Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

ESTIMATING PARAMETERS WITH DATA FARMING FOR CONDITION-BASED MAINTENANCE IN A DIGITAL TWIN

Alexander Wuttke
Joachim Hunker
Markus Rabe

Department IT in Production and Logistics
TU Dortmund University
Leonhard-Euler-Strasse 5
Dortmund, 44227, GERMANY

Jan-Philipp Diepenbrock

IVA Schmetz GmbH

Holzener Strasse 39
Menden, 58708, GERMANY

ABSTRACT

Nowadays, vast amounts of data can be collected by sensors and used for data-driven approaches. Digital twins provide a framework to exploit these data for solving various issues. For many companies in the industrial sector, machine maintenance is a significant issue. Maintenance is essential for high overall equipment efficiency, but it can also be costly. Therefore, it should only be performed when necessary, based on the machine's condition. Condition monitoring is used to assess a machine's condition periodically, allowing for condition-based maintenance. In this paper, a simulation-based approach for parameter estimation is presented that contributes to condition-based maintenance. It introduces condition indicators for certain features of machines and demonstrates how to evaluate them using data farming, which employs simulation models as data generators. Additionally, the implementation of this approach in digital twins is discussed.

1 INTRODUCTION

Maintenance is a crucial part of industry to retain or restore the intended functionality of machines. Effective maintenance greatly contributes to resilient manufacturing systems. It can also be included in the after-sales portfolio of manufacturing companies for their complex products. After-sales services, such as maintenance, are a key point for customer retention, increasing customer loyalty and generating income through different service offerings. However, up to 20 % of the running costs in the industry are accounted for by maintenance (Isermann 2011). Therefore, the improvement of maintenance is of great interest. In the advent of the progressing digitalization, new opportunities are emerging with regard to data-driven approaches for maintenance. Data that are collected by sensors at the machine itself can be used to estimate the condition, enabling condition-based maintenance (CBM). The main goal for manufacturing companies is to predict when maintenance is necessary before the equipment fails and, thus, causes costs due to unexpected downtime. The overall goal is to shift from a reactive process by diagnostic methods such as failure mode and effect analysis to a proactive process design (Lee et al. 2014). Lately, the use of digital twins (DTs) as a framework to collect, process, and utilize data for CBM has been proposed (Aivaliotis et al. 2019).

A key aspect of CBM is an effective condition monitoring (CM) that is used to detect faults and, if possible, provide a detailed analysis for targeted maintenance actions. The utilized approaches for CM are usually based on analytical calculations or machine learning programs. However, analytical calculations might not be available for complex machines and machine learning programs are dependent on lots of data, including those of faults that might (and ideally should) not occur. This paper proposes an alternative approach for CM that is simulation-based and can be used for complex machines, even if there are little

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

to no data about faults and breakdowns. The novelty of this paper consists in the application of a data farming approach (Sanchez 2018), which uses a simulation model to generate data that can be used to estimate process parameters. In addition, the approach is embedded in a concept that breaks the assessment of conditions down to condition indicators (CIs), which are used to measure the health status of certain features of machines. Also, the implementation in a DT designed for maintenance that constitutes a suitable framework is discussed.

The structure of the paper is as follows. The background on CBM, CM, sensors, data farming, and DTs with an emphasis on maintenance is given in Section 2. The simulation-based approach is presented in Section 3. This includes the introduction to the concept of CIs, a detailed explanation of the simulation-based parameter estimation, and the integration of the concept in a DT for CBM. The approach is put into operation in Section 4, where a use case of an industrial furnace is presented. The paper closes in Section 5 with a summary and an outlook on further research opportunities.

2 RELATED WORK

In this section, related work and its background are introduced. First, a short overview of CBM as well as CM is given. Sensors and observational data are briefly discussed, as they enable CBM and CM. A framework for handling observational data can be realized by DTs, which are discussed with a focus on maintenance. The section closes with an introduction to data farming and design of experiments (DoE).

2.1 Condition-based Maintenance and Condition Monitoring

Maintenance is defined as the "combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function" (DIN 2018, p. 8). Important terms related to maintenance are "condition" and "faults". The condition of a machine refers to the current state regarding the health status, where a good health means the absence of faults and a bad health means the presence of faults. A fault is understood as the deviation of at least one feature of a machine from the acceptable standard condition, where a feature describes one specific characteristic of a machine. The assessment of the state of health does not have to be limited to the two states good and bad, but can contain gradations that enable a differentiated consideration with regard to the severity of the entirety of faults.

Maintenance is generally conducted according to one of the following strategies:

- *Reactive maintenance* (breakdown maintenance): Maintenance is only carried out when a fault has already occurred.
- *Preventive maintenance* (planned maintenance): The maintenance is carried out in periodic intervals, independently of the condition of the machine.
- *Condition-based maintenance* (predictive maintenance): Maintenance will be carried out when the condition of the machines indicates its necessity.

Out of these three strategies, CBM offers the greatest potential in terms of cost savings and efficiency, as the number of maintenance actions aims to be as low as possible but as high as needed (Isermann 2011). To enable CBM, the condition must be constantly analyzed. This process is called CM. Due to the complexity of modern machines and the amount of data involved, CM is mostly performed automatically. According to Jardine et al. (2006), the key steps of CBM are data acquisition followed by data processing and maintenance decision making based on the processed data.

Two fundamental tasks in CBM are fault detection and diagnosis (FDD) (Isermann 2011). Fault detection is used to recognize faults as soon and as precisely as possible. In fault diagnosis, more information on the fault is compiled, e.g., the cause of the fault or which specific part is faulty. Plenty methods are available for solving these tasks, which can be divided into the following two categories or are a hybrid form between both (Badihi et al. 2022):

- *Signal-based*: Methods use raw or processed sensor output with high resolution, e.g., limit checking and trend checking.
- *Model-based*: Methods utilize mathematical or physical models without necessarily high-resolution sensor data, e.g., parameter estimation and parity equations.

The creation of a model for model-based approaches requires a-priori knowledge of the machine at hand. This includes the physical and mathematical description of the processes taking place in the machine. With such models, data can be calculated and then compared to observational data from the real-world application by calculating the deviation. The deviation between calculated data and observational data is also referred to as the residual. The general process of model-based methods for fault diagnosis according to Jardine et al. (2006) consists of the three steps of generating residuals from observational data compared to the model data, the following evaluation of the residuals, and their diagnosis. The process is visually summarized in Figure 1.

Figure 1: General principle of model-based methods by Jardine et al. (2006).

A machine usually conducts processes that can be described by process parameters. The knowledge about specific values of process parameters can greatly enhance the ability to detect faults and diagnose their origin and severity. However, values of process parameters are often not known or not measurable. Then, the popular model-based method for FDD called parameter estimation can be used. This method is especially suitable to detect multiplicative faults, as other methods are not suitable to detect those (Isermann 2011). The general procedure of the method consists in choosing the process parameters of the model in such a way that the residuals compared to the observational data are minimal. For example, the minimization of the residuals can be achieved by the method of least squares (Gama 2010). Methods based on Monte Carlo simulation are also available, for which a detailed overview can be found in Luengo et al. (2020).

To quantify the physical processes of the real world, sensors are used. They observe machines and collect data, which are called observational data in this paper to emphasize their nature of origin. The physical values that can be detected by sensors include, e.g., temperature, pressure, and vibration. Sensors measure in intervals. Therefore, timestamps are often used in combination with each value. Data in such format are referred to as time series data. The collected observational data are never a perfect picture of reality. They are severely dependent on the quality of the sensors and influenced by disturbances to a varying degree. Other flaws of observational data include, e.g., missing values or outliers, which lower the overall data quality (García et al. 2015). To improve the data quality and minimize flaws, data cleaning is used. An overview of sensors in the industrial internet of things is reported by Misra et al. (2021). Another important aspect when utilizing sensors is data management. The amount of data that accumulates over time must be logged and stored. An extensive overview on the storage of data in databases is provided by Meier and Kaufmann (2019). A framework to handle and utilize observational data collected by sensors are DTs, which are introduced in the next section.

In an extensive literature review on the topic of CBM, Quatrini et al. (2020) have identified three challenges for future technological developments in this area. Firstly, specific degradation models for a broader variety of components of machines should be developed. Secondly, the integration of CBM in multi-component multi-sensor machines should be further investigated. Lastly, CBM in the context of human-machine interfaces should be researched. This paper contributes to the second challenge by proposing and discussing a simulation-based approach, especially suitable for multi-component machines, as further discussed in Section 2.3. Also, by using a DT as a framework, this paper contributes to the third challenge, as DTs are proven implementation of a human-machine interface (van der Valk et al. 2020).

2.2 Digital Twins for Condition-Based Maintenance

Currently, DTs are a popular topic in research and industry. Many definitions of a DT can be found, but no clear definition has been agreed upon. This work follows the basic understanding given by Grieves and Vickers (2017), who define a DT along three basic dimensions, which are a physical product in real space, a virtual product in virtual space, and a bi-directional connection of data and information between both physical and virtual product. This bi-directional connection distinguishes a DT from a digital shadow, which implements a uni-directional data flow (Kritzinger et al. 2018). Furthermore, services are also mentioned as another dimension of a DT (Enders and Hoßbach 2019). Examples for common services in DTs are maintenance, process optimization, and data analytics. A comprehensive overview and taxonomy of DTs is given by van der Valk et al. (2020).

According to Tao and Qi (2019), two essential elements are relevant for DTs. On the one hand, observational data of the physical product in the real space are relevant in order to parameterize and instantiate a model. Typically, sensors are used for data acquisition on a physical product. DTs offer a framework for data-driven methods by managing, storing, processing, and utilizing observational data. On the other hand, models for the physical product are necessary that are used in the virtual space. The models are often simulation models, e.g., for multi-physics simulation, in order to represent the temperatures and pressure of an industrial furnace for FDD. Simulation enables DTs to utilize data that cannot be measured by sensors and, therefore, increase the data basis for data-driven approaches. Further information on simulation is given in the next section.

2.3 Simulation, Data Farming, and Design of Experiments

Simulation is a method that models systems in so-called simulation models and experiments with them to gain insights to reality (Verein Deutscher Ingenieure 2014). Especially for complex systems, where analytical methods are hardly or not at all applicable, simulation is a well-established method. While simulation is a powerful tool for analyzing complex systems, the simulation models can also be used for data farming. The term was coined in 1998 in a project called "Albert" by the United States military (Brandstein and Horne 1998). Data farming is a process that utilizes a simulation model and an efficient experiment design to conduct large-scale experiments with the intention to support decision makers by analyzing and visualizing the output, called response (Sanchez 2018). In particular, data farming is a process that can be described by a procedure model, dividing a data farming study into a series of different phases. Established procedure models are the "Loop of Loops" (Horne and Seichter 2014), "Knowledge Discovery in Simulation Data" (Feldkamp et al. 2015), and "Farming for Mining" (Hunker et al. 2022), with the latter two in particular establishing concepts from knowledge discovery and machine learning for data farming in the domains of production as well as logistics.

One of the core phases across all procedure models to conduct purposeful large-scale experiments with a simulation model is DoE. The result of an experiment design is a design matrix X, which specifies the values for the parameters of a simulation model, called factors, k (Kleijnen 2015). A single row, x, in X is called a design point, which sets the values, called levels, for each factor used in a single simulation run, $i, i = 1..n$, with n denoting the number of runs. A single column, thus, describes the different values of a single factor k for every design point. Furthermore, a replication is the repetition of X using a different starting value, called seed (Sanchez et al. 2020).

To create a design matrix, different designs have been proposed, such as grid-based designs and space-filling designs. Typical factorial designs such as 2^k (coarse grids) or n^k (finer grids) quickly prove unsuitable for complex simulation models due to the high number of factors, which lead to a large number of simulation runs that must be performed (Cioppa and Lucas 2007). This problem is known under the term "curse of dimensionality", which, generally speaking, describes the exponential correlation between factors and the amount of data that is required to adequately explore the input space. Space-filling designs, such as latin hypercubes, are much more efficient for quantitative factors. They offer a reduced number of simulation runs while ensuring balanced values of factors as well as a balanced model output (Sanchez et al. 2020). Many variants of latin hypercubes have been proposed, for example, nearly orthogonal latin hypercubes (Cioppa and Lucas 2007) and stacked nearly orthogonal latin hypercubes (Parker 2022). Latin hypercubes have gained a high popularity among researchers and practitioners due to their versatile, general-purpose design and their flexibility in fitting metamodels for output analysis (Kleijnen 2015; Sanchez et al. 2020).

Following Sanchez and Lucas (2002), the overall goals of a data farming study are to gain a deep understanding of the simulation model and underlying system, robust decision-making, and comparing as well as evaluating different decision alternatives. Typically, response-surface methods (e.g., first-order or second-order polynomial metamodels) or data mining (e.g., classification and segmentation) are used to identify important factors and their interaction (Sanchez et al. 2020).

3 CONCEPT FOR SIMULATION-BASED CONDITION MONITORING FOR CONDITION-BASED MAINTENANCE IN A DIGITAL TWIN

This section is divided into three parts. In Section 3.1, an overview is given of the concept for CM based on the processing of CIs that describe individual features of a machine. With the general concept in mind, Section 3.2 introduces a simulation-based approach on how to determine CIs with the estimation of process parameters. Finally, the integration of the concept in a DT of a machine is discussed in Section 3.3.

3.1 Condition Monitoring Exploiting Condition Indicators for Condition-based Maintenance

According to the authors' understanding, CM is an iterative process that is capable of assessing the condition of machines over a period of time and providing condition reports based on quantifiable data. The process includes perpetual fault detection and, if possible, the provision of further information on the fault. This paper proposes to determine condition reports based on CIs that are used to monitor the health of specific features of machines. The authors define CIs as a measurable characteristic or feature of a machine or process that changes with machine or process degradation and can be monitored to indicate the onset or severity of machine or process degradation. For example, assessing the condition of an industrial furnace can be based on CIs like heating-up time and vibration. The CIs are determined on the basis of observational data collected by sensors during the machine's operation. Moreover, additional data, e.g., data of enterprise resource planning systems or simulation results, can optionally be used for the CIs. The assessment of both, the condition report and the CIs, is conducted in dedicated assessment units, which encapsulate the assessment logic that is used to determine an output based on the available input data. The concept of determining the condition report based on CIs is visualized in Figure 2.

A crucial part of the introduction of a CM system according to this concept is the selection of CIs. Therefore, an in-depth analysis of the machine at hand should be made regarding how to quantify the condition and, thus, to decide which CIs are needed. A higher number of CIs increases the information base for the condition report on the one side, but on the other side also increases the required amount of observational data and the complexity of the condition assessment unit (CAU). Thus, it could be beneficial to limit the CIs to the most crucial ones. This also aids to avoid the curse of dimensionality, since fewer factors are taken into account (see Section 2.3). Depending on the complexity of the machine, decomposition into components with their own CIs can be appropriate. With the CIs identified, it can be derived which observational data of sensors and which additional data are needed for their assessment.

The CIs are used in the CAU together with additional data. The assessment is not necessarily restricted to the most recent data, but can also request or maintain data collected earlier for, e.g., trend analysis. The implementation can be realized in different ways that are to be chosen by the users. Approaches range from simple ones, e.g., defining corridors for each CI that may not be exceeded, to more complex ones, e.g., machine learning models like deep neural networks, which determine a quantifiable condition report. The CIs assessment units (CIAUs) generally have to use observational data as a basis and can optionally depend on additional data. Just as in the CAU, previous data can be utilized. On the output side, CIs as

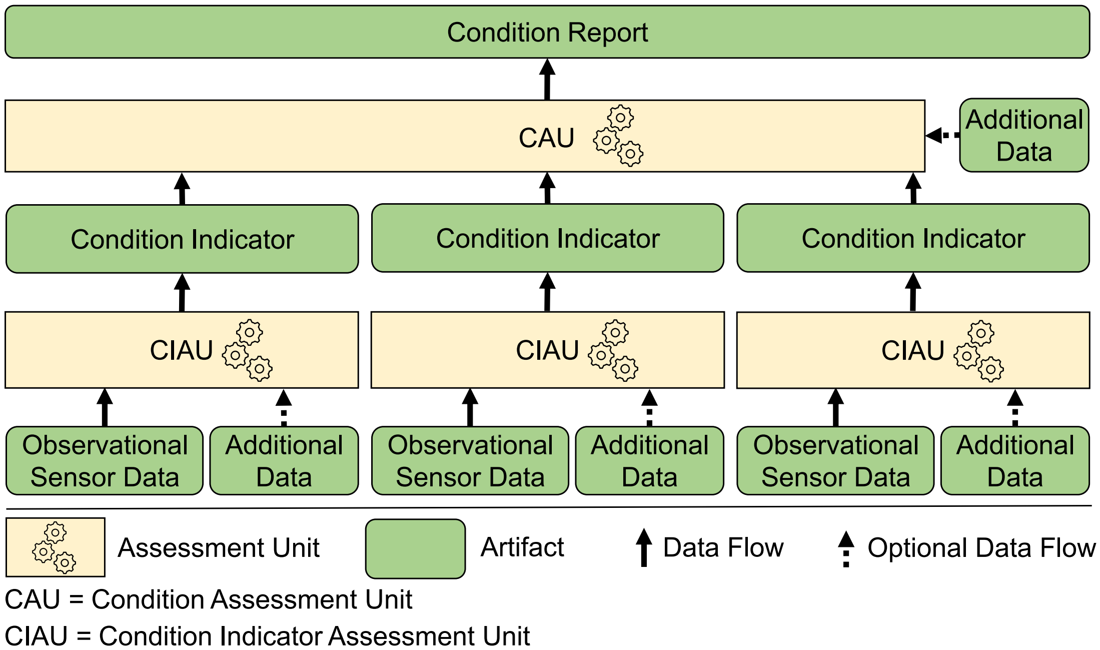

Figure 2: Assessment of condition reports based on condition indicators.

quantifiable values must be yielded. Similar to the CAU, the implementation of CIAUs can be achieved by approaches of varying complexity. Given the multitude of implementation possibilities for the CAU and CIAUs, this paper acknowledges the limitations of exploring each one in detail and defers this aspect to future research. However, in the next section, one novel approach for the implementation of CIAUs based on simulation is introduced.

3.2 Using Data Farming to Estimate Process Parameters for Condition Indicators

As described in the previous section, CIs are used to process observational data of machine features for FDD. The processing can be achieved by implementing parameter estimation (see Section 2.1). Although analytical approaches for solving these tasks already exist, they are limited to less complex models and require a high implementation effort. This assessment is consistent with statements in the literature. For example, (Jardine et al. 2006, p. 1494) state that “explicit mathematical modeling may not be feasible for complex systems since it would be very difficult or even impossible to build mathematical models for such systems”. One possibility to investigate complex models with emergent behavior where analytical methods reach their limits is simulation (see Section 2.3). In this section, a simulation-based approach for estimating process parameters is introduced. The approach is based on the assumption that if the simulation result data are similar to the collected observational data, the parameters used for the simulation resemble the real process parameters closely.

Thus, the approach is to determine the parameters for simulation that lead to the best-matching generated data compared to the observational data. However, it is not possible to calculate the optimal parameters for the best possible match by simulation in advance. Therefore, several parameter sets must be tested and the corresponding generated data compared to the sensor data to find a good solution. The testing of parameter sets should be a guided process aided by a proper DoE, where parameters are referred to as factors and a set of values for these factors as a design point (see Section 2.3). Depending on the number of factors and the complexity of the simulation model, high performance computing might be required to achieve results in an acceptable amount of time. The simulation of one design point will result in simulation result data. The combination of all simulation result data of all design points is seen as the solution space. The process to generate the solution space corresponds to the data farming procedure model of Hunker et al. (2022). To the best of the authors' knowledge, the usage of data farming for estimating process parameters is a novelty.

After the solution space is generated, the position of the observational data in it must be determined. The position in the solution space means the relative location of observational data to the simulation result data of all design points. This can be done by brute force, comparing the observational data to each set of simulation result data for each design point. Heuristic approaches are viable as well. The similarity can be determined using measures such as the mean absolute error (mae) on a number of sampling points in time. These simulation result data with the greatest similarity approximate the real behavior of the system best. Since each set of simulation result data is linked to a design point and its values for factors, which are the parameters to estimate, conclusions can be drawn. Depending on the problem, it could be sufficient to just use the parameters of the best or a mean of a group of the design points with best-matching simulation result data. A more-complex analysis could be conducted by identifying and analyzing clusters. There is also the possibility that none of the design points is a good match for the given observational data. In that case, the real process is not explainable by the simulation model under the given DoE. This means that an unforeseen behavior was detected, the simulation model is faulty, or the DoE was not conducted carefully enough. The similarity matching between observational data and the simulation result data of a design point is schematically visualized in Figure 3 on an example of an evacuation process (creation of a vacuum).

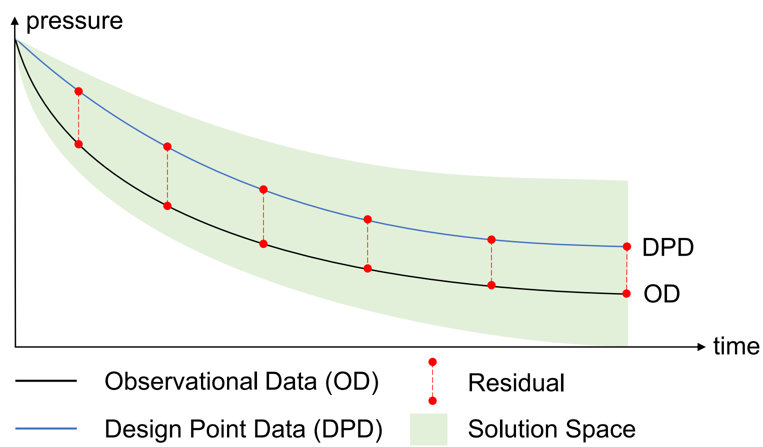

Figure 3: Qualitative figure of similarity matching in an evacuation process.

In Figure 3, the solution space is depicted as an area that is formed by the simulation result data of all design points in the DoE. One of these design points is highlighted in the figure. Given that observational data were collected, the similarity to this specific set is determined by measuring the residual at certain sampling points, which is indicated with the dotted line in the figure.

From this approach, requirements for the observational data are derived. First, they must be comparable to the simulation result data in the design points. Generally speaking, observational data are univariate time series with either varying or equally-spaced points of time. Given that simulation result data are available on the same points in time, similarity can be calculated according to measures like the total absolute error for all elements in the time series, the mae or their Euclidean distance. When the points of time in the sensor and simulation time series differ, other means like dynamic time-warping can be used (Gama 2010). It is also important only to select a suitable and interesting subset of data from the continuous flow of sensor data. For example, a subset of data for a CI for heating systems should include data from the start of a heating process until the end rather than including data outside the heating process. Moreover, the subset data should have a high quality that can be achieved by cleaning the data to handle missing, noisy, and wrong data (see Section 2.1).

There are also requirements on the simulation model. It must be able to generate time series data at specific points in time, according to the available observational data and the chosen method of similarity measurement. Therefore, time-continuous simulation models are to be used preferably for many technical systems, as data can be simulated for every given point in time. For other systems, like in material logistics and especially in supply chains, time-discrete or discrete event simulation are more efficient. Another requirement is that the simulation model must only use the process parameters that are degradable and that should be estimated as factors. If other process parameters, which should not be estimated, change in value, the simulation model must be adapted accordingly and a new data farming process for a solution space must be conducted that reflects the changes. For example, the heating process of an industrial furnace is dependent on the volume of air in the heating chamber. Although the volume of the heating chamber is constant, the volume of air will decrease depending on the volume of the batch, which is process-dependent. Therefore, a new variant of the simulation model must be created that adopts the volume of air, which is a form of dynamic simulation model tuning (Lugaresi and Matta 2020).

The approach presented in this section is summarized in Figure 4.

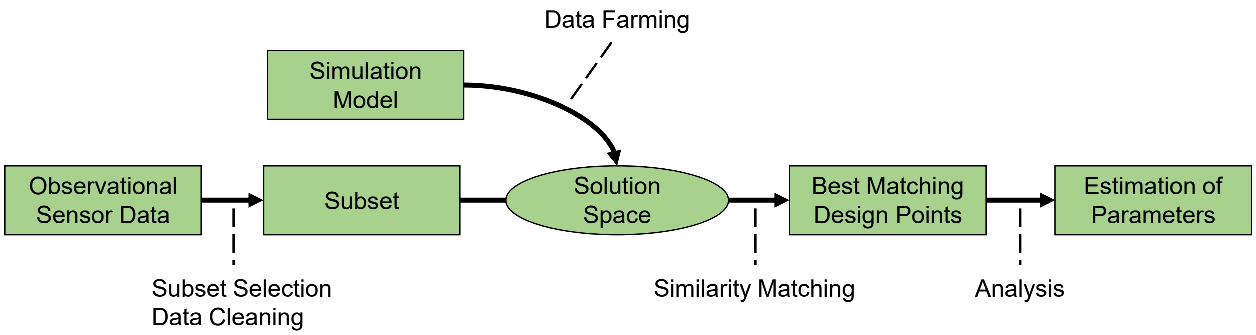

Figure 4: Approach of using data farming to estimate parameters.

As can be seen from Figure 4, the approach requires observational data from sensors and a simulation model as inputs. This corresponds to the inputs shown in Figure 2. A popular framework for handling data and models, including observational data and simulation models, are DTs. In the next section it is shown how the approach and the general concept of CM based on CIs can be integrated into a DT.

3.3 Integration into a Digital Twin

A proven framework in science and industry to leverage observational data collected by sensors and models is the DT (see Section 2.2). In the concept of this paper, DTs provide the inputs requested by the CAU and the CIAUs. The CIAUs request the observational data from the DT before selecting a suitable subset. Depending on the implementation of the DT, data cleaning steps might not be needed if similar steps are already taken when importing the data into the DT. The DT also provides simulation models required for data farming or the capabilities to use simulation as a service to generate the needed data. If dynamic simulation model instancing is needed, it must be ensured that the DT has this capability. An advantage that can be exploited is the sharing of simulation models for the use for multiple CIAUs at once. This also applies to sharing the simulation models with other software components.

DTs not only help with the input but also aid in utilizing the output. The concept of CM based on CIs can be recognized as a service inside the DT, making it easier to be used for further processes. This will mainly refer to CBM. However, other applications, like asset management, process optimization, and general condition-based decision making, can also utilize the output. Furthermore, the utilization of this output extends beyond automated processing, as it enables accessibility of information in the human-machine interface of the DT. By providing value-added information for CBM, the concept presented in this paper contributes to targeted maintenance measures, which are applied on the real space of the DT. Although the use of the concept as part of a DT is encouraged by the authors, it can also be used in a stand-alone system or incorporated in other software landscapes, which likely requires a higher implementation effort compared to the integration in an existing DT.

4 USE-CASE-BASED PROOF OF CONCEPT

In this section, the previously introduced simulation-based approach is tested on a case study in the domain of industrial furnaces. It will be shown how a CI for the vacuum process in a heating chamber can be calculated with the use of data farming.

The industrial furnace examined in this example is used for the post processing of steel products like annealing, tempering, and hardening. The furnaces consist of a heating chamber for the charge, heating elements, fans for better distribution of possibly present gassing atmospheres as well as cooling elements. To prevent the steel from reacting with the oxygen in the air when heated up, the furnace contains a vacuum pump. It is used for initial evacuation of the furnace and during the ongoing process.

The concept shown in Section 3 proposes to analyze the machine regarding its crucial CIs. For this use case, only one exemplary CI will be examined, although an industrial furnace would require more for an appropriate condition report. In the following, the initial evacuation process in a heating chamber of an industrial furnace will be assessed with a CI. As degradable factors for this CI, the speed of the vacuum pump and the leakage rate through the hull and piping were chosen. The condition of the evacuation process has proven to have a crucial impact on the entire system in industrial practice as the furnace's process quality is highly dependent on the vacuum and structural damage at the furnace as well as vacuum pump failures can be easily detected.

A mathematical description can be used to characterize the physical processes involved in the evacuation process including the degradable factors identified for the CI. The change of the pressure p [mbar] in the heating chamber is dependent on the chamber volume V [m^3], the speed of the pump s [$\frac{\text{m}^3}{\text{s}}$], the leakage rate l [$\frac{\text{mbar}}{\text{s}}$], and the pressure difference of inside and outside the heating chamber $p_{at} - p$, where the standard atmosphere pressure is defined as $p_{at} = 1013.25$ mbar. The mathematical description results in the following Equation (1):

$$\frac{\partial p}{\partial s} = -p\frac{s}{V} + l\sqrt{\frac{p_{at} - p}{\text{mbar}}} \tag{1}$$

To conduct data farming, a simulation model must be created that can be used to simulate Equation (1). The results of the simulation must be available for each point of time i for which observational data exist. Therefore, a time-continuous simulation model is built. In order to evaluate Equation (1) in the simulation model, the equation must be transformed to be incremental with time steps of Δt_i. Utilizing the Euler method, the following Equation (2) can be obtained that is used in the simulation:

$$p_i = p_{i-1}e^{\left(-\frac{s}{V}\Delta t_i\right)} + l\sqrt{\frac{p_{at} - p_{i-1}}{\text{mbar}}}\Delta t_i \tag{2}$$

Equation (2) is subject to the assumption that the pump speed is approximately constant throughout the time increment Δt_i. The volume of the chamber was chosen according to the real industrial furnace minus the batch volume.

The CI will be calculated for two evacuation processes with recorded observational data and not in the running processes. The observational data for both processes were collected in a real industrial application. The observational data are referred to as time series 1 (TS1) and time series 2 (TS2). The data for TS1 and TS2 were selected as subsets from the recorded data, beginning with the start of the vacuum pump until the desired level of vacuum was reached. Necessary data cleaning steps were conducted on the subsets as well. In the data farming, only two factors are considered. Therefore, a full factorial DoE with 1,000 levels for each factor in an appropriate range was used, resulting in 1,000,000 sets of simulation result data in the solution space. More information on the DoE can be taken from Table 1. As a measure for the distance between the observational time series data and the design points, the mae was chosen.

Table 1: Full factorial DoE for the generation of the solution space.

Factor	Lowest	Highest	Levels
pump speed (s)	0.025	0.15	1,000
leakage rate (l)	0.0	0.35	1,000

The experiments, including the generation of a solution space with data farming and the matching of both time series data to each design point in it, were conducted on a personal computer with standard hardware in less than two seconds. The results with the best matching design points for TS1 are given in Table 2 and those for TS2 are given in Table 3.

Table 2: Best matching design points for TS1.

rank	s	l	mae
1	0.0882	0.0181	13.0346
2	0.0882	0.0182	13.0350
3	0.0881	0.0181	13.0359
4	0.0879	0.0183	13.0377
5	0.0882	0.0182	13.0377

Table 3: Best matching design points for TS2.

rank	s	l	mae
1	0.0856	0.3101	15.0442
2	0.0856	0.3102	15.0450
3	0.0855	0.3101	15.0459
4	0.0856	0.3103	15.0477
5	0.0878	0.3102	15.0477

Table 2 and Table 3 show similar results for the speed of the pump, indicating not much wear and tear. The estimated values for the leakage rate are roughly seventeenfold higher in TS2 than in TS1, which indicates a change in the health status of the machine. The mae of both time series of 13 and 15 millibars are in an acceptable range. The results of the experiments were evaluated in two ways. First, a comparison with existing approaches in the company, which provided the data for this use case, was made. The existing approach makes use of machine learning that trained a model, which analyses for undesired trends and conducts limit checking for unusual values of data. The model creates alarms and assesses the health condition in ten discrete steps. The application to TS1 and TS2 with the machine learning program yielded similar results to the experiments in this section. The machine learning program rated TS1 with a health level of 9, which is the second best, and TS1 was rated 1, which is the second worst rating. The second way of evaluation was conducted by analyzing the real situation. While the customer noticed no problems in the time period while and around TS1, there were noticeable problems just before and after the data of TS2 were collected. After examining the industrial furnace, it was found that a crack in the retort had contributed massively to the inability to build up an effective vacuum.

Compared to other fault detection methods, just as the briefly described machine learning program of the company in Section 3.2, the simulation-based approach provides additional information about specific process parameters, allowing for a better fault prognosis. In the experiment, the machine learning program did not help in fault prognosis. However, the experiments in this section clearly show an increase in the leaking rate in TS2 compared to TS1. Given low-complexity simulation models, like the one used in the experiments, the parameter estimation can be achieved quickly and can contribute to real-time applications. The predominant share of work is the creation of a suitably detailed simulation model. Given that a simulation model already exists, the given approach can be implemented fast and easily. Although this paper proves the general feasibility of using data farming for parameter estimation, it is only shown for a rather simple time-continuous simulation model. For such cases, analytical methods could have also be used to estimate parameters. However, the authors assume that the simulation-based approach should be transferable to more complex problems, for which analytical methods will not be feasible anymore. The verification of this assumption should be carried out in subsequent research.

5 SUMMARY AND OUTLOOK

In this paper, a simulation-based condition monitoring approach for use in CBM in the framework of DTs is introduced. After the theoretical background and related work were given, a concept for CM based on CIs is presented. CIs describe the health of certain features in machines and are based on observational data. A simulation-based approach for estimating parameters is introduced that makes use of data farming. As a framework, the integration of the whole concept into a DT is discussed. Finally, the approach of simulation-based parameter estimation is tested on a use case of an industrial furnace. Exemplarily, one CI for the vacuum process is assessed with real-world data. It is shown how a physical simulation model could be built and how to utilize it in a data farming-process. The CI is assessed for two processes, in which one process was known to be in good health and the other one in bad health. It was shown that the novel simulation-based parameter estimation is able to detect the fault and, additionally, gave a good prognosis by specifically stating the estimated parameters.

In further research, the novel approach should be examined under different aspects. First, the parameter estimation with data farming should be tested on more-complex simulation models with more than two process parameters. Second, the assessment units of CIs should be further investigated. Finally, the assessment of the condition report should also be investigated in detail to allow for a holistic proof of concept of the proposed concept in this paper.

REFERENCES

Aivaliotis, P., K. Georgoulias, and G. Chryssolouris. 2019. “The Use of Digital Twin for Predictive Maintenance in Manufacturing”. *International Journal of Computer Integrated Manufacturing* 32(11):1067–1080.

Badihi, H., Y. Zhang, B. Jiang, P. Pillay, and S. Rakheja. 2022. “A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis”. *Proceedings of the Institute of Electrical and Electronics Engineers, Inc* 110(6):754–806.

Brandstein, A. G., and G. E. Horne. 1998. *Data Farming: A Meta-technique for Research in the 21st Century*. Quantico, Virginia: Marine Corps Combat Development Command Publication.

Cioppa, T. M., and T. W. Lucas. 2007. “Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes”. *Technometrics* 49(1):45–55.

DIN 2018. *DIN EN 13306:2018-02, Maintenance Terminology; Trilingual Version*. Berlin: Beuth Verlag.

Enders, M. R., and N. Hoßbach. 2019. “Dimensions of Digital Twin Applications. A Literature Review”. In *Proceedings of the 25th Americas Conference on Information Systems*. August 15th–17th, Cancun, Mexico, 1–10.

Feldkamp, N., S. Bergmann, and S. Straßburger. 2015. “Visual Analytics of Manufacturing Simulation Data”. In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, Chan, W. K. V., I. Moon, Roeder, T. M. K., C. Macal, and M. D. Rossetti, 779–790. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gama, J. 2010. *Knowledge Discovery from Data Streams*. Boca Raton, FL: CRC Press.

García, S., J. Luengo, and F. Herrera. 2015. *Data Preprocessing in Data Mining*. Cham: Springer International Publishing.

Grieves, M., and J. Vickers. 2017. “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems”. In *Transdisciplinary Perspectives on Complex Systems*, edited by F.-J. Kahlen, S. Flumerfelt, and A. Alves, 85–113. Cham: Springer International Publishing.

Horne, G., and S. Seichter. 2014. “Data Farming in Support of NATO Operations – Methodology and Proof-of-Concept”. In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and Miller J. A., 2355–2363. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hunker, J., A. A. Scheidler, M. Rabe, and H. van der Valk. 2022. “A New Data Farming Procedure Model for a Farming for Mining Method in Logistics Networks”. In *Procedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Peng Y., S. Shashaani, E. Song, C. G. Corlu, L. H. Lee, E. P. Chew, T. Roeder, and P. Lendermann, 1461–1472. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Isermann, R. 2011. *Fault Diagnosis of Technical Processes, Applications: Model Based Methods for Actuators, Sensors, Drives, Machinery, and Industrial Plants*. Berlin: Springer.

Jardine, A. K., D. Lin, and D. Banjevic. 2006. “A Review on Machinery Diagnostics and Prognostics Implementing Condition-based Maintenance”. *Mechanical Systems and Signal Processing* 20(7):1483–1510.

Kleijnen, J. P. 2015. *Design and Analysis of Simulation Experiments*. Cham: Springer International Publishing.

Kritzinger, W., M. Karner, G. Traar, J. Henjes, and W. Sihn. 2018. “Digital Twin in Manufacturing: A Categorical Literature Review and Classification”. *IFAC-PapersOnLine* 51(11):1016–1022.

Lee, J., F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel. 2014. "Prognostics and Health Management Design for Rotary Machinery Systems Reviews, Methodology and Applications". *Mechanical Systems and Signal Processing* 42(1):314–334.

Luengo, D., L. Martino, M. Bugallo, V. Elvira, and S. Särkkä. 2020. "A Survey of Monte Carlo Methods for Parameter Estimation". *EURASIP Journal on Advances in Signal Processing* 2020(1):1–62.

Lugaresi, G., and A. Matta. 2020. "Generation and Tuning of Discrete Event Simulation Models for Manufacturing Applications". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, and T. R. T. Roeder, 2707–2718. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Meier, A., and M. Kaufmann. 2019. *SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management*. Wiesbaden, Germany: Springer Fachmedien.

Misra, S., C. Roy, and A. Mukherjee. 2021. *Introduction to Industrial Internet of Things and Industry 4.0*. Boca Raton, FL: CRC Press.

Parker, J. D. 2022. *Extending and Improving Designs for Large-Scale Computer Experiments*. Dissertation, Naval Postgraduate School, Monterey.

Quatrini, E., F. Costantino, G. Di Gravio, and R. Patriarca. 2020. "Condition-Based Maintenance. An Extensive Literature Review". *Machines* 8(2):31–59.

Sanchez, S. M. 2018. "Data Farming: Better Data, not Just Big Data". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, A. Mustafee, S. J. Skoogh, and B. Johansson, 425–439. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sanchez, S. M., and T. W. Lucas. 2002. "Exploring the World of Agent-based Simulations: Simple Models, Complex Analyses". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 116–126. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sanchez, S. M., P. J. Sanchez, and H. Wan. 2020. "Work Smarter, not Harder: A Tutorial on Designing and Conducting Simulation Experiments". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, and T. R. T. Roeder, 1128–1142. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tao, F., and Q. Qi. 2019. "Make More Digital Twins". *Nature* 573(7775):490–491.

van der Valk, H., H. Haße, F. Möller, M. Arbter, J.-L. Henning, and B. Otto. 2020. "A Taxonomy of Digital Twins". In *Proceedings of the 26th Americas Conference on Information Systems*. August 10th–14th, Salt Lake City, Utah, 1–10.

Verein Deutscher Ingenieure 2014. *VDI 3633 – Simulation of Systems in Materials Handling, Logistics and Production: Part 1, Fundamentals*. Berlin, Germany: Beuth.

AUTHOR BIOGRAPHIES

ALEXANDER WUTTKE is a researcher at the department IT in Production and Logistics at the TU Dortmund University. He holds a Master of Science in Mechanical Engineering with a focus on IT in Production and Logistics from TU Dortmund University. His research interests are digital twins and condition-based maintenance. His email address is alexander2.wuttke@tu-dortmund.de.

JOACHIM HUNKER is a researcher at the department IT in Production and Logistics at the TU Dortmund University. He holds a Master of Science in Logistics, Infrastructure, and Mobility with a focus on IT in Logistics from the Technical University of Hamburg. He graduated with a master thesis on a hybrid scheduling approach of assembly lines of car manufacturers. His research focuses on simulation-based data generation and data analytics in logistics. His email address is joachim.hunker@tu-dortmund.de.

MARKUS RABE is a full professor for IT in Production and Logistics at the TU Dortmund University. Until 2010 he had been with Fraunhofer IPK in Berlin as head of the corporate logistics and processes department, head of the central IT department, and a member of the institute direction circle. His research focus is on information systems for supply chains, production planning, and simulation. Markus Rabe is vice chair of the "Simulation in Production and Logistics" group of the simulation society ASIM, member of the editorial board of the Journal of Simulation, member of several conference program committees, has chaired the ASIM SPL conference in 1998, 2000, 2004, 2008, and 2015, Local Chair of the WSC'2012 in Berlin and Proceedings Chair of the WSC'2018 and WSC'2019. More than 250 publications and editions report from his work. His email address is markus.rabe@tu-dortmund.de.

JAN-PHILIPP DIEPENBROCK is an engineer at IVA Schmetz GmbH, Germany. He holds a Ph.D. from the University of Magdeburg and researched in the field of simulation and data processing. He also holds a diploma in Physics from TU Munich University and studied at the University of Cambridge under a scholarship. His research focuses on machine learning for condition monitoring and condition-based maintenance. His email address is jan.diepenbrock@iva-schmetz.de.