Neural networks like *deep convolutional neural networks* or *long short-term memory networks* provide significant advancements in natural language processing, speech recognition, computer vision, and other areas [10], [11], [12]. In diagnostics and prognostics, there is also a rapidly rising number of studies on DL [13]. The fundamental concept of DL is to create a model that includes several layers in which data representations are learned at different levels of abstraction. Through nonlinear

transformation, each layer gradually increases the abstraction of the input data. Overall, this results in a highly complex transformation function, requiring a lot of data for training [10].

The general problem of a shortage of training data when using data-driven methods for diagnosis and prognosis is particularly severe in DL [11]. In order to exploit the potential of DL in domains where there is a lack of data, approaches exist to at least reduce the effort of data generation. Common approaches are transfer learning, active learning, and data augmentation, for which there is already research in the field of diagnostics and prognostics as well. Transfer learning aims to use data or models from other domains to increase predictive accuracy or training efficiency in the domain concerned. In diagnostics and prognostics, transfer learning can be used to leverage data from other operating conditions or even from similar engineering systems [14]. In active learning, it is estimated which data points are the most informative for training. By labeling only these data points, the effort of data labeling is kept as low as possible [15]. The data augmentation approach consists of making small modifications to the original data. These modified data form additional training data and are useful in avoiding overfitting [13], [16].

B. FUSING MACHINE LEARNING WITH KNOWLEDGE

The integration of knowledge is not a fundamentally new topic in machine learning, but in the last five years in particular, this research segment has experienced a holistic consideration and, especially with the development of new methods, a considerable increase in research activities. As a result of the short period of intensive study so far, the establishment of a uniform terminology is still in progress. Thus, the segment is referred to in the literature as (*physics-*) *informed machine learning* [17], [18], [19], *physics-based (machine) learning* [20], [21], *physics guided machine learning* [22], *knowledge-embedded machine learning* [23], and *theory-guided data science* [5]. In the following, the term *physics-informed machine learning* (PML) is used, as the authors see this term becoming more and more prevalent.

The guiding principle for fusing knowledge and data-driven methods is to include as much information as possible from both sources. To achieve such fusion of data and knowledge, a wide range of methods has been developed [19]. The spectrum starts with the integration of partial knowledge that is not sufficient for a complete physics-based modeling. For instance, it could be knowledge about valid bounds, monotonicity constraints, or correlations of intermediate or target variables. An overview of PML methods for this purpose is given in [24]. The spectrum of PML methods continues with the existence of sufficient knowledge to create an entire PM. This PM, although often containing considerable errors, provides an estimate of the actual value of the target variable. Such a PM can be combined with the DM in various ways, as studied in this paper. The existence of a highly precise PM could be regarded as the end of the spectrum. In such a case,

the often small amount of available data is not used to learn a data-driven prediction of the target variable. Instead, the data are used to narrow down the imprecise knowledge about the distribution of model parameters [2], [22].

A comprehensive survey on the whole topic of knowledge integration is provided in [19], which not only considers the process of integrating knowledge into the learning process itself but also the way in which knowledge is represented. An alternative form of survey can be found in [22], which focuses on the modeling of cyber-physical systems. The main criteria for categorization in [22] are physics-based preprocessing, physics-based network architectures, and physics-based regularization. A cornerstone in the research segment has been laid in [5] with its survey and the introduction of the term *theory-guided data science*. Further examples of studies on this subject using a different classification are provided in [25] and [26]. The reviews listed in this section address PML in general, unrelated to the subject of PHM. Therefore, in Section III, an overview of PML with regards to PHM is given.

What are the main advantages of knowledge integration over approaches that reduce the effort of data generation? With the latter, the fundamental source of information remains data. Hence, the inherent advantages of knowledge integration over a purely data-driven approach do not change. One advantage of PML is that knowledge of causal relationships applies beyond the support of the collected data, reducing the general dependency on data availability [12]. Moreover, a deep neural network like other DMs may provide predictions that do not comply with physical laws or other boundary conditions. A behavior that can be inhibited by knowledge integration. Another advantage is that PML can improve the explainability of models and their predictions [19]. All three advantages mentioned are highly relevant in diagnostic and prognostic applications and emphasize the relevance of fusing machine learning with knowledge for PHM.

III. STATE OF THE RESEARCH ON HYBRID PROGNOSTIC METHODS

The problems of insufficient training data and the complexity of physics-based modeling are identified in [27] as the two major challenges in the application of RUL prediction. For this reason, PML, which promises improvement through a combination of both approaches, is of great importance for PHM. The literature on diagnostics and prognostics has already begun to consider the topic of PML, mostly under the designation *hybrid approach*, which is also the term used in the subsequent overview. In the following, the state of research on hybrid methods in prognostics is presented along with references to related approaches in diagnostics. Thereby, methods are only considered in which data-driven and physics-based modeling explain the same relationship between input and target variables. This paper does not address the numerous hybrid methods where the DM and PM perform different subtasks and thus complement each other.

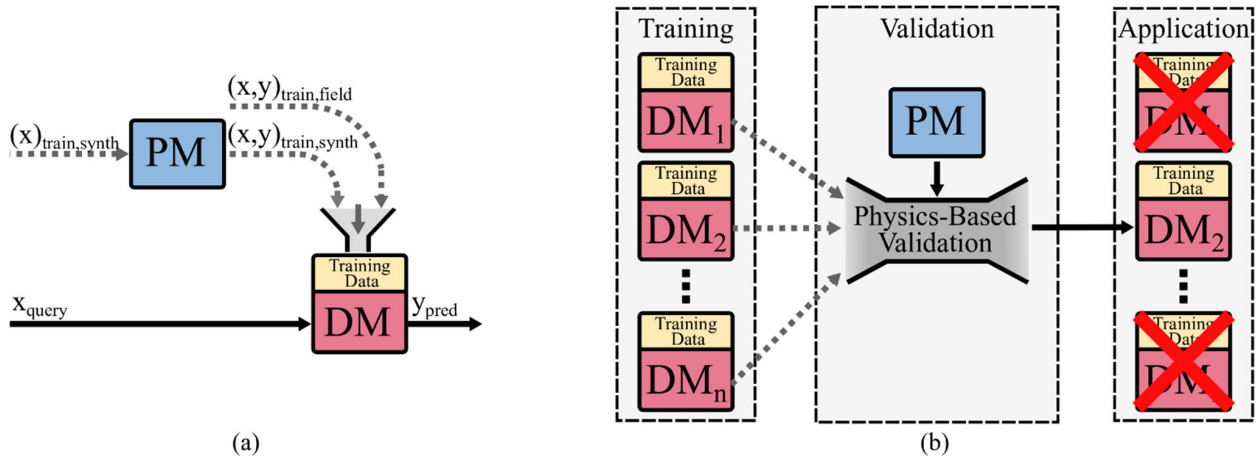


FIGURE 1. Illustration of the functional principle of the hybrid prognostic methods (a) physics-based generation of synthetic training data and (b) final hypothesis set validation.

An example of such an excluded method would be estimating the future load profile with a DM and using a PM to generate an RUL prediction based on that estimated load profile. This type of hybrid method has been studied extensively compared to those where PM and DM share the same objective [28].

To differentiate between hybrid methods in the following, the characteristic of a passive and an active use of the PM is introduced. In the case of a passive use, the PM is not involved in the formation of the overall prediction; instead, its outputs foster the training and validation of DMs. Thus, for these methods, computation of the PM is only required during the training phase of a prognostic application. When the PM is actively used, its output is involved in the formation of the overall result for a specific query point. Consequently, these methods also require the PM to be computed during the application phase. Based on the distinction between passive and active, Sections III-A and III-B present hybrid methods. Thereby, the functionality of a method is first introduced. Subsequently, its applications in prognostics and, if relevant, related approaches in diagnostics are outlined. The functionality is presented with the help of Figs. 1 to 3. In these figures, PMs are visualized by blue boxes and DMs along with their training data are visualized by red-yellow boxes. Operations performed in the training phase are represented by dashed lines and those in the application phase by solid lines. Input data are denoted by x and the target variable or its estimates are denoted by y . The index *train* is used to specify training data. To express that data are from a query during the application phase, the index *query* is used for input data and *pred* is used for output data.

A. HYBRID METHODS FOR RUL PREDICTION WITH PASSIVE USE OF THE PHYSICS-BASED MODEL

In the presence of one or more PMs as well as DMs, two passive methods for model combination exist. These are designated in this paper as *physics-based generation of synthetic training data* and *final hypothesis set validation*.

1) PHYSICS-BASED GENERATION OF SYNTHETIC TRAINING DATA

This method focuses on the availability of data. Therefore, the PM is computed to generate additional labeled data. These synthetic data are utilized to extend the set of actual training data, e.g., field data. Fig. 1a illustrates this approach. Specific features of this method are, on the one hand, that data can be generated in all areas of the state space in which the PM is valid. Thus, areas where less actual training data is available can be selectively covered by synthetic data. On the other hand, the synthetic data can be used for a pre-training of a DM in the sense of a physics-guided initialization [25]. The training with actual data, especially for small data sets, is used to subsequently fine-tune the pre-trained model [29].

Regarding the application of *physics-based generation of synthetic training data* in prognostics, [18] and [30] are to be highlighted. Both use a PM that can be described as a low-fidelity PM. The model is used to pre-train a recurrent neural network whose structure is specifically adapted to the given problem. In addition to these works on prognostics, there are also various studies on diagnostics in which a PM enhances the training data. As shown by [31], [32], and [33], these studies often use high-fidelity PMs and apply transfer learning approaches to combine synthetic and actual data.

2) FINAL HYPOTHESIS SET VALIDATION

Another form of the passive use of entire PMs is a final hypothesis set validation. The entirety of the possible parameterizations of a DM constitute its so-called hypothesis set. The objective of the training process is to select one parameterization from this set as the final hypothesis. The development of a diagnostic and prognostic application often involves the training of several models with different initializations, hyperparameters, or even entirely different learning methods. A sufficient generalization of these models cannot be guaranteed. In order to validate trained models, extensive test data are usually required, which is specifically retained from the

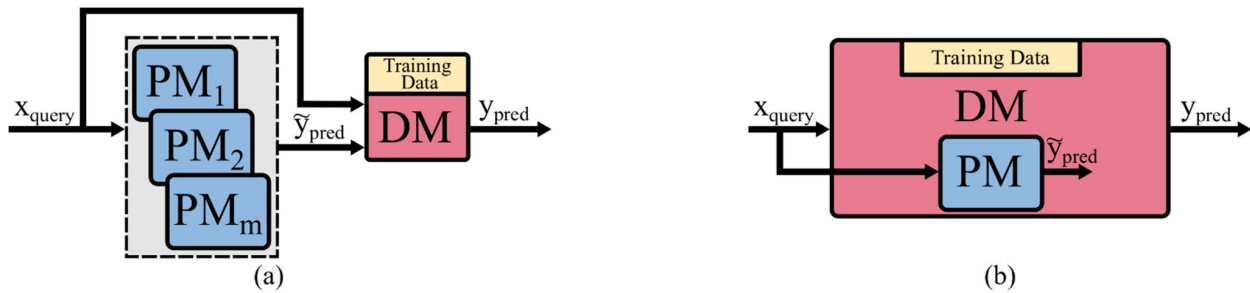


FIGURE 2. Illustration of the functional principle of the hybrid prognostic methods (a) physics-based model as input and (b) physics-based model within the data-driven model.

training. If knowledge or even a complete PM is available, it can be used for an extended validation as well [19], [34]. Given one or more complete PMs, agreement with those can be used as one of the validation aspects for the training results. This method of knowledge integration is shown in Fig. 1b.

The authors are not aware of a final hypothesis set validation in diagnostics and prognostics that uses an entire PM. A check against partial knowledge on the problem, however, exists. This falls within the scope of post-hoc explanation in explainable machine learning, where the fundamental admissibility of a learned model based on human knowledge is examined [35].

B. HYBRID METHODS FOR RUL PREDICTION WITH ACTIVE USE OF THE PHYSICS-BASED MODEL

The following hybrid methods are characterized by the fact that the output of a PM at the given query point contributes directly to the overall output of the hybrid method. Some of these methods can be further subdivided by their topology, using the designations serial and parallel. In serial approaches, the output of one model becomes the input of a subsequent model. In parallel methods, DMs and PMs generate their outputs independently of each other, which are then joined by a mathematical function [36]. The methods of active use also share many similarities with the conventional ensemble methods of machine learning, as noted below.

1) PHYSICS-BASED MODEL AS INPUT

The method *physics-based model as input* corresponds to the definition of the serial approach. The outputs of one or more PMs form additional input features of the DM. Thus, in the case of prognostics, the estimates of the PMs on the future health or the RUL, are contained in the input vector of the DM alongside the regular input values [37]. This functionality, which is shown in Fig. 2a, is similar to the ensemble method of stacking. Stacking uses the outputs of an ensemble of DMs as an input of a higher-level meta learner. The main difference compared to the hybrid variant is that no PM provides input to the higher-level learner.

So far, there is limited research on the use of *physics-based model as input* regarding a hybrid prediction of health or RUL. The closest to this hybrid method is the work of [28], where the PM provides a prediction for short-term effects

on health. This short-term prediction is taken up by a data-driven, similarity-based approach that generates a long-term prediction. Additionally, in [28], it is stated that the presented approach is the first hybrid prediction in which both types of models relate to the same task and are connected serially. In contrast, there are various papers on *physics-based models as input* in diagnostics. However, the two model types usually have different prediction tasks in that the PM describes the system behavior in the normal state. The actual diagnosis is only performed by the DM. This type of hybrid diagnosis is studied in [38], [39], [40], and [41].

2) PHYSICS-BASED MODEL WITHIN THE DATA-DRIVEN MODEL

This hybrid method incorporates the PM into the structure of the DM, as illustrated in Fig. 2b. It is the only hybrid method in this section that cannot be compared to conventional ensemble methods and also cannot be designated as serial or parallel. Here the PM provides its prediction of the future health or the RUL, not as a part of the input features but at an intermediate stage of the DM's data processing. For this, the use of probabilistic graphical models is particularly suitable. This is because of the inherent interpretability of these models, which allows elements of the model to be assigned a specific meaning like damage level or future health. It enables the specification of edges and nodes based on individual knowledge but also the inclusion of entire PMs.

Whether *physics-based model as input* or *physics-based model within the data-driven model*, both methods are similar in that insights about the quantity to be estimated are available to the DM at a certain stage of its data processing. This can be shown by expressing these methods as equations, using the same variable designations as in Fig. 2. For the method *physics-based model as input*, it results in

$$y_{\text{pred}} = f([\tilde{y}_{\text{pred}}, \mathbf{x}_{\text{query}}]). \quad (1)$$

The square brackets denote that \tilde{y}_{pred} and $\mathbf{x}_{\text{query}}$ are combined within one input vector. The method *physics-based model within the data-driven model* can be written as

$$y_{\text{pred}} = f(\tilde{y}_{\text{pred}}, \mathbf{x}_{\text{query}}). \quad (2)$$

Hence, the learning task for both methods is to generate a mapping from the estimate of the PM \tilde{y}_{pred} , in conjunction

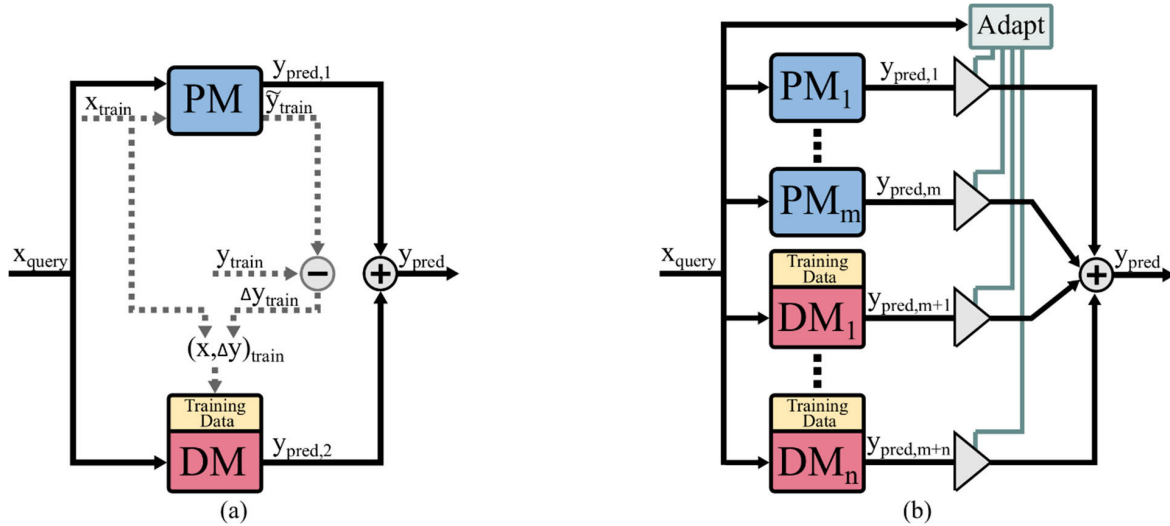


FIGURE 3. Illustration of the functional principle of the hybrid prognostic methods (a) residual modeling and (b) regions of competence.

with additional input values \mathbf{x}_{query} , to the target variable’s actual value [19], [22]. Both methods are therefore related to knowledge-based and highly in-depth feature engineering, as applied in [42]. While the first method matches a regular ensemble approach, the second does not. The second one is particularly suitable if the chosen structure of the DM and the embedding of the PM are based on a physical context.

There are several examples of the integration of knowledge into probabilistic graphical models in diagnostics and prognostics. The incorporation of entire PMs, however, is scarce. The only example of this in the case of prognostics is presented in [20]. It introduces a research project whose objective is to develop a predictive system that not only assesses the safety status of aircraft but also that of the entire airspace. As a central element of information fusion, a probabilistic graphical model is utilized that incorporates PMs, among others.

3) RESIDUAL MODELING

Another hybrid method for an active use of the PM is *residual modeling*. It intends to compensate for deviations of the PM’s predictions from the target value using a DM. When developing the prognostic application, first the PM’s residuals regarding the training data are determined

$$\Delta y_{train} = \tilde{y}_{train} - y_{train}. \tag{3}$$

Subsequently, the DM is trained to predict these residuals. During the application phase, both models calculate predictions independently of each other, which are then added [25]

$$y_{pred} = y_{pred,1} + y_{pred,2}. \tag{4}$$

The described procedure is also visualized in Fig. 3a.

In *residual modeling*, the outputs of the DM and the PM refer to the same quantity, e.g., the RUL. However, these models do not share the same prognostic objective, with the

DM predicting residuals of the PM. Thus, *residual modeling* is a borderline case regarding the definition of hybrid methods considered in this paper. *Residual modeling* can be considered a parallel hybrid method. Furthermore, it is similar in function to the basic principle of the ensemble method of boosting. In boosting, the model ensemble is trained sequentially. During the training of one ensemble element, the objective is to compensate for errors of the previously trained models [43].

The only paper on *residual modeling* in prognostics known to the authors is [44]. In this, the voltage degradation of fuel cells is predicted. A PM is used to describe the overall degradation trend. Added to this is the output of the DM. It serves as a model for specific degradation effects, such as reversible degradation. Besides, the study on the diagnosis of heat exchangers in [45] is particularly relevant. On the one hand, the method’s basic effectiveness in comparison to the purely physics-based approach is confirmed. On the other hand, it is also reported that residual modeling is inferior to the serial hybrid approach.

4) REGIONS OF COMPETENCE

The method *regions of competence* is based on the presence of a model ensemble that consists of one or more PMs and DMs. The fusion of these models is done by a weighted sum of their outputs, as shown in Fig. 3b. Expressing the hybrid method as an equation yields

$$y_{pred} = \sum_{i=1}^{m+n} w_i(\mathbf{x}_{query}) \cdot y_{pred,i}, \tag{5}$$

where w_i is the weight of the i -th output and $m + n$ is the amount of models. The weighting is based on a local competence assessment, for which various approaches exist. Thereby, the value of w_i is adapted to how high the

competence of the corresponding model is estimated for the position $\mathbf{x}_{\text{query}}$.

Fundamentally, the designation and functionality of the hybrid method are the same as for the corresponding purely data-driven ensemble method [43]. The difference is that the hybrid method also includes PMs within the model ensemble. The motivation behind the hybrid and the data-driven variant is that no model by itself performs best throughout the entire state space. Otherwise, the effort of having multiple models would be unnecessary [46]. Since the models within the ensemble provide estimates independently of each other, *regions of competence* is to be designated as a parallel method.

Regarding the application in prognostics, [46] presents a variant of *regions of competence* that enables embedding knowledge about the costs of a misestimate into the competence assessment. Using the example of a turbofan engine degradation data set, it is shown how an underestimation of the RUL can be preferred to an overestimation. Further work on combining different models based on local competence includes, for example, [47], [48], and [49].

IV. CASE STUDY ON THE USE OF HYBRID METHODS FOR RUL PREDICTION OF FILTERS

In the following, a case study on the predictive accuracy of hybrid RUL predictions for a filtration process is conducted. As a foundation for the hybrid prediction, a physics-based filtration model and a Gaussian process as DM are created. For the case study, four of the previously presented hybrid methods are applied to the RUL prediction of filters. For this purpose, the subject of the case study is first presented in detail. Next, the development of the PM and DM as baseline models is explained. Thereafter, the implementation of the different hybrid methods is presented. Finally, the predictive accuracy of these methods is analyzed.

A. SUBJECT OF THE CASE STUDY

Filtration to separate solids from a fluid, in this case gas, is a process that can be found in almost every branch of industry and therefore has already been the subject of research on diagnostics and prognostics, as [9], [50], and [51] show. In the present case, a test rig is used for performing automated life testing of filter mats. During testing, the differential pressure across the filter increases as a result of the filter being loaded with dust particles. The filter is considered to have failed as soon as its differential pressure exceeds a threshold of 600 Pa.

The parameters and measured values of the test rig, together with their range of values within the case study, are

- the particle size distributions of the test dust {dust types ISO 12103-1 A2, A3 and A4},
- the amount of dust feed per time {60, 80, 119, 159, 178, 237, 238, 317 mm³/s},
- the nominal flow rate {54, 82 l/min},
- the current test time [0, 336 s],
- the measured flow rate [0, 84 l/min], and
- the measured differential pressure [0, 600 Pa].

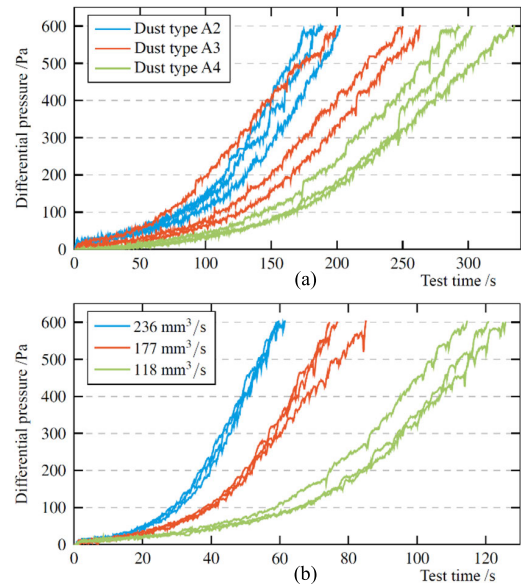


FIGURE 4. Differential pressure trajectories of filter loading under identical conditions except for (a) the type of dust and (b) the dust supply per time. There are (a) three types of dust and (b) three levels of dust feed with three trajectories each.

The differential pressure trajectories of run-to-failure tests with different dust types and different levels of dust feed are shown in Fig. 4. The differential pressure trajectories show a typical characteristic despite disturbing influences. They start from a convex slope due to the initially prevailing depth filtration and transition into a linear slope of the cake filtration [52]. Further information on the filtration test bench and the life tests can be found in [53] and [54].

B. PHYSICS-BASED MODELING OF FILTER LOADING

The scientific literature contains various PMs for the calculation of filter loading. The models range from low- to high-fidelity approaches, as shown by [55] and [56]. For the modeling of the test scenario, a model is chosen that can be classified as a middle ground between low- and high-fidelity. The model is based on the works of [52], [55], [57], and [58].

The fundamental concept of the model is to divide the filter into separate layers. In the present case, the number of layers J is five. At one simulation time, the current filter efficiency and the mass of absorbed particles in the foremost layer are calculated as a function of the previous loading in accordance with the equations given by [52]. These equations account for the particle collection mechanisms of diffusion and interception. Based on the calculated absorbed particle mass, the reduced particle concentration of the aerosol flowing into the subsequent layer is determined by

$$f_{t,j+1,i} = f_{t,j,i} - m_{t,j,i}, \quad (6)$$

where $f_{t,j}$ represents the particle mass flowing into the j -th layer at the discrete simulation time t . The particle mass collected in the j -th filter layer at time t is designated as m .

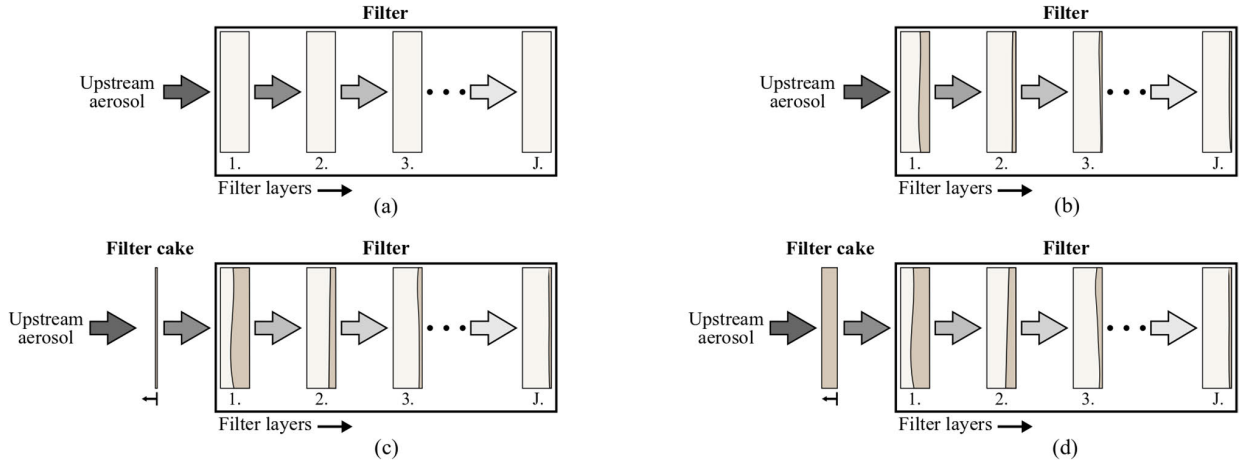


FIGURE 5. Visualization of the physics-based modeling of the filter loading process. The progressive filter loading is illustrated as (a) start of the filtration process, (b) pure depth filtration, (c) onset of cake filtration, and (d) dominating cake filtration.

Since the distribution of particle sizes contained in the dust is known, the filter efficiency and collected mass are calculated for different particle sizes, which are indicated by i . According to this scheme, the loading of each layer from the front to the back of the filter is calculated iteratively for the current simulation time. Since the incoming particle concentration is highest in the first layer, the most particles are captured there. This effect is intensified by the fact that the filter efficiency of a layer rises with its increasing loading. The packing density of a layer $\alpha_{t,j}$ is the sum of the constant packing density of filter fibers $\alpha_{f,j}$ and the collected particles $\alpha_{p,t,j}$. As soon as $\alpha_{t,1}$, the packing density of the front layer, reaches the loading limit α_{lim} , a filter cake is formed in front of it. The calculation of its filtration effect and growth due to its increasing loading is performed ahead of the first layer following the iterative calculation of (6). The overall procedure for filter loading calculation is illustrated in Fig. 5.

Once the amount of dust collected by the filter cake and filter layers is calculated for time t , the resulting differential pressure is determined. If a filter cake has already formed at t , its differential pressure is calculated in accordance with [57] and [59] as

$$\Delta P_{t,cake} = \frac{UM_{t,0}}{\rho} \cdot \frac{\alpha_c}{(1 - \alpha_c)^3} \cdot \frac{36K_{ck}\mu}{C\bar{d}_p^2}. \quad (7)$$

In the equation, U is the flow velocity, $M_{t,0}$ is the total mass of the filter cake per area at time t , and ρ is the density of the dust. The fraction in the middle reflects the influence of the cake packing density α_c . The latter part of the formula includes the Kozeny constant K_{ck} , the gas dynamic viscosity μ , the slip correction factor C and the mean particle diameter \bar{d}_p .

The differential pressure of a filter layer is calculated using a modified equation from [52]

$$\Delta P_{t,j} = 64\mu UZ \frac{\alpha_{p,t,j}^{3/2}}{\bar{d}_p^2} \left(1 + 56(\alpha_{t,j})^3\right). \quad (8)$$

The differential pressure at time t across the j -th layer is $\Delta P_{t,j}$. The width of a filter layer in the direction of the air flow is denoted as Z . The total differential pressure is then calculated as the sum of the differential pressures of the filter cake and the filter layers

$$\Delta P_t = \Delta P_{t,cake} + \sum_{j=1}^J \Delta P_{t,j}. \quad (9)$$

A simulation of filter loading with the PM described requires detailed knowledge of various parameters. This involves the fiber diameter, the packing density of the fibers within the filter, the filter geometry, the inflowing particle mass per time, the statistical distribution of the particle size, the volumetric flow rate, and the density of the dust. However, three parameters with great importance for the predicted trajectory are highly difficult to measure or are subject to strong fluctuations between individual life tests. These are the packing density of the fibers in the filter α_f , the loading limit α_{lim} , and the packing density of the filter cake α_c . The first parameter α_f is subject to significant manufacturing tolerances due to the spunbond manufacturing of the filter mats. The effect of manufacturing tolerances is particularly pronounced due to the fact that a comparatively small filter area of 61.3 cm² is tested. The second parameter α_{lim} , represents the loading limit above which the formation of a filter cake starts. The method of calculation proposed by [52] is suitable for thin high-efficiency particulate air filters (HEPA) and ultra-low penetration air filters (ULPA) with an early onset of cake filtration. The given filter, instead, has a significantly lower average packing density and greater depth filtration. Therefore, no precise knowledge is available for the loading limit. Furthermore, this parameter depends on the filter packing density, which is affected by production tolerances. Consequently, both parameters also fluctuate significantly between different run-to-failure tests. The third parameter α_c describes the density with which the filter cake builds up in front of the filter, which is also difficult to measure directly and strongly depends on the test setup [57], [59].

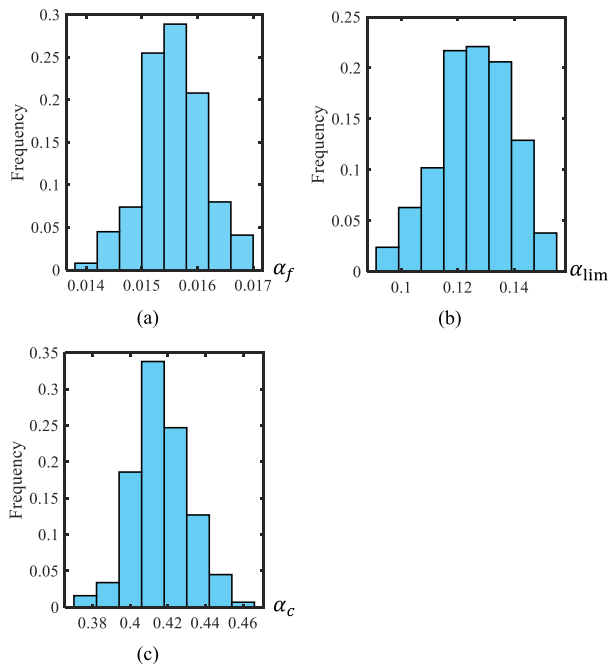


FIGURE 6. The marginal distributions of the unitless parameters (a) α_f , (b) α_{lim} , and (c) α_c obtained by the MCMC sampling for a prediction of the PM.

Besides the actual modeling, determining uncertainties is also a central element in the development of a diagnostic or prognostic application. In the present case, the two governing sources of uncertainty are, first, the imprecise information on the value of α_f , α_{lim} , and α_c , and second, uncertainties due to measurement noise and other disturbing influences. The former is designated as epistemic uncertainty and the latter as aleatory uncertainty. This assignment is based on the fact that the parameter uncertainty, in contrast to that caused by the noise, can be reduced with additional data [4].

In this case study, Markov chain Monte Carlo (MCMC) sampling is used, which has already been employed in diagnostic and prognostic applications [60]. Thus, for every prediction of the PM, an MCMC sampling is performed with the differential pressure measurements available up to the point of prediction. The result is a posterior distribution of the parameters α_f , α_{lim} , and α_c containing 1500 samples. Based on this distribution, the differential pressure prediction of the PM is calculated. As an example of an MCMC sampling, the marginal distributions of α_f , α_{lim} , and α_c for a prediction with the intermediate dust size A3 are illustrated in Fig. 6. The prediction itself is presented below and shown in Fig. 7b.

Besides the parameter uncertainty, another major source of uncertainty in this case is measurement noise and other disturbing influences, resulting in aleatory uncertainty. They are assumed here to be independent and normally distributed. Both epistemic and aleatory uncertainties are an inherent part of diagnostic and prognostic applications, which, more than being reduced, must be properly quantified [61]. In this case, the MCMC samples of the three model parameters are used to calculate differential pressure trajectories, which in turn

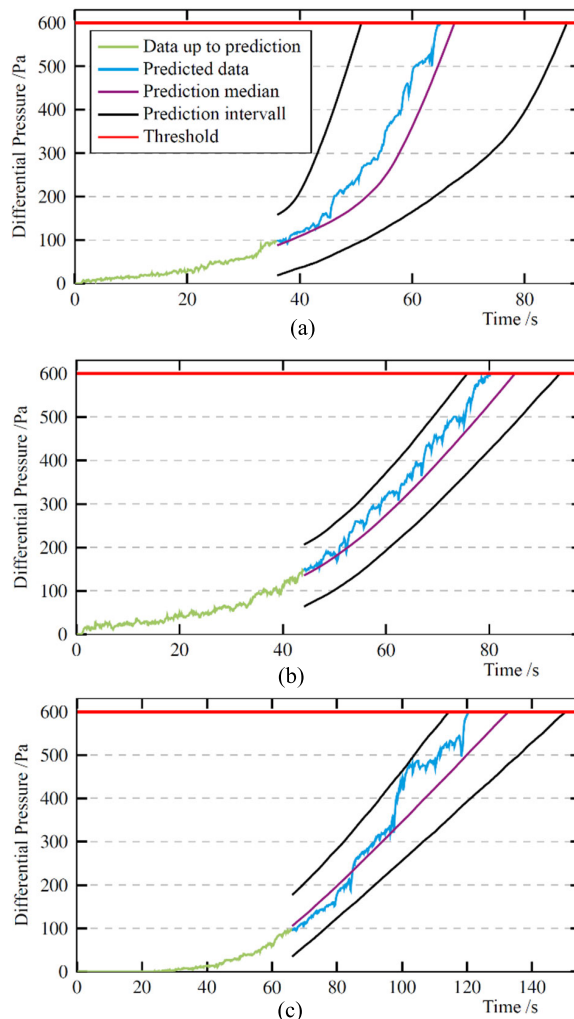


FIGURE 7. Exemplary differential pressure predictions of the PM for run-to-failure trajectories of the three dust types ISO 12103-1 a) A2, b) A3, and c) A4. The pressure readings colored green are those available at the time of prediction that are used for MCMC sampling. The further progression of the readings, which are colored blue, is to be predicted. The prediction of the PM is represented by a prediction median and a prediction interval (PI). The selected significance level of the PI is 90%.

are additively superimposed by normally distributed noise. The two-sided prediction interval is determined as the 5th and 95th percentiles of the noisy trajectories at each calculation step. Examples of model-based predictions, including uncertainty quantification, are shown in Fig. 7 for each dust type. Based on this, the RUL prediction of the PM is determined as the intersection of the differential pressure prediction with the threshold of 600 Pa. From the pressure measurements shown, it is evident that the influence of disturbances or noise tends to increase with rising filter loading. Here, this so-called heteroskedasticity is neglected in the uncertainty analysis of the PM, as a constant noise level is assumed.

C. DATA-DRIVEN MODELING OF THE REMAINING USEFUL LIFE OF FILTERS

For data-driven RUL prediction, a Gaussian process (GP) is used due to its rigorous uncertainty representation.

Data-driven methods can be used for direct or indirect RUL prediction. On the one hand, a direct prediction of the failure time or the RUL can be learned. This is equivalent to the application of multivariate pattern mapping with RUL as the target variable. On the other hand, an indirect RUL prediction can be learned by iteratively predicting the evolution of the damage-determining variable over time. In order to obtain an RUL estimate, the time when the predicted damage trajectory reaches the failure criterion is computed [62]. In this case study, the hybrid prediction is studied only for the RUL of filters. The additional complexity of an indirect prediction is not required, so the GP is implemented in the form of a direct prediction.

Due to the direct prediction, the GP's target variable y_{pred} is the RUL. The input features \mathbf{x} are based on parameters and sensor data of the test rig that are listed in Section IV-A. The vector \mathbf{x} contains the dust type as dummy coding, the current test time, the mass of dust supplied per time, the measured values of flow rate and differential pressure, as well as the logarithmized differential pressure. Fundamentally, a GP regression is structured as

$$y_{\text{pred}}(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (10)$$

with the mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ [63]. Thereby, $m(\mathbf{x})$ reflects an overall trend of the target variable and thus allows the integration of knowledge to a certain degree, comparable to feature engineering. In the present case, the highest predictive accuracy with respect to the test data is obtained by a zero-mean function $m(\mathbf{x})$. Additionally, in this configuration, a model that is "as data-driven as possible" is obtained. The similarity of $y_{\text{pred}}(\mathbf{x})$ to the observed values in the training data is represented by $k(\mathbf{x}, \mathbf{x}')$. Here, an isotropic exponential kernel is used for this.

The filter test rig enables run-to-failure tests to be carried out with comparatively little experimental effort. As a result, 50 run-to-failure trajectories are available just for the training of the GP. This amount of data is extensive enough that it allows the GP to perform significantly better than the PM, with an RMSE of 7.89 s for the test data. The PM, in contrast, shows an RMSE of 19.4 s. However, this neither corresponds to typical prognostic applications nor to the case discussed in the paper. Therefore, the number of trajectories for the GP's training is reduced to 15 by sampling without replacement. To minimize the influence of the sampling, it is repeated 30 times, so that the hybrid methods are also calculated correspondingly many times. The results of these 30 repetitions are averaged. Moreover, the training trajectories selected during sampling are the same for each hybrid method.

As a representative test criterion, the censored test data from the "Preventive to Predictive Maintenance" data set provided on the machine learning platform Kaggle are used [54]. These are 50 additional run-to-failure trajectories of the test rig. Each of the test trajectories is censored at a random time for which the RUL is to be predicted.

```

1: TrainingData ← 50 run-to-failure trajectories
2: TestData ← 50 run-to-failure trajectories
3: for i = 1 to 30 do
4:   TrainingSample ← randomly draw 15 trajectories
      from TrainingData
5:   GP ← GPTraining(TrainingSample)
6:   for j = 1 to 50 do
7:     TestResult(i,j) ← Hybrid(GP, PM, TestData(i))
8:   end
9: end
10: MeanResult(j) ← mean_i(TestResult(i,j))

```

FIGURE 8. General procedure of data handling within the case study.

The described general handling of training and test data is summarized in Fig. 8. Changes to this procedure that arise depending on the concept of a hybrid method are explicitly stated in Section IV-D.

D. IMPLEMENTATION OF HYBRID METHODS FOR COMBINING PHYSICS-BASED AND DATA-DRIVEN MODELS

Using the PM presented in Section IV-B and the DM from Section IV-C, hybrid methods for RUL prediction are implemented. This implementation includes the four hybrid methods *physics-based generation of synthetic training data*, *physics-based model as input*, *residual modeling*, and *regions of competence*. As the method *final hypothesis set validation* is highly dependent on the selected search space of possible DMs, the comparability with the other methods is limited. Therefore, it is not included in this case study. Furthermore, due to the similarity between *physics-based model as input* and *physics-based model within the data-driven model*, only the former is applied. In the following, important characteristics and special features concerning the implementation of the four hybrid methods are discussed.

1) PHYSICS-BASED GENERATION OF SYNTHETIC TRAINING DATA

This method intends to address a lack of training data by using the PM to generate synthetic training data. In the present case, the operating conditions of the filter are determined by three parameters: dust type, dust feed per time, and nominal flow rate. The values of these parameters occurring within the case study are listed in Section IV-A. The combinatorics of these values results in 48 different operating conditions. However, only 15 run-to-failure trajectories are taken for training, for which even duplications of the operating conditions can occur. Consequently, as in many diagnostic and prognostic applications, only a part of the possible operating conditions can be covered by the training data. As an implementation of the method *physics-based generation of synthetic training data*, on average 36 run-to-failure trajectories are computed using the physics-based filter model for those operating conditions that are not present in the training data.

These trajectories are then used as synthetic training data along with the actual training data. Both types of data are weighted equally in the training of the GP. Regarding the procedure in Fig. 8, this hybrid method requires the generation of synthetic training data to be added after line 4. For the testing of the hybrid method in line 7, however, the PM is not needed.

2) PHYSICS-BASED MODEL AS INPUT

The first step in the implementation is to compute the RUL prediction of the PM for the 15 randomly selected run-to-failure trajectories taken as training data. Within one training trajectory, the RUL prediction of the PM is recalculated at intervals of 2 s, including the MCMC sampling with the data accumulated up to that point. In the subsequent training of the GP, the original input features remain. They are only extended by the current RUL prediction of the PM as an additional input feature. For the procedure in Fig. 8, this method would entail adding after line 4 the additional computation of PM predictions for the sampled training data.

3) RESIDUAL MODELING

In the same way as for the method *physics-based model as input*, the PM predictions are calculated for the training data. Instead of an additional input feature, the predictions are used for transforming the labels of the training data. The GP's target variable becomes the deviation of these PM predictions from the actual RUL values. Therefore, this method would require adding the calculation of the PM's residuals after line 4 in Fig. 8. In the application phase, the output of residual modeling is calculated by summing the outputs of the GP and the PM. This also has to be considered in the uncertainty analysis. An analytical summation of the output distributions of both model types is not feasible due to the PM and its MCMC sampling. Therefore, a distribution reflecting the uncertainty of the overall output is obtained by a Monte-Carlo simulation.

4) REGIONS OF COMPETENCE

The model ensemble used for the method *regions of competence* consists of the GP and the PM. The local competence of both models is assessed during runtime based on the 40 validation data points closest to the query point. The proximity of validation data points is determined using the Euclidean distance norm. For the competence measure, the sum of the residual squares is applied. For the competence assessment, separate validation data – in the best case, entire trajectories – have to be withheld from the GP's training. The consequence of such a reduction of training data would be a less accurate GP and, thus, a different data-driven baseline model compared to the other hybrid methods. Hence, for comparability among the hybrid methods, 15 additional validation trajectories are intentionally sampled from the 50 training trajectories.

The model assessed as most competent locally is used for the overall output. This corresponds to a model selection where the weighting of the more competent model is 1 and

the output of the other model is weighted with 0. In the literature, such an approach is often referred to as *dynamic model selection* [64]. The choice of using a selection is based on the heuristic recommendation on ensemble methods for classification tasks in [43]. This recommendation can be transferred to the regression problem of an RUL prediction. For heterogeneous ensembles consisting of a few strong models, as here, a selection is recommended. For a large number of weak learners, a combination of the output values should be performed, e.g., by bagging. The described procedure for the implemented method *regions of competence* is identical to the one shown in Fig. 8.

E. ANALYSIS OF THE PREDICTIVE ACCURACY OF THE IMPLEMENTED HYBRID METHODS

The predictive accuracy of the hybrid methods is determined by using 15 randomly selected run-to-failure trajectories as training data and 50 censored trajectories as test data, as the procedure in Fig. 8 shows. The first metric used for assessing the predictive accuracy is the root-mean-square error (RMSE). To calculate the RMSE, a representative point prediction needs to be determined from the prediction distribution. In the present case, this is the median. In addition to the accuracy of the point prediction, the uncertainty estimates are also evaluated. For this purpose, the coverage rate of the prediction interval is calculated as a second metric. It specifies the rate at which the observed RUL values fall within the calculated prediction interval. Since the test data impose a specific time of prediction, no metrics can be used for assessments that consider the behavior of the RUL prediction within a run-to-failure, such as the prognostic horizon [7]. Due to the variation of the 15 trajectories used for the training, the accuracy of the GP also changes, as does the hybrid method's accuracy. To account for this variation, the mean and standard deviation of the RMSE values and coverage rate are determined.

The results of implementing four hybrid methods as well as the results of the physics-based and data-driven baseline models are provided in Table 1. On the test data, the GP has an RMSE of 14.9 s. Despite the reduced amount of 15 run-to-failure trajectories for the GP's training, the physics-based filter model has a 30% higher RMSE of 19.4 s. The RMSE values of the hybrid methods show, in two cases, a significant improvement over both baseline models. The method *physics-based generation of synthetic training data* achieves the lowest RMSE of 10.4 s, which is a 30% improvement compared to the purely data-driven GP. With a standard deviation of 0.8 s, its range of variation is also significantly smaller than the GP's. The method *physics-based model as input* also provides a significant improvement over the GP with an RMSE of 12.8 s. The *residual modeling* in contrast, does not provide any improvement over the GP. However, due to its mode of operation, it may rather be compared to the PM. In *residual modeling*, the GP is specifically trained to compensate for the errors of the PM, which it does to some extent

TABLE 1. Test results of the two baseline models and the implemented hybrid methods. The values of the two metrics, RMSE and coverage rate, are given as the arithmetic mean and empirical standard deviation.

Prognostic method	RMSE / s	Coverage rate / %
Physics-based model	19.4	90.0
Gaussian process	14.9 ± 4.0	90.7 ± 8.3
Phys. generation of synthetic training data	10.4 ± 0.8	90.0 ± 2.4
Phys. model as input	12.8 ± 3.5	92.1 ± 8.2
Residual modeling	19.1 ± 0.4	99.5 ± 1.2
Regions of competence	14.7 ± 3.6	91.9 ± 4.6

with an RMSE of 19.1 s and its small standard deviation of 0.4 s. The method *regions of competence* performs better than the GP and the PM as well. The improvement over the GP, though, is only 0.2 s.

Considering the metric coverage rate, the PM and DM show minor or no deviations from the targeted coverage of 90%. Thus, both models provide a suitable baseline for the hybrid methods. These methods also yield only small deviations, except for the *residual modeling*. This hybrid method yields a coverage rate of 99.5%. It is found that the compensating GP returns such wide uncertainty bounds that those of the overall output also become unnecessarily wide. Looking at the calculated standard deviation of the coverage rate, the GP has the highest range of variation at 8.3%. Except for *physics-based model as input*, the hybrid methods also reduce the standard deviation by at least 3.7%. However, since the mean coverage rate of *residual modeling* is close to the maximum of 100%, its range of variation is only comparable to a limited extent.

Taking into account the RMSE and the coverage rate, the preliminary conclusion is that for the case study, the method *physics-based generation of synthetic training data* provides the highest predictive accuracy. Regarding both metrics, *residual modeling* provides the worst accuracy, which also confirms the results mentioned in Section III-B about the inferiority of this hybrid method reported in [45].

V. EVALUATION OF THE HYBRID METHODS

Following the no-free-lunch-theorem that no predictive algorithm universally performs best, the observed performance differences between the hybrid methods cannot be considered universal across prognostic applications [65]. Nevertheless, conclusions about individual attributes of the hybrid methods can be drawn based on their functionality and supported by the case study. Section V-A assesses the ability to provide accurate RUL predictions in areas of the state space with little or no training data. Section V-B analyzes which hybrid methods are particularly affected by a high variance of the DM. In Section V-C, the computational efficiency of the hybrid methods is examined. A summary of the results and a discussion of their scope are provided in Section V-D.

A. PREDICTION IN AREAS OF THE STATE SPACE WITH LITTLE OR NO DATA

In the section above, the predictive accuracy of the hybrid model is analyzed with respect to the given set of test data. As argued in Section I, due to the common shortage of data, a key challenge in using data-driven methods is when query points occur that fall in regions of the state space where little or no training data are available. Therefore, this section specifically discusses how the hybrid methods handle this challenge.

DMs are capable of extrapolation to a limited extent [66]. Nevertheless, there is a strong trend that as the distance to training data increases, the accuracy of the DM decreases. In this context, extrapolation is defined as a query point falling outside the convex hull of the training data [66]. However, the predictive accuracy of the PM is often independent from training data. Under the assumption that its validity extends to areas with little or no training data, it provides an additional source of information. To illustrate this on the basis of this study, a correlation analysis is performed. The correlation is determined between the prediction error and the distance to the nearest point in the training data. The calculated correlation factors are listed in Table 2. A special case is the method *physics-based generation of synthetic training data* where synthetic training data is generated. In order to still capture the reduction in correlation due to the PM, these synthetic data are omitted from the calculation of the distance to the nearest training data points.

The average correlation for the GP of 0.31 supports the argued accuracy decline of DMs with increasing distance to training data. In contrast, the PM even shows a negative correlation of -0.12 , which does not have to be the case generally. Nevertheless, it reduces the correlation between error and distance to data points for the applied hybrid methods and thus reduces the dependency on nearby data.

The methods that employ the PM in a manner that specifically counteracts the problem of insufficient coverage of the state space with data are, from the authors' point of view, particularly *physics-based generation of synthetic training data* but also *residual modeling* and *regions of competence*. The first method, as implemented in the present study, enables the generation of additional synthetic training data in areas that are not covered by the actual training data. *Residual modeling* can also ensure reliance on the PM's predictions in such areas, particularly by using a parameterization or training of the DM that causes its predictions to tend to zero at points far from the training data. In the present study, such behavior is obtained by using a zero-mean function for the GP. A similar approach can be applied to the method *regions of competence*. For this purpose, when evaluating competencies, the weighting of the PM would increase in proportion to the distance to training data. However, this approach is not applied to the method *regions of competence* within this case study. The other three hybrid methods continue to use the DM to make predictions for data points that are distant from the training data, even if

TABLE 2. Analysis of the error correlation for the two baseline models and the implemented hybrid methods. The Pearson correlation coefficient (PCC) is determined between the prediction error on the test data and the mean Euclidean distance to the nearest five points of the standardized training data. The table lists the mean and standard deviation of the PCC.

Prognostic method	PCC
Physics-based model	-0.12 ± 0.15
Gaussian process	0.31 ± 0.10
Phys. generation of synthetic training data	0.10 ± 0.10
Phys. model as input	0.21 ± 0.13
Residual modeling	-0.09 ± 0.16
Regions of competence	0.18 ± 0.13

TABLE 3. Test results of the two baseline models and the implemented hybrid methods with 5 run-to-failure trajectories as training data as well as the change in the RMSE values compared to Table 1 with 15 trajectories.

Prognostic method	RMSE / s	RMSE change / %	Coverage rate / %
Physics-based model	19.4	0	90.0
Gaussian process	24.8 ± 5.6	+66	88.3 ± 7.5
Phys. generation of synthetic training data	10.7 ± 0.5	+3	88.3 ± 2.7
Phys. model as input	21.6 ± 4.6	+69	90.2 ± 6.5
Residual modeling	19.4 ± 0.3	+2	98.8 ± 1.8
Regions of competence	19.8 ± 2.4	+35	91.3 ± 2.2

these predictions are supported, for example, by a physics-based input feature.

To support the statements made above about the individual hybrid methods, they are calculated for comparison with a reduced and an increased amount of training data. Instead of the 15 run-to-failure trajectories used for the results in Table 1, 5 and 25 trajectories are taken as training data. The procedure of randomly selecting run-to-failure trajectories as training data and its 30 times repetition is the same as in Section IV-E. The results for the modified amount of training data and the differences from the initial results with 15 run-to-failure trajectories are shown in Tables 3 and 4. Overall, these results show a dependency on training data volume for both the GP and the hybrid methods. The only result that does not correspond to the expected behavior is that of the *residual modeling*, which does not provide any further improvement when increasing the training data set to 25 trajectories.

Additionally, the results strongly support the statements on the suitability of the hybrid methods for predictions in areas with little or no training data. Especially when reduced to 5 run-to-failure trajectories, the methods *physics-based generation of synthetic training data* and *residual modeling* show a significantly smaller increase in RMSE than the GP or the method *physics-based model as input*. The method *regions of competence* also has a relatively small increase in RMSE, even without an increased weighting of PMs with

TABLE 4. Test results of the two baseline models and the implemented hybrid methods with 25 run-to-failure trajectories as training data as well as the change in the RMSE values compared to Table 1 with 15 trajectories.

Prognostic method	RMSE / s	RMSE change / %	Coverage rate / %
Physics-based model	19.4	0	90.0
Gaussian process	12.1 ± 1.6	-19	91.7 ± 4.2
Phys. generation of synthetic training data	9.9 ± 0.7	-5	91.5 ± 3.2
Phys. model as input	10.2 ± 1.3	-19	92.6 ± 3.0
Residual modeling	19.1 ± 0.4	0	99.9 ± 0.5
Regions of competence	12.0 ± 1.8	-18	92.9 ± 3.6

growing distance to training data, like proposed above. With 25 trajectories much more data are available; still *physics-based generation of synthetic training data* performs best, but the gap in particular to the method *physics-based model as input* is significantly smaller than for 5 and 15 trajectories.

In the tests with 5 and 25 run-to-failure trajectories, the coverage rate of the hybrid methods shows only slight deviations from the target of 90%. The only exception is still the method *residual modeling*, which yields for the test with 25 trajectories a coverage rate of 99.9%. This almost complete coverage of the test data points thus deviates significantly from the target value.

B. EMPLOYING DATA-DRIVEN MODELS WITH HIGH VARIANCE

Another attribute that systematically influences the effectiveness of some hybrid methods is the variance of the DM. Fundamentally, the accuracy of a DM is affected by the DM's variance. According to the concept of the bias-variance trade-off, the bias and the variance of a model are counterparts, which together with the noise result in the squared prediction error [67]. Therefore, the variance of the DM fundamentally affects its accuracy, as well as the accuracy of all hybrid methods. However, besides this, an excessive variance also has a further negative impact on the hybrid methods *regions of competence* and *final hypothesis set validation*.

For the former method, the local competence of the baseline models is typically assessed using labeled validation data close to the query point. If the predictions of these models show a high degree of variability that is unrelated to system dynamics, there is only a small correlation between the accuracy on the validation data and the accuracy at the query point. As a result, in the case of high variance, the competence assessment would not be able to infer the model performance at the query point based on the validation data and thus would be unable to identify the local best model [64], [68]. Especially for complex DMs, such a high variance might occur.

Another form of competence assessment takes place in the *final hypothesis set validation*, but in this case globally and not locally. The PM is used to determine the suitability

TABLE 5. Implementations of the method regions of competence, each with a different DM.

	Linear model	Gaussian process	Random forest
No. of training trajectories	10	5	5
RMSE of the PM / s	19.4	19.4	19.4
RMSE of the DM / s	25.3 ± 12.1	24.8 ± 5.6	24.6 ± 6.7
RMSE of the hybrid method / s	19.4 ± 3.4	20.0 ± 3.0	20.1 ± 3.8
Number of correct model selections for test data	31.3 ± 6.3	29.5 ± 5.1	28.3 ± 4.5
Number of wrong model selections for test data	18.7 ± 6.3	20.5 ± 5.1	21.7 ± 4.5

of different training results. If the accuracy of the DM is highly variable, this validation is affected in the same way. As a result, it is either necessary to perform the validation much more densely in the feature space, which increases the validation effort, or it causes the validation to be less accurate.

In the remaining hybrid methods, no consecutive step is based on the output of the DM. Thus, for them, no further negative impact of the variance of the DM beyond its mere accuracy is apparent.

Since the method *regions of competence* is also used in the case study, the statements made can also be empirically studied. For this purpose, the hybrid method is implemented based on three different DMs. This includes a linear regression model, a GP, and a random forest (RF). The input features of the three DMs are the same as described in Section IV-C. Since the linear model has a higher RMSE than the GP and RF, 10 trajectories are used for its regression, while only 5 trajectories are used for the training of the other two DMs. The validation data for the local competence assessment are the same for each of the three variants. Furthermore, in order to achieve a high variance for the RF, it consists of only five decision trees. To further enhance the influence of the model variance, different from the implementation described in Section IV-D, the 10 nearest validation data points are used for the competence assessment. Other than that, the procedure is the same as for the other studies on *regions of competence*.

The results of the three implementations of *regions of competence*, each with a different DM, are listed in Table 5. The RMSE of the three models is at a similar level, with the linear model having the largest standard deviation. Nevertheless, the RMSE of the hybrid method is the lowest for the linear model. In addition to the RMSE, another aspect that can be analyzed is how often the competence assessment correctly selects the most accurate model for the 50 trajectories of the test data. Again, the linear model shows the best performance. Even though it has the largest variation in terms of RMSE in this study, locally, a linear model has a small variance, which is why these results support the statement that variance affects the method *regions of competence* [67]. In contrast to the linear model, the competence assessment using the RF as DM selects the correct model least frequently. Given that the RF, with its few decision trees, is likely to have a high variance,

TABLE 6. Frequency of calculating a PM prediction when applying the hybrid methods within the case study. The given frequency refers to the training and the test phases of one data sample with 15 trajectories.

Hybrid method	Frequency of PM calculations in the	
	training phase	test phase
Phys. generation of synthetic training data	36	0
Phys. model as input	768	50
Residual modeling	768	50
Regions of competence	0	2050

this also supports the statement about *regions of competence*. Summarized, although the differences in the results are small, the statements about the method *regions of competence* being affected by the model variance are supported by these empirical results.

C. EXAMINATION OF THE COMPUTATIONAL EFFORT OF THE HYBRID METHODS

The computational capacity in prognostic applications is often highly limited. Therefore, another attribute that decisively influences the choice of a prognostic method to be used is the computational effort [69]. In this case study, the calculation of one PM prediction, including an MCMC sampling with 1500 samples, requires an average of 523.8 s of computation time. By using the same computing infrastructure, the GP takes less than 10^{-2} s for one prediction. This computational efficiency of machine learning models is also repeatedly discussed in the literature on diagnostics and prognostics. Data-driven models are used as surrogate models of the PM in order to reduce computation time. Examples of such work include [70] and [71].

Due to the different combination strategies of DMs and PMs, the hybrid methods differ significantly in the computational effort they require for training and making new predictions. As training mostly takes place during the development phase and the computation of predictions mostly occurs in the application phase of diagnostic or prognostic applications, conclusions can be drawn about the computational efficiency in these phases.

If an attempt is made to minimize the computational effort during training, the method *regions of competence* is advisable. In accordance with the case study, the local competence estimation can be performed dynamically when a prediction of the hybrid method is calculated. Thus, it is the only hybrid method considered where no computation of the PM is necessary for the DM's training. If, however, the computational capacity for the application phase is limited, which occurs often in industrial diagnostic and prognostic applications [72], the methods *physics-based generation of synthetic training data* and *final hypothesis set validation* instead are well suited. Consistent with the characterization

TABLE 7. Summary on the beneficial attributes of individual hybrid prognostic methods, denoted by a dot. The distinction between the training and application phases used in the analysis of the computational efficiency of hybrid methods is indicated by the numbers 1 and 2.

Aspects	Saving computing capacity in training (1) or application (2)	
	Unaffected by a high variance of the DM	
	Accurate prognosis in areas with little or no training data	
Hybrid method	Phys. generation of synthetic training data	● ● 2
	Final hypothesis set validation	● ● 2
	Physics-based model as input	● ●
	Phys. model within the data-driven	● ●
	Residual modeling	● ●
	Regions of competence	● ● 1

of a passive use of the PM from Section III, these methods only employ the PM during the development phase.

The statements on computational efficiency are also confirmed by the case study. Since the effort of calculating the PM dominates, Table 6 lists how often a PM prediction is computed during the training and test phases of the hybrid methods. As stated, of the hybrid methods implemented, the method *regions of competence* requires the least computational effort during the training phase. During testing, the method *physics-based generation of synthetic training data* is the most computationally efficient.

D. SUMMARY OF THE RESULTS OBTAINED AND DISCUSSION OF THEIR GENERALIZATION

Hybrid methods for combining entire DMs and PMs are qualitatively analyzed for their attributes regarding three aspects. These aspects are

- the ability to provide accurate RUL predictions in areas of the state space where training data is scarce,
- the influence of a high variance of the DM on the hybrid method itself, and
- the computational efficiency of the hybrid methods during the development and application phases due to their different use of PMs.

A summary of the results regarding these aspects for each hybrid method is given in Table 7. Overall, the method *physics-based generation of synthetic training data* possesses positive attributes with respect to all three aspects. *Residual modeling* is assigned two positive attributes. Only one positive attribute is determined for the methods *final hypothesis set validation*, *physics-based model as input*, and *physics-based model within the data-driven*.

In the case study on the RUL prediction of filters, the applied hybrid methods show different predictive accuracies. The results achieved by the hybrid methods possess no claim of transferability to prognostic applications in general. At most, they can be regarded as an indicator for the

applicability of individual methods. The identified attributes of the individual hybrid methods, in contrast, are based on a qualitative analysis of these methods and their functionality. Thus, the results listed in Table 7 characterize the hybrid methods beyond the case study and therefore have a claim to transferability to other prognostic applications. However, this transferability is limited to the qualitative statements and does not include any claims about the degree to which, for example, a hybrid method is more computationally efficient. Furthermore, although the three aspects considered are of high relevance, it must be assumed that there are many more challenges and requirements when developing a prognostic application. One such requirement, which goes beyond the scope of this study but is also highly relevant, could be the explainability of methods and models [73]. Thus, the choice of a prognostic method will not be solely based on the investigation of this work.

VI. CONCLUSION

The paper focuses on improving data-driven RUL predictions by integrating knowledge using hybrid methods. Advantages of this approach over purely data-driven approaches, where the effort of data generation is reduced, are highlighted. As this paper addresses the combination of complete DMs and PMs, a new categorization of corresponding hybrid methods is presented, and previous works utilizing these methods are identified.

Four hybrid methods are applied within a case study on the RUL prediction for a filter loading process. The results of these methods show that although the PM has a 30% higher RMSE than the GP, its incorporation as a complementary source of information in addition to data provides a significant improvement in predictive accuracy. In terms of the metrics RMSE and coverage rate, the method *physics-based generation of synthetic training data* performs best. The mean RMSE of this method is 10.4 s, while the GP shows an RMSE of 14.9 s. The coverage rate of the hybrid method is on average consistent with the targeted 90% and has a standard deviation of 2.4%, which is also significantly lower than the GP with 8.3%. Of the hybrid methods applied, *residual modeling* performs the worst. It has an 84% higher RMSE than *physics-based generation of synthetic training data*, making it the only hybrid method with a higher RMSE than the DM.

The analyses of hybrid methods based on the case study provide insights on individual hybrid methods that have not yet been studied in the literature. Three methods are identified that are less affected by a shortage of training data. These methods rely on the PM's prediction in areas with little or no training data and thus hold advantages over the other hybrid methods. This finding is supported by tests performed with different amounts of training data. Another observation of these tests is that, also in this case, the method *physics-based generation of synthetic training data* consistently has the lowest RMSE.

In addition, two hybrid methods are identified where the use of a DM with high variance is of particular disadvantage. One of them is the method *regions of competence*. The tests on this method with different DMs underline this. The greater the variance of the DM, the less likely it is that the correct model is selected during the competence assessment.

It is observed that, for the case study, the calculation of the PM's prediction requires a $5 \cdot 10^4$ times higher computational effort than the DM's prediction. Based on this, conclusions are drawn about the computational efficiency of the hybrid methods during the development and application phases.

Overall, attributes of the individual hybrid methods are discovered. Thereby, the method *physics-based generation of synthetic training data* provides the highest number of positive attributes. These attributes are, on the one hand, highly relevant for prognostic applications and, on the other hand, since they are derived from qualitative analyses, not limited to the scope of the case study.

REFERENCES

- [1] X. Jia, B. Huang, J. Feng, H. Cai, and J. Lee, "Review of PHM data competitions from 2008 to 2017," in *Proc. Annu. Conf. PHM Soc.*, Philadelphia, PA, USA, 2018, pp. 1–10.
- [2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
- [3] K. Javed, R. Gouriveau, and N. Zerhouni, "State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels," *Mech. Syst. Signal Process.*, vol. 94, pp. 214–236, Sep. 2017.
- [4] E. Hullermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, 2021.
- [5] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2318–2331, Jun. 2017.
- [6] T. Gao, C. Lu, and M. Hao, "Design requirements of PHM system fault diagnosis capability," in *Proc. Chin. Autom. Congr. (CAC)*, Hangzhou, China, Nov. 2019, pp. 2622–2626.
- [7] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for offline evaluation of prognostic performance," *Int. J. Prognostics Health Manag.*, vol. 1, no. 1, Mar. 2021, pp. 4–23.
- [8] P. M. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, p. 78–87, Oct. 2012.
- [9] O. F. Eker, F. Camci, and I. K. Jennions, "Physics-based prognostic modelling of filter clogging phenomena," *Mech. Syst. Signal Process.*, vol. 75, pp. 395–412, Jun. 2016.
- [10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [11] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, "An introductory review of deep learning for prediction models with big data," *Frontiers Artif. Intell.*, vol. 3, p. 4, Feb. 2020.
- [12] B. Scholkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, "Toward causal representation learning," *Proc. IEEE*, vol. 109, no. 5, pp. 612–634, May 2021.
- [13] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Eng. Appl. Artif. Intell.*, vol. 92, Jun. 2020, Art. no. 103678.
- [14] M. Braig and P. Zeiler, "Using data from similar systems for data-driven condition diagnosis and prognosis of engineering systems: A review and an outline of future research challenges," *IEEE Access*, vol. 11, pp. 1506–1554, 2023.
- [15] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A survey of deep active learning," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–40, Oct. 2021.
- [16] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.
- [17] M. Raissi and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *J. Comput. Phys.*, vol. 357, pp. 125–141, Mar. 2018.
- [18] Y. A. Yucesan and F. A. C. Viana, "A physics-informed neural network for wind turbine main bearing fatigue," *Int. J. Prognostics Health Manag.*, vol. 11, no. 1, pp. 1–17, 2020.
- [19] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, M. Walczak, J. Pfrommer, A. Pick, R. Ramamurthy, J. Garcke, C. Bauckhage, and J. Schuecker, "Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 614–633, Jan. 2023.
- [20] Y. Liu and K. Goebel, "Information fusion for national airspace system prognostics: A NASA ULI project," in *Proc. Annu. Conf. PHM Soc.*, Philadelphia, PA, USA, 2018, pp. 24–27.
- [21] R. Swischuk, L. Mainini, B. Peherstorfer, and K. Willcox, "Projection-based model reduction: Formulations for physics-based machine learning," *Comput. Fluids*, vol. 179, pp. 704–717, Jan. 2019.
- [22] R. Rai and C. K. Sahu, "Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus," *IEEE Access*, vol. 8, pp. 71050–71073, 2020.
- [23] F. Farbiz, M. S. Habibullah, B. Hamadicharef, T. Maszczyk, and S. Aggarwal, "Knowledge-embedded machine learning and its applications in smart manufacturing," *J. Intell. Manuf.*, vol. 33, pp. 1–18, Jun. 2022.
- [24] S. Hagemeyer, P. Zeiler, and M. F. Huber, "On the integration of fundamental knowledge about degradation processes into data-driven diagnostics and prognostics using theory-guided data science," in *Proc. PHM Soc. Eur. Conf.*, Turin, Italy, 2022, pp. 156–165.
- [25] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating scientific knowledge with machine learning for engineering and environmental systems," 2020, *arXiv:2003.04919*.
- [26] M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. van Vlijmen, M. D. Berliner, M. Z. Bazant, R. D. Braatz, W. C. Chueh, and B. D. Storey, "Perspective—Combining physics and machine learning to predict battery lifetime," *J. Electrochem. Soc.*, vol. 168, no. 3, Mar. 2021, Art. no. 030525.
- [27] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Rel. Eng. Syst. Saf.*, vol. 217, Jan. 2022, Art. no. 107961.
- [28] O. F. Eker, F. Camci, and I. K. Jennions, "A new hybrid prognostic methodology," *Int. J. Prognostics Health Manag.*, vol. 10, no. 2, pp. 1–13, Jun. 2019.
- [29] X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, "Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles," in *Proc. SIAM Int. Conf. Data Mining*, Calgary, Canada, 2019, pp. 558–566.
- [30] A. Dourado and F. A. C. Viana, "Physics-informed neural networks for corrosion-fatigue prognosis," in *Proc. Annu. Conf. PHM Soc.*, Scottsdale, AZ, USA, 2019, pp. 1–12.
- [31] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, "A digital-twin-assisted fault diagnosis using deep transfer learning," *IEEE Access*, vol. 7, pp. 19990–19999, 2019.
- [32] W. Li, S. Gu, X. Zhang, and T. Chen, "Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106904.
- [33] Y. Gao, X. Liu, H. Huang, and J. Xiang, "A hybrid of FEM simulations and generative adversarial networks to classify faults in rotor-bearing systems," *ISA Trans.*, vol. 108, pp. 356–366, Feb. 2021.
- [34] L. von Rueden, T. Wirtz, F. Hueger, J. D. Schneider, N. Piatkowski, and C. Bauckhage, "Street-map based validation of semantic segmentation in autonomous driving," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Milan, Italy, Jan. 2021, pp. 10203–10210.
- [35] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *J. Artif. Intell. Res.*, vol. 70, pp. 245–317, Jan. 2021.
- [36] F. Mangili, "Development of advanced computational methods for prognostics and health management in energy components and systems," Doctoral dissertation, Energy Dept., Politecnico di Milano, Italy, 2013.
- [37] S. K. Singh, R. Yang, A. Behjat, R. Rai, S. Chowdhury, and I. Matei, "PI-LSTM: Physics-infused long short-term memory network," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Boca Raton, FL, USA, Dec. 2019, pp. 34–41.

- [38] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Hybrid deep fault detection and isolation: Combining deep neural networks and system performance models," *Int. J. Prognostics Health Manag.*, vol. 10, no. 4, pp. 1–25, 2019.
- [39] N. Markovic, T. Stoetzel, V. Staudt, and D. Kolossa, "Hybrid fault detection in power systems," in *Proc. IEEE Int. Electric Mach. Drives Conf. (IEMDC)*, San Diego, CA, USA, May 2019, pp. 911–915.
- [40] D. Jung, K. Y. Ng, E. Frisk, and M. Krysanter, "Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation," *Control Eng. Pract.*, vol. 80, pp. 146–156, Nov. 2018.
- [41] S. Frank, M. Heaney, X. Jin, and J. Robertson, "Hybrid model-based and data-driven fault detection and diagnostics for commercial buildings," in *Proc. ACEEE Summer Study Energy Efficiency Buildings*, Pacific Grove, CA, USA, 2016, pp. 1–16.
- [42] P. Kundu, A. K. Darpe, and M. S. Kulkarni, "A review on diagnostic and prognostic approaches for gears," *Struct. Health Monitor.*, vol. 20, no. 5, pp. 2853–2893, Sep. 2021.
- [43] L. Kuncheva, *Combining Pattern Classifiers—Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2014.
- [44] R. Pan, D. Yang, Y. Wang, and Z. Chen, "Performance degradation prediction of proton exchange membrane fuel cell using a hybrid prognostic approach," *Int. J. Hydrogen Energy*, vol. 45, no. 55, pp. 30994–31008, Nov. 2020.
- [45] R. Penha, B. Upadhyaya, and J. W. Hines, "Monitoring and diagnosis of a heat exchanger using hybrid system modeling," U.S. Dept. Energy NEER Program, Dept. Nucl. Eng., Univ. Tennessee, Knoxville, TN, USA, Tech. Rep. DE-FG07-01ID14114/UTNE-07, 2004. [Online]. Available: https://www.researchgate.net/publication/2572701_Monitoring_And_Diagnosis_Of_A_Heat_Exchanger_Using_Hybrid_System_Modeling and <https://www.osti.gov/servlets/purl/832721>
- [46] M. Hoening, S. Hagemeyer, and P. Zeiler, "Enhancing remaining useful life-time prediction by an advanced ensemble method adapted to the specific characteristics of prognostics and health management," in *Proc. 29th Eur. Saf. Rel. Conf.*, Hannover, Germany, 2019, pp. 1155–1162.
- [47] Y. Zang, W. Shanguan, B. Cai, H. Wang, and M. G. Pecht, "Hybrid remaining useful life prediction method. A case study on railway D-cables," *Rel. Eng. Syst. Saf.*, vol. 213, Sep. 2021, Art. no. 107746.
- [48] P. Baraldi, F. Mangili, G. Gola, B. H. Nystad, and E. Zio, "A hybrid ensemble-based approach for process parameter estimation and degradation assessment in offshore oil platforms," *Int. J. Performability Eng.*, vol. 10, no. 5, pp. 497–509, 2014.
- [49] K. Goebel, N. Eklund, and P. Bonanni, "Fusing competing prediction algorithms for prognostics," in *Proc. IEEE Aerosp. Conf.*, Mar. 2006, p. 10.
- [50] Z. Skaf, O. F. Eker, and I. K. Jennions, "System component degradation: Filter clogging in a UAV fuel system," in *Proc. PHMA*, Jeju, Republik Korea, 2017, pp. 1–11.
- [51] O. Saarela, J. E. Hulsund, A. Taipale, and M. Hegle, "Remaining useful life estimation for air filters at a nuclear power plant," in *Proc. PHM Soc. Eur. Conf.*, Nantes, France, 2014, pp. 1–7.
- [52] D. Thomas, P. Penicot, P. Contal, D. Leclerc, and J. Vendel, "Clogging of fibrous filters by solid aerosol particles experimental and modelling study," *Chem. Eng. Sci.*, vol. 56, no. 11, pp. 3549–3561, Jun. 2001.
- [53] F. Mauthe, S. Hagemeyer, and P. Zeiler, "Creation of publicly available data sets for prognostics and diagnostics addressing data scenarios relevant to industrial applications," *Int. J. Prognostics Health Manag.*, vol. 12, no. 2, pp. 1–20, Nov. 2021.
- [54] S. Hagemeyer, F. Mauthe, M. Dutt, and P. Zeiler. (2021). *Preventive to Predictive Maintenance*. [Online]. Available: <https://www.kaggle.com/prognosticshse/preventive-to-predictive-maintenance/metadata>
- [55] V. Novick, P. Higgins, B. Dierkschiede, C. Abrahamson, W. Richardson, P. Monson, and P. Ellison, "Efficiency and mass loading characteristics of a typical HEPA filter media material," in *Proc. Nuclear Air Cleaning Conf.*, San Diego, CA, USA, 1990, pp. 782–798.
- [56] H. Bai, X. Qian, S. Fan, Y. Shi, Y. Duo, C. Guo, and X. Wang, "Theoretical model of single fiber efficiency and the effect of microstructure on fibrous filtration," *Ind. Eng. Chem. Res.*, vol. 60, no. 1, pp. 3–36, 2021.
- [57] S. Callé, P. Contal, D. Thomas, D. Bémer, and D. Leclerc, "Description of the clogging and cleaning cycles of filter media," *Powder Technol.*, vol. 123, no. 1, pp. 40–52, Feb. 2002.
- [58] W. Bergman, R. D. Taylor, H. H. Miller, A. H. Bierman, H. D. Hebard, R. A. Da Roza, and B. Y. Lum, "Enhanced filtration program at LLL. A progress report," in *Proc. Nuclear Air Cleaning Conf.*, Boston, MA, USA, 1978, pp. 1058–1099.
- [59] V. J. Novick, P. R. Monson, and P. E. Ellison, "The effect of solid particle mass loading on the pressure drop of HEPA filters," *J. Aerosol Sci.*, vol. 23, no. 6, pp. 657–665, Sep. 1992.
- [60] N.-H. Kim, D. An, and J.-H. Choi, *Prognostics and Health Management of Engineering Systems: An Introduction*. Berlin, Germany: Springer, 2016.
- [61] K. Goebel, M. Daigle, A. Saxena, S. Sankararaman, I. Roychoudhury, and J. Celaya, *Prognostics: The Science of Making Predictions*. Scotts Valley, CA, USA: CreateSpace Independent, 2017.
- [62] K. Goebel, B. Saha, and A. Saxena, "A comparison of three data-driven techniques for prognostics," in *Proc. 62nd Meeting MFPT Soc.*, Virginia Beach, VA, USA, 2008, pp. 119–131.
- [63] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [64] A. S. Britto, R. Sabourin, and L. E. S. Oliveira, "Dynamic selection of classifiers—A comprehensive review," *Pattern Recognit.*, vol. 47, no. 11, pp. 3665–3680, Nov. 2014.
- [65] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001.
- [66] R. Balestrierio, J. Pesenti, and Y. LeCun, "Learning in high dimension always amounts to extrapolation," 2021, *arXiv:2110.09485*.
- [67] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin, Germany: Springer, 2009.
- [68] L. Didaci, G. Giacinto, F. Roli, and G. L. Marcialis, "A study on the performances of dynamic classifier selection based on local accuracy estimation," *Pattern Recognit.*, vol. 38, no. 11, pp. 2188–2191, Nov. 2005.
- [69] Y. A. Yucesan, A. Dourado, and F. A. Viana, "A survey of modeling for prognosis and health management of industrial equipment," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101404.
- [70] S. Sankararaman, Y. Ling, C. Shantz, and S. Mahadevan, "Uncertainty quantification in fatigue crack growth prognosis," *Int. J. Prognostics Health Manag.*, vol. 2, no. 1, pp. 1–15, 2011.
- [71] Y. Yu, H. Yao, and Y. Liu, "Physics-based learning for aircraft dynamics simulation," in *Proc. Annu. Conf. PHM Soc.*, Philadelphia, PA, USA, 2018, pp. 1–9.
- [72] S. Hong, Z. Zhou, E. Zio, and W. Wang, "An adaptive method for health trend prediction of rotating bearings," *Digit. Signal Process.*, vol. 35, pp. 117–123, Dec. 2014.
- [73] S. Vollert, M. Atzmueller, and A. Theissler, "Interpretable machine learning: A brief survey from the predictive maintenance perspective," in *Proc. 26th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Vasteras, Sweden, Sep. 2021.

SIMON HAGMEYER received the B.Eng. and M.Sc. degrees in mechatronics from the Reutlingen University of Applied Science, Germany. He is currently pursuing the Ph.D. degree in cooperation with the University of Stuttgart. He is currently a Research Assistant with the Reliability Engineering and Prognostics and Health Management Research Group, Esslingen University of Applied Science, Germany. His research interests include diagnostics and prognostics, data-driven methods, hybrid methods, and data science.

PETER ZEILER received the Dipl.-Ing. and Dr.-Ing. degrees in mechanical engineering from the University of Stuttgart. He was the Team Leader of Research and Development with Robert Bosch GmbH. Thereafter, he became the Head of the Department of Reliability Engineering and the Department of Drive Technology, Institute of Machine Components, University of Stuttgart. In 2017, he joined the Esslingen University of Applied Sciences, Germany. Since 2021, he has been associated with the Faculty of Engineering Design, Production Engineering, and Automotive Engineering, University of Stuttgart. He is currently a Professor with the Faculty of Mechanical and Systems Engineering, Esslingen University of Applied Sciences, where he is the Head of the Reliability Engineering and Prognostics and Health Management Research Group. His research interests include prognostics and health management, methods for calculating reliability under operating conditions, modeling and simulation of reliability, and availability for complex systems. He is committed to Verein Deutscher Ingenieure (VDI) (Association of German Engineers), where he is also the Chairperson of the Safety and Reliability Advisory Board and the Technical Committees for Prognostics and Health Management and Monte Carlo Simulation. He is also the Conference Chair of the German Conference Fachtagung Technische Zuverlässigkeit (TTZ) (Technical Reliability Conference).

• • •