Simulation-Based Inference for the Amortized Determination of Physical Model Parameters of a Mechanical System for Diagnostics and Prognostics

Cédric Schenkel[1,2] and Kai Hencken[3]

[1,3] *ABB Switzerland Ltd, Corporate Research Center, Baden-Dättwil, 5405, Switzerland*
cedric.schenkel@gmail.com
kai.hencken@ch.abb.com

[2] *Current address: University of Southampton, Southampton, SO17 1BJ, United Kingdom*
C.J.Schenkel@soton.ac.uk

ABSTRACT

Many diagnostic and prognostic applications rely on complex measurements, including time series and high-dimensional data. In a common approach, one extracts key features that capture the system degradation while ignoring nuisance effects. Physical system properties are of interest due to their (direct) relation and relevance to degradation and failure modes, often allowing for superior interpretability compared to purely data-driven approaches. However, determining them from the observed data is difficult due to the inherent non-identifiability in already rather simple models. Recently developed simulation-based inference (SBI) approaches, based on neural posterior estimates (NPE) and conditional invertible neural networks (cINN), allow the incorporation of domain knowledge in the form of simulation capabilities to extract model parameters as physics-based features for diagnostics and prognostics. This is demonstrated in the case of a mechanical actuator that operates medium voltage breakers. Simulations of a simplified multi-body model are used as input for the training of a cINN that not only provides a set of physical parameter values but also their respective uncertainties. By using simulated data as synthetic measurements and conducting a number of statistical checks, the performance of the trained cINN is confirmed. We demonstrate that accurate multi-body parameter estimation is possible for some parameters, whereas those that cannot be identified from either the opening or closing motion remain distributed according to the prior distribution, without significantly affecting the others. We further show that the opening and closing motions are sensitive to different parameters, complementing each other in this respect. Data from a real mechanical endurance test are used to demonstrate the method's effectiveness in a real-world application. Its integration into a diagnostics or prognostics framework is discussed as an outlook.

1. INTRODUCTION

With the development of new sensors, the amount of data that can be measured for diagnostics and prognostics has increased dramatically. As a result, data are often complex, comprising high-resolution time series or high-dimensional representations. Key features are then extracted from the data to track relevant degradation or fault initiation in the asset. Whereas these features should be sensitive to those effects, they should at the same time not be influenced by other effects, for example, changes in the operational or environmental conditions.

Finding such features in systems exhibiting complex dynamics is challenging, especially when dealing with highly reliable systems, where failure data are not abundant enough to detect such features through a purely data-driven approach. Consequently, incorporating domain knowledge about the asset is frequently necessary to guide the feature selection. A common way to do this is to use physics-based simulation models. These are often already available from product development and are therefore expected to incorporate key knowledge about the system dynamics. The parameters of these models can then naturally be used as features. Their link to system degradation is often straightforward, whereas dependencies on other parameters, unrelated to degradation or faulty behaviour, can be ignored. In some cases, parameter changes can also be directly associated with specific failure modes. Finally, in connection with prognostic applications, the evolution of such parameters, as well as critical thresholds, can be found more easily due to this interpretability.

For several reasons, one often faces challenges when using this approach in practice: Fitting of the model parameters to the measured data can be mathematically complex and prone to instabilities due to the nonlinear nature of such models. Even worse, realistic simulations require a larger number of adjustable parameters. Consequently, even simplified models are often affected by problems of nonidentifiability. Such nonidentifiability can be either of a structural or practical nature (Wieland, Hauber, Rosenblatt, Tönsing, & Timmer, 2021): In some cases (denoted as structural), more than one set of parameters will lead to exactly the same results even without the introduction of measurement uncertainties. But in most cases (called practical), even small measurement errors or noise can lead to large variations of the determined parameters. These large, random fluctuations in the model parameters over time are a major obstacle for applications such as anomaly detection or tracking degradation dynamics.

Integrating available information in the form of prior distributions on the model parameters within a Bayesian framework is a suitable way to overcome these issues. Often, variations of the underlying parameters are already known from production tolerances or from a fleet of devices or measurements. A simple approach is then to replace, for example, the maximum-likelihood estimator (MLE) with the maximum-a-posterior (MAP) one. Additionally, it is often equally important not only to obtain point estimators, but also to quantify their uncertainty. Under certain operating conditions, the parameters may not be precisely identifiable, resulting in increased uncertainty. Accounting for this uncertainty is important because it can provide confidence in detecting anomalies and evaluating degradation over time.

A full Bayesian analysis, using, for example, a Markov Chain Monte Carlo (MCMC) approach, is computationally demanding and typically needs to be repeated whenever new measurements become available. Even though this might be lifted using schemes that reuse results if there is only a small shift in the data and therefore the posterior distribution. Such an approach may also be infeasible for other reasons: If the underlying model is too complex, the likelihood, a key component of any Bayesian analysis, may be impossible to evaluate or computationally intractable. Instead, an approach using simulation models capable of generating large amounts of synthetic data representing both healthy and faulty or degrading cases can be used to then serve as a basis for a purely data-driven classification or regression approach (L. Stanek & Fröhlich, 2000; M. Stanek, Morari, & Fröhlich, 2001). It is therefore of interest to understand whether some newly developed methods in the area of simulation-based inference (SBI), especially combined with neural posterior estimation (NPE), are a possibility for this approach to be used in practical applications.

These simulation-based inference approaches can also be compared with physics-informed neural networks. Whereas the latter can also incorporate a physical model and determine model parameters, even including their uncertainties, they still require that the physical model be specified in a form that can be integrated into the loss function. In contrast, simulation-based approaches only require access to a simulator capable of generating samples, without relying on a specific form of the underlying model.

The paper is organized in the following way: In Sec. 2 we discuss the idea behind SBI, focusing especially on the method behind *Bayesflow*, a library that was used for the study. As an application example, the mechanical actuator of a medium voltage breaker was studied. Details of the simulation model are discussed in Sec. 3. Specifics of the implementation in *Bayesflow* are then given in Sec. 4. In a first step, simulated data are used to validate the model performance in Sec. 6, which is critical for establishing trust in the model. Which is then applied to real data from an endurance test in Sec. 7. As an outlook, Sec. 8 discusses the embedding of such an approach within a prognostics framework, before summarizing the key findings in Sec. 9.

2. OVERVIEW OF SIMULATION-BASED STATISTICAL INFERENCE METHODS

A common problem when using simulation models is the fact that the likelihood of the measurements x given the parameters θ, $p(x|\theta)$, might be computationally intractable, even though it formally exists. This can be either because its calculation is computationally too demanding for further statistical analysis or because it is not known in a computable form at all.

In contrast, one often has access to a simulator capable of producing different simulated measurements given the model parameters. Such problems are known as simulation-based inference or likelihood-free methods. These have been developed over several years; for a review, see (Cranmer, Brehmer, & Louppe, 2020). One well-known method is the so-called "Approximate Bayesian Computation" (Sisson, Fan, & Beaumont, 2018), which has already been explored for PHM applications in (Hencken, Borrelli, Ceccarelli, & Krivda, 2022).

In this approach, a large number of model parameters θ_i are sampled from the prior distribution $p(\theta)$ and the simulator is used to generate simulated measurement data x_i according to $p(x|\theta_i)$. The measurement data are then compressed into a low dimensional summary statistic $s_i = s(x_i)$. Only simulations, where s_i is close enough to the observed one s_M are then kept. The corresponding samples θ_i can then be shown to be from the posterior distribution $p(\theta|x)$.

This is a computationally demanding approach, as most simulations are not used and are instead rejected by the method. Although several improvements have been developed to re-

duce the number of computations required, the method remains computationally expensive. In addition, for each new measurement x_M, the calculation needs to be repeated. This makes this approach only suitable for applications that are slow enough, and the high computational cost is warranted.

In recent years, a new approach to SBI has gained increasing interest. Due to newly developed methods in machine learning, it is possible to use special neural networks that learn to produce samples θ_i from the posterior distribution $p(\theta|x)$ efficiently, conditional on x_M. These are often called neural posterior estimator (NPE). Originally, they were based on conditional invertible neural networks (cINNs), which make use of so-called normalizing flows to learn a generative model of posterior samples that are trained conditionally on the measurement data.

The main idea is to train a flow $f(z)$ that transforms a simple statistical base distribution $p(z)$ to one that approximates the posterior distribution. The network is trained to be conditional on the measurement data. A big advantage of this approach is that it can be trained on simulation data only, giving it the name simulation-based inference. No real data are required during this stage. In the inference phase, the real measurement is fed as a condition, and the networks generate samples from the posterior distribution of the parameters, to be further analyzed.

A major advantage of this approach, also with respect to PHM application, is its amortization feature. This means that the model is trained offline using simulated data only once. Afterward, the evaluation does not require any additional retraining, for example, to adapt the method to some specific asset. This also means that posterior samples can be generated very quickly.

In this work, we have made use of the available *"Bayesflow"* Python package, as it readily implements a flexible framework and allows making use of a number of available submodels. For details of this package, we refer to (Radev, Mertens, Voss, Ardizzone, & Köthe, 2020; Radev, Schmitt, Schumacher, et al., 2023; Radev, Schmitt, Pratz, et al., 2023) and (BayesFlow authors (lead maintainer: Stefan T. Radev), 2023-2026).

3. MECHANICAL MODEL OF THE BREAKER ACTUATOR

As an example to demonstrate the performance of this approach for PHM applications, we use a mechanical actuator of a medium voltage breaker. Medium voltage breakers are important devices in electrical distribution networks to interrupt the current in case of a fault, but are also used in applications to connect and disconnect currents on a regular basis, leading to a large number of operations. They are typically based on a spring-operated mechanical mechanism. It is known that actuator failure is one of the main failure modes of such breakers (Janssen, Makareinis, & Sölver, 2014), making the tracking of these devices during operations of interest.

A schematic view of such an actuator is shown in Fig. 1(a). To close the contacts, an additional spring applies a force pushing the mechanism upward. This is translated via the main shaft and a lever to a pushrod, which then closes the contacts. To reduce the electrical resistance of the closed contacts, they are pressed together with a sufficiently large force. This force is provided by an additional spring inside the pushrod. When opening the breaker, the external additional force is removed, and the system opens through the compressed spring in the pushrod and some additional spring attached to the main shaft.

Different sensors have been proposed to monitor these actuators, among them travel sensors, vibration sensors, and sensors providing timing information (Hoffmann et al., 2020). Here, we make use of the measurements of a travel sensor, which records the travel curve, that is, the position of the moving mechanism during the opening or closing motion, as a basis for the analysis.

Detailed models have been studied for such devices, mostly making use of a multi-body simulation approach; see, e.g., (Baniasadi, Darijani, & Matin, 2025). Its main purpose is to ensure the robust operation of the mechanism, but they can also be used to study changes due to parameter evolution. One difficulty is the nonlinearity introduced by the rotary motion. In addition, when the contacts close or open, the system dynamics change, changing from a system with two degrees of freedom to one with only a single one, and vice versa. The real mechanism is even more complex, introducing a larger number of both geometrical and physical parameters, like spring constants and friction forces, that need to be set. This, as mentioned above, makes the system nonidentifiable.

A simpler, but still realistic model, is a two spring-mass system, which is used here. The nonlinearity of the system is assumed to introduce minor deviations due to the small angle by which the main shaft is moving. The different moving parts are assumed to be described by two degrees of freedom due to the linkage of all components. This simulation model is shown in Fig. 1(b), the ODE describing its dynamics is given by

$$\begin{aligned} M_a\ddot{X}_a &= -K_1(X_a - X_1) + K_2(X_b - X_a - X_2) \\ &\quad -\Gamma_1\dot{X}_a + \Gamma_2(\dot{X}_b - \dot{X}_a) + F \\ M_b\ddot{X}_b &= -K_2(X_b - X_a - X_2) - \Gamma_2(\dot{X}_b - \dot{X}_a) \end{aligned} \quad (1)$$

if the two masses are moving. In case the contacts are closed, the second mass is assumed to be fixed at $X_b = L_0$ and cor-

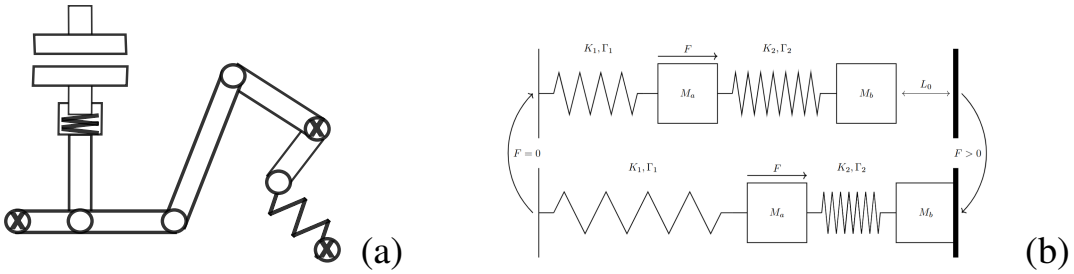

Figure 1. Schematic drawing of the actuator mechanism of the MV breaker during the opening phase (with $F = 0$) (a). The system consists of the contact system, which is linked via a pushrod containing a spring to the lever. The lever is moved up or down, leading to a rotation of the main shaft. Its representation as a linear ODE of a two spring-mass is depicted in (b). The system has two dynamic regimes, depending on whether the contacts are open or closed.

respondingly $\dot{X}_b = 0$, and the remaining dynamics is

$$\begin{aligned} M_a\ddot{X}_a &= -K_1(X_a - X_1) + K_2(L_0 - X_a - X_2) \\ &\quad -\Gamma_1\dot{X}_a - \Gamma_2\dot{X}_a + F \\ X_b &= L_0 \end{aligned} \tag{2}$$

In addition, the collision of the contacts is assumed to be fully inelastic. In the application, we can only observe the position of a, x_a, whereas the position of the second mass b, x_b, and its velocity are inaccessible.

The transition between the two dynamical modes of open and closed contacts is found in the following way: During closing, the change happens upon the moving contact reaching the fixed one. During opening, on the other hand, the total force on the moving contact is calculated, and it is assumed to open when this force becomes zero or changes sign. Note that forces due to the current flowing through the system are ignored. This is because the application is mostly interesting for operations under nominal current, where the current and, therefore, Lorentz forces are small. Also note that the real system consists of three contact systems, which are assumed here to be moving in sync.

Even this very simple system faces some nonidentifiability issues. A first, trivial one is the fact that the equations are invariant under the rescaling of all parameters by a constant factor. This can be removed by dividing the equations by the first mass M_a. Another issue is that the origin of the coordinate system can be chosen arbitrarily for each coordinate. Defining new coordinates $x_a = X_a - X_1$, $x_b = X_b - X_1 - X_2$ removes this arbitrariness and simplifies the equations by identifying the origin of the coordinate system to coincide with the neutral position of the two springs. This also leads to the gap distance in the new coordinate system defined as $L = L_0 - X_1 - X_2$. The equations of motion are then given by

$$\begin{aligned} \ddot{x}_a &= -k_1 x_a + k_2(x_b - x_a) \\ &\quad -\gamma_1\dot{x}_a + \gamma_2(\dot{x}_b - \dot{x}_a) + f \\ \ddot{x}_b &= -\mu k_2(x_b - x_a) - \mu\gamma_2(\dot{x}_b - \dot{x}_a) \end{aligned} \tag{3}$$

and

$$\begin{aligned} \ddot{x}_a &= -k_1 x_a + k_2(L - x_a) \\ &\quad -\gamma_1\dot{x}_a - \gamma_2\dot{x}_a + f \\ x_b &= L \end{aligned} \tag{4}$$

respectively. Here, f has a positive value during closing and is equal to zero during opening.

This leaves us with a system without obvious nonidentifiability and a total of seven parameters:

$$\begin{aligned} &\mu = M_b/M_a, \quad k_1 = K_1/M_a, \quad k_2 = K_2/M_a, \\ &\gamma_1 = \Gamma_1/M_a, \quad \gamma_2 = \Gamma_2/M_a, \quad f = F/M_a, \quad L \end{aligned} \tag{5}$$

A simulation is initialized from either of the equilibrium or the stationary initial states. For the open state, one finds

$$x_{a,open} = 0, \qquad x_{b,open} = 0 \tag{6}$$

and for the closed one

$$x_{a,open} = \frac{f + k_2 L}{k_1 + k_2}, \qquad x_{b,open} = L \tag{7}$$

where we only consider travel curves that start with the contacts closed.

Using these initial conditions, the ODE system is solved using a standard solver, and the condition for the change of the dynamic mode is checked. In the case of closing, we change it the moment x_b reaches L, whereas for the opening motion, the sum of the spring force and friction force is checked for becoming zero or changing sign. In a last step the exact position of the first mass x_a is modified by adding some normally distributed noise to it. The size of this noise was adapted to correspond to the one found in the real measurement system.

4. IMPLEMENTATION IN BAYESFLOW

BayesFlow is a Python library that provides a convenient framework for simulation-based, likelihood-free, amortized Bayesian inference using neural networks, enabling the imple-

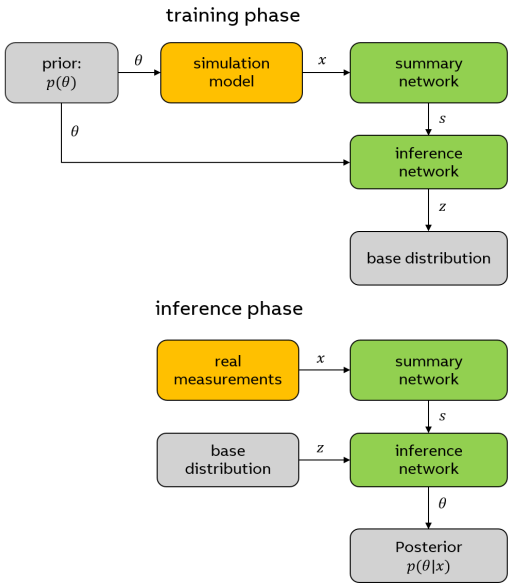

Figure 2. Schematics of the SBI approach as used in *Bayesflow*: During the training phase, model parameters θ are sampled from the prior distribution $p(\theta)$ and measurement data x are simulated. These data are then fed into a summary network to compress it into a fixed-size, theoretically optimal summary statistic s, before being used as conditional input to the inference network. The inference network then learns the mapping of the base distribution $p(z)$ to the parameter space. During the inference phase, the real measurement data x are again fed into the summary network. The resulting summary statistics are then used to condition the inference network. Samples z from the base distribution $p(z)$ are transformed into posterior samples of the model parameters. Adapted from (Radev et al., 2020).

mentation of neural posterior estimators (NPEs).

In this section, we examine in more detail the neural network components underlying the amortized Bayesian inference in *BayesFlow* as used in this application, namely the summary and inference networks depicted in Fig. 2. We are building on Sec. 6, where we introduced the simulator-based framework, of sampling model parameters from the joint prior and using them to generate data, thereby implicitly defining the likelihood. Sec. 3 describes the specific physical model used in this work, which provides the simulated (travel curves/ circuit breaker) data for training.

Given the pairs of parameter samples θ from prior $p(\theta)$ and the corresponding simulated data x from the implicit likelihood $p(x|\theta)$, the potentially high-dimensional observations are first processed by a summary network. Its role is to compress the data into a fixed-size, maximally informative representation, which serves as input to the inference network. This avoids the need for a manually designed, potentially suboptimal, summary statistics and allows for flexible adaptation to the structure of the data. Generally, the summary network is a *simple* feed-forward neural network. In our case, the observations are time series, and we therefore decided on a suitable architecture, a *TimeSeriesNetwork*, based on the LSTNet architecture. Alternative choices include set-based architectures such as *DeepSet*, containing pairs of both time and distance, which would be preferred in case of unordered or missing data, or convolutional neural networks for measurements based on image data.

The inference network constitutes the core of the posterior approximation. It is implemented as an invertible neural network and employs a flow-based approach capable of transforming a simple base distribution into a more complex, potentially multi-modal distribution. The idea is based on transportation maps, which are chosen to be implemented as an invertible neural network, which are represented by normalizing flows (Papamakarios, Nalisnick, Rezende, Mohamed, & Lakshminarayanan, 2021; Kobyzev, Prince, & Brubaker, 2020), or subsequent flow matching (Xie & Cheng, 2025; Wildberger et al., 2023; Lipman, Chen, Ben-Hamu, Nickel, & Le, 2023), which were introduced as a follow-up on some of the shortcomings of normalizing flow. Importantly, the transformation needs to be invertible and differentiable in both directions, i.e., a *diffeomorphism*, ensuring a valid probabilistic mapping. During training, the network learns to map parameters θ, conditioned on the summary statistics of the data s, to a simple latent space z with a known base distribution $p(z)$. At inference time, this mapping is inverted: samples z drawn from the base distribution $p(z)$ are transformed into those from the posterior distribution conditioned on the observed data $p(\theta|s)$. Flow matching methods are used in this work, which extend classical normalizing flows and can offer improved training efficiency and stability while retaining the flexibility of invertible transformations.

The summary and inference networks are jointly trained on a large number of simulated cases. For this, we use a standard stochastic gradient (SGD) approach.

5. PRIOR DISTRIBUTIONS

A key ingredient to the inference is the definition of a suitable prior distribution for the seven dimensional parameter space. Because the real geometry is transformed into the simplified one, the parameters lose some of their direct interpretation. Instead of converting the physical parameters like spring constants and masses to this new model, we have made use of earlier studies, where parameter fitting of a number of travel curves was already done, and some good mean values were found. This was then used for determining both the location and the spread parameter of a log-normal distribution. This distribution was chosen because all parameters have to be positive.

Confirming that these priors are suitable was done by performing a ”prior-predictive check”. For this, a number of travel curves are generated using parameters sampled from

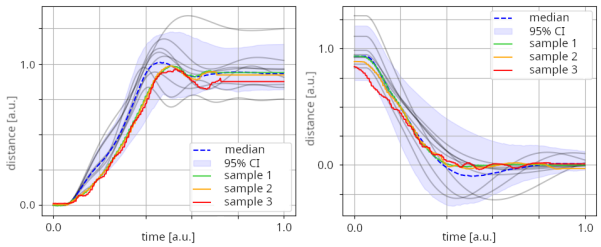

Figure 3. Prior predictive check for the closing and opening travel curves. Shown are the results of simulated travel curves using the model and prior distributions of its parameters, both individual travel curves and the 95% credibility interval. These are compared to three real travel curves. Overall, the real travel curves are well represented within the possible range covered by the model and prior distributions.

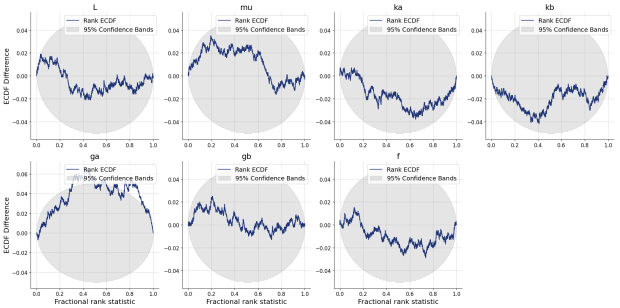

Figure 4. Simulation-based calibration (SBC) plots for the closing case, showing that the model is well-trained. For details on the interpretation of these graphs, refer to the main text. Apart from the parameter γ_b, all ECDF differences are within the 95% confidence interval.

the prior and the simulator. These generated travel curves are then compared to some of the real ones measured. If there is a good overlap and agreement between those two, it gives some confidence that the model, including the priors, is suitable for the analysis. The results of this prior-predictive check are shown in Fig. 3, which confirms the suitability.

6. SIMULATION RESULTS

In a next step, simulated data are used as measurement (or inference) input to evaluate the performance of both the approach and the trained model. Such an analysis is important to confirm not only that the training of the model has converged, as indicated by the evolution of the loss function, but also that the resulting model agrees with the expected posterior distribution. These steps can be seen as following the "Bayesian workflow" (Gelman et al., 2020), which is a structured approach to get confidence in the statistical analysis.

A first step in the analysis uses the "simulation-based calibration" (SBC) (Talts, Betancourt, Simpson, Vehtari, & Gelman, 2020). This is an approach commonly used to test the internal consistency of the model. It is based on the fact that, if $\hat{x}$ are synthetic data generated from parameters $\hat{\theta}$ drawn from the prior, then the posterior distribution $p(\theta|\hat{x})$ marginalized over $\hat{x}$ and $\hat{\theta}$ should coincide with the prior distribution $p(\theta)$. By plotting the ECDFs of both sample types for the parameters, together with their confidence intervals, this consistency can be assessed. It is a strong test of the performance of the posterior distribution. Results from this test are shown in Fig 4 for the closing case; similar results are also found in the opening one. The difference between the two curves is almost always within the 95% confidence interval. This shows that the model is well calibrated.

A second analysis is to check the distribution of the posterior samples with respect to the z-score, that is, the normalized distribution around the true value, against the "contraction", defined as $1-\sigma^2_{post}/\sigma^2_{prior}$, which is an indicator of how well the measurement can reduce the uncertainty of the parameter

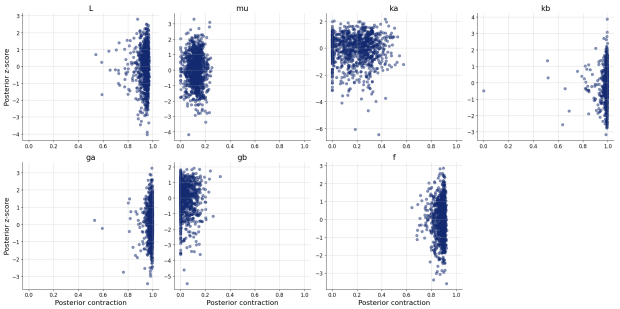

Figure 5. Scattering of the z-score and the contraction for the different model parameters are shown for the closing case. The z-score distribution shows that the model is distributed without bias and large outliers. The contraction shows that only part of the parameters can be informed in this case.

compared to the prior one. Results are given in Fig. 5 again only for the closing case as an example. For the z-score, most posterior samples fall into the range of three standard deviations and are well centered around zero, indicating that there is no bias in the method. For the contraction, on the other hand, the parameters fall into two groups: some are close to one, indicating accurate estimation, while others are much lower, meaning they only weakly improve with respect to the prior.

In the case of the simulated data, we also have access to the ground truth, so we can plot the relation between the posterior samples and it. Apart from confirming that the model is working as expected, it can also indicate whether certain parameter regions are better suited than others for parameter estimation. Results are shown in Fig. 6 now for both the opening and closing cases. The same behaviour as before for the contraction is seen: certain parameters are determined with a high accuracy, whereas some seem to be sampled from the prior distribution only, without any improvement due to the measurement. One can also see that the opening and closing analysis are sensitive to different parameters, partially complementing each other. This shows that the most likely explanation for the two

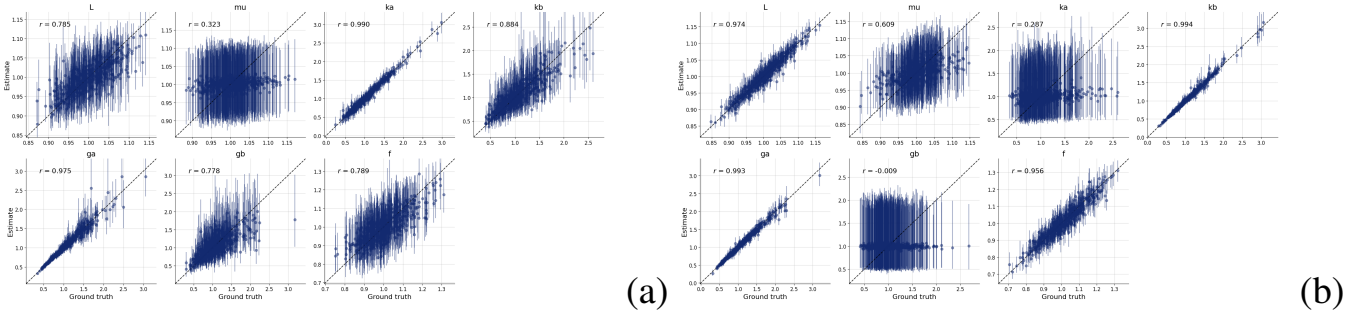

Figure 6. Comparison of the posterior samples with the ground truth value for the opening (a) and closing (b) case. The diagonal corresponds to the case where the ground truth and posterior agree. As seen before, only some of the model parameters can be inferred. Others remain distributed according to the prior distribution. The sensitivities differ between the opening and closing measurements.

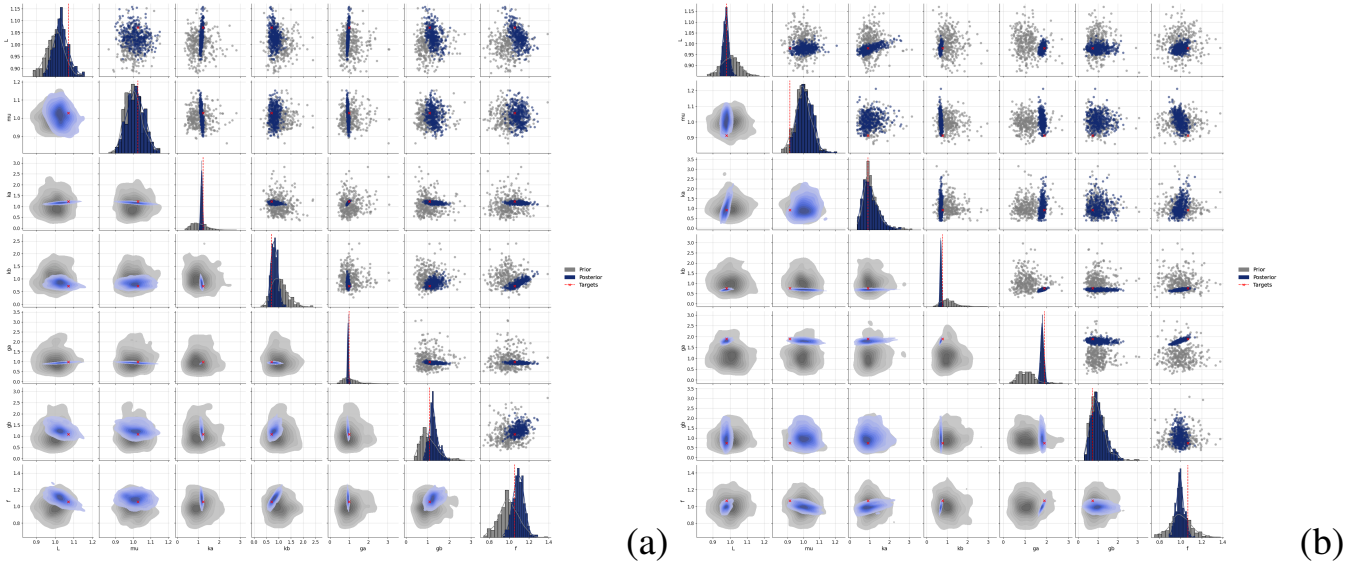

Figure 7. Pairs Plot showing prior, posterior distribution, and ground truth of one simulated case. Shown for the opening (a) and closing (b) case. No strong correlation between the parameters is found for almost all pairs. The posterior distributions are also found to include the ground truth values in all cases.

groups is that some parameters are not identifiable from the specific travel curve. The informed priors used, of course, also restrict them, so that weak identification would still be possible but is then overruled by the prior.

A regular analysis to be performed is the pairs plot for some specific cases of simulated data. Here, one shows the pairwise probability distribution, which can indicate strong correlations between parameters. In the simulation case, three properties can be shown: the prior and posterior distributions, as well as the ground truth value. Results for one representative case are given in Fig. 7. Most pairs of posterior distributions seem to be almost uncorrelated, with a notable exception between the applied force f and the spring constant k_b. One can also see that the ground truth in general falls within the high-probability region of the posterior distribution. Finally, a posterior predictive check can be made by taking some simulated travel curves and comparing it with the distribution of the predicted ones from the posterior distribution. Results for four cases each are shown in Fig. 8, with a good agreement, that is, the simulated curves falling within the uncertainty range of the predicted curves. In the closing case, one also sees that for some cases the curve yields substantial uncertainty towards the end in time, indicating that in these cases the parameters are not determined very accurately.

7. RESULTS FROM ENDURANCE TEST DATA

We demonstrate the successful application to real data from a mechanical endurance test. Data from both opening and closing travel curves were collected over a large number of operations until failure. Posterior predictive checks for some cases at different stages of the testing are shown in Fig. 9 for

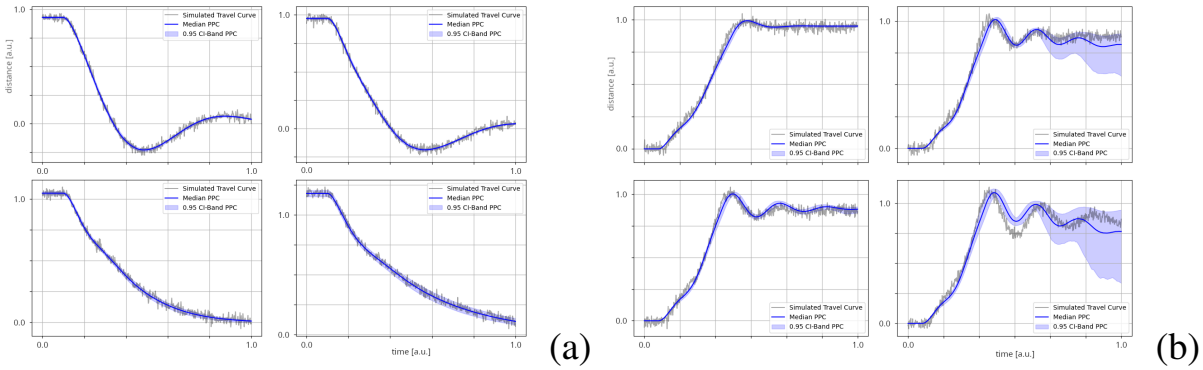

Figure 8. Results of a posterior predictive check are shown: four examples of simulated and predicted travel curve for the opening (a) and closing (b) case. The agreement is quite good, with the last stage of the closing travel curve showing some larger uncertainty.

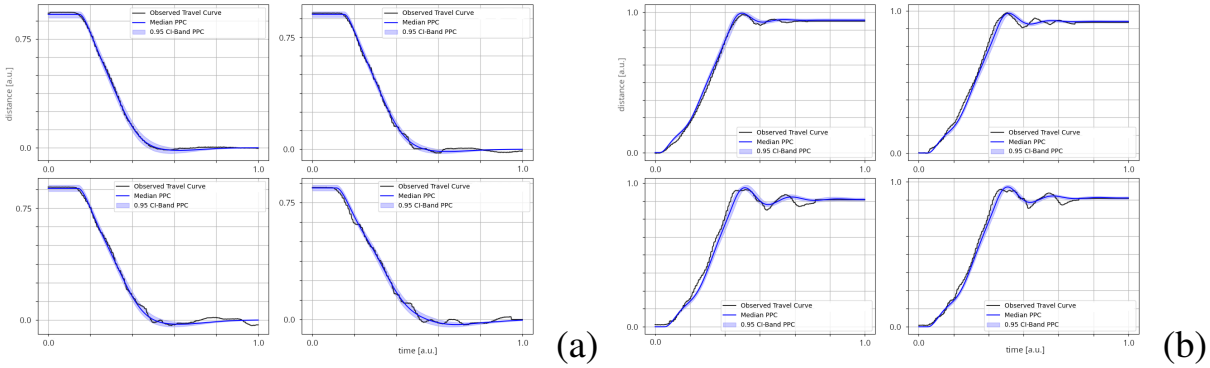

Figure 9. Posterior predictive check for both the opening (a) and closing (b) travel curves using four different real measurements at different stages of the degradation of the breaker mechanism. The travel curves are generally well predicted by the model, with some deviations seen in the final phase of the closing curve.

both opening and closing curves. In both cases, the main part of the motion is captured accurately, whereas the two-spring model seems to be unable to capture some details of the system's damping at the end of the motion. This is clearly a limitation of the simplified model. Whereas the main motion is dominated by the movement of the contact system, the final part depends more strongly on details of the mechanical drive, which is a part that was not modeled. It is therefore interesting to see whether this model inadequacy leads to issues when tracking the evolution of individual model parameters.

The evolution of all seven parameters for both the opening and closing case is given in Fig. 10.

Consistent with the simulated data, the travel curves do not allow certain model parameters to be determined. Again, the affected parameters differ between the opening and closing scenarios. Of the model parameters that can be determined, some not only show a linear trend, but also a clearly visible change point, where the slope changes strongly towards the end of the test, corresponding to the end of life of the device. This indicates that the method is able to detect not only normal degradation but also a potential transition to a faulty state before the final failure of the mechanism. The uncertainty bounds are of interest to confirm that there is a real change in slope rather than just a normal fluctuation in them.

8. OUTLOOK

The parameters determined, including uncertainties are now the basis for further analysis. A simple application is for diagnostics. Using a static threshold, set for example by simulation studies on a more complex multi-body simulation, the probability of exceeding this threshold can be determined. This can be done either using a single distribution for individual parameters, or by looking at the full (correlated) posterior distribution.

A second use case is for prognostics. The parameters, including uncertainties, are used to inform a degradation model, which is then again used to calculate the probability of crossing a threshold, given the RUL distribution. As we have seen, the degradation of the circuit breaker also shows some change point behaviour. Different change point detection algorithms are available in the literature (Li, Lin, Lau, & Zeng, 2019). One approach suitable especially for prognostics application

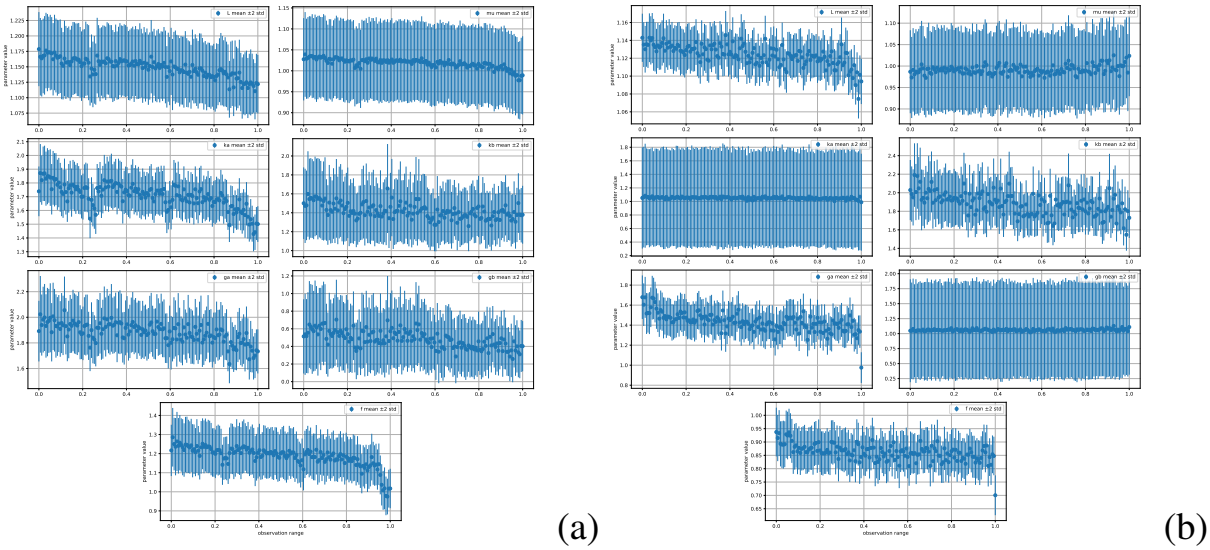

Figure 10. Development of the different model parameters over the full endurance test for the opening (a) and closing (b) case. Especially for some variables, both a linear trend and in addition a change point close to the final failure of the actuator can be seen.

is the "Bayesian Online Changepoint Detection" algorithm (Adams & MacKay, 2007). For an application of this algorithm, see (Mukin, Hencken, & Cilibrasi, 2026).

9. SUMMARY AND CONCLUSIONS

In this work, we discussed the problems that arise when using simulation models to track the evolution and degradation of assets and using the model parameters as features for diagnostics or prognostics. Often, such an approach is not easily doable due to the nonidentifiability of such systems, which needs to be overcome. The Bayesian approach solves this by introducing prior distributions, constraining the potential parameter values to reasonable ones.

Nevertheless, a full Bayesian approach is computationally demanding, and calculating the likelihood needed, for example, for an MCMC approach, might not be tractable. Using a likelihood-free or simulation-based approach overcomes these problems. Recent NPE or SBI approaches are available to implement such an approach. Training the model requires only simulated data, and the amortized nature of the approach means that the system needs to be trained only once, upfront and offline.

Using as an example the mechanical actuator of an electrical breaker, it was shown that this approach, using only data from a simplified spring-mass model, is possible. The performance of the model was verified using a number of statistical checks. Finally, using real data from an endurance test showed that the approach can track the degradation of some parameters, whereas those that cannot be easily determined from the specific measurement are not causing problems.

The use of simulation-based inference approaches is an interesting and promising approach to PHM. On the one hand, one often has rather detailed simulation models for the technical assets one is interested in. On the other hand, using the knowledge incorporated in these models is important for identifying relevant features via model parameters and for becoming robust to other parameters. The availability of NPE methods is still rapidly evolving, but the current state is already suitable for real applications.

REFERENCES

Adams, R. P., & MacKay, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

Baniasadi, A. A., Darijani, H., & Matin, M. (2025). Analytical multi-body approach for component selection and timing analysis in spring-actuated vacuum circuit breakers. *Scientific Reports*, *15*(1), 42478. doi: 10.1038/s41598-025-26609-6

BayesFlow authors (lead maintainer: Stefan T. Radev). (2023-2026). *BayesFlow web page.* (`https://www.bayesflow.org`)

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055-30062. doi: 10.1073/pnas.1912789117

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ... Modrák, M. (2020). Bayesian workflow. *arXiv preprint arXiv:2011.01808*.

Hencken, K., Borrelli, E.-M., Ceccarelli, D., & Krivda, A. (2022). Approximate Bayesian computation as a new

tool for partial discharge analysis of partial discharge data. In *PHM Society European conference* (Vol. 7, pp. 182–192).

Hoffmann, M. W., Wildermuth, S., Gitzel, R., Boyaci, A., Gebhardt, J., Kaul, H., ... Tornede, T. (2020). Integration of novel sensors and machine learning for predictive maintenance in medium voltage switchgear to enable the energy and mobility revolutions. *Sensors*, *20*(7), 2099. doi: 10.3390/s20072099

Janssen, A., Makareinis, D., & Sölver, C.-E. (2014). International surveys on circuit-breaker reliability data for substation and system studies. *IEEE Transactions on Power Delivery*, *29*(2), 808-814. doi: 10.1109/TPWRD.2013.2274750

Kobyzev, I., Prince, S. J., & Brubaker, M. A. (2020). Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, *43*(11), 3964–3979.

Li, Y., Lin, G., Lau, T., & Zeng, R. (2019). A review of changepoint detection models. *arXiv preprint arXiv:1908.07136*.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2023, 2). Flow matching for generative modeling. *11th International Conference on Learning Representations, ICLR 2023*.

Mukin, R., Hencken, K., & Cilibrasi, M. (2026). Application of the Bayesian online change-point detection algorithm to the degradation of an electrical device. *Proceedings of the PHM Europe Conference*.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, *22*, 1-64.

Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Köthe, U. (2020). BayesFlow: Learning complex stochastic models with invertible neural networks. *arXiv preprint arXiv:2003.06281*.

Radev, S. T., Schmitt, M., Pratz, V., Picchini, U., Köthe, U., & Bürkner, P.-C. (2023). JANA: Jointly amortized neural approximation of complex Bayesian models. In *Proceedings of the thirty-ninth conference on uncertainty in artificial intelligence* (Vol. 216, p. 1695-1706).

Radev, S. T., Schmitt, M., Schumacher, L., Elsemüller, L., Pratz, V., Schälte, Y., ... Bürkner, P.-C. (2023). BayesFlow: Amortized Bayesian workflows with neural networks. *The Journal of Open Source Software*, *8*(89), 5702.

Sisson, S. A., Fan, Y., & Beaumont, M. (2018). *Handbook of approximate Bayesian computation*. CRC press.

Stanek, L., & Fröhlich, K. (2000). Model-aided diagnosis-a new method for online condition assessment of high voltage circuit breakers. *IEEE Transactions on Power Delivery*, *15*(2), 585–591.

Stanek, M., Morari, M., & Fröhlich, K. (2001). Model-aided diagnosis: An inexpensive combination of model-based and case-based condition assessment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *31*(2), 137–145.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2020). Validating Bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*.

Wieland, F.-G., Hauber, A. L., Rosenblatt, M., Tönsing, C., & Timmer, J. (2021, March). On structural and practical identifiability. *Current Opinion in Systems Biology*, *25*, 60–69. doi: 10.1016/j.coisb.2021.03.005

Wildberger, J., Dax, M., Buchholz, S., Green, S., Macke, J. H., & Schölkopf, B. (2023). Flow matching for scalable simulation-based inference. *Advances in Neural Information Processing Systems*, *36*, 16837–16864.

Xie, Y., & Cheng, X. (2025). Flow-based generative models as iterative algorithms in probability space. *arXiv preprint arXiv:2502.13394*.

BIOGRAPHIES

Cédric Schenkel obtained a Bachelor of Science in Pharmaceutical Sciences from the University of Bern in 2021 and a Master of Science in Drug Sciences from the University of Basel in 2023. He subsequently pursued additional studies in Biomedical Engineering at the University of Basel, focusing on the mathematical foundations and implementation of machine learning and deep learning. In 2025, he joined ABB for an internship in Simulation-Based Statistical Inference. Since January 2026, he has been pursuing a PhD in Computational Chemistry at the University of Southampton.

Kai Hencken is a corporate research fellow at the ABB corporate research center. He obtained a PhD in Theoretical Physics from the University of Basel in 1994. He was a post-doc at the University of Washington from 1995 to 1997 and at the University of Basel from 1997 to 2005, where he received his Habilitation in 2000 and is a lecturer since. In 2005, he joined the theoretical Physics group at ABB corporate research. His research interests include the combination of physical modelling with statistical methods to solve problems related to industrial devices, as well as developing diagnostics and prognostics approaches.