HOKKAIDO UNIVERSITY

| | |
|---|---|
| Title | Bayesian Model Calibration Using Temporal Process Features and Streamlined Active Learning |
| Author(s) | Verkens, B.; Faes, M.; Van Rymenant, P. et al. |
| Relation | Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026), June 28 - July 1, 2026, Sapporo, Japan |
| Citation | Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026) |
| Issue Date | 2026-07-01 |
| Doc URL | https://hdl.handle.net/2115/100447 |
| Type | conference paper |
| File Information | C000087_1.pdf |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

ISRERM 2026

*10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026)*
28 June–1 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

# Bayesian Model Calibration Using Temporal Process Features and Streamlined Active Learning

B. Verkens[1,2*], M. Faes[3,4], P. Van Rymenant[1] and D. Moens[1,2]

[1)] *KU Leuven, Department of Mechanical Engineering*
*Jan Pieter de Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium*
*email: {bouwe.verkens, patrick.vanrymenant, david.moens}@kuleuven.be*
[2)] *FlandersMake@KU Leuven, Belgium*
[3)] *TU Dortmund, Chair for Reliability Engineering*
*Leonhard-Euler-Strasse 5, 44227 Dortmund, Germany*
*email: matthias.faes@tu-dortmund.de*
[4)] *Tongji University, International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Shanghai 200092, China*

Accurate multi-physics finite element modelling of resistance spot welding (RSW) remains challenging due to the uncertainties in contact and material parameters that shape the coupled thermal–electrical–mechanical process. This study introduces a Bayesian calibration framework, in contrast to the deterministic parameter optimisation approaches commonly adopted in existing RSW literature, which neglect the inherent variability of the welding process. The proposed framework performs calibration using multivariate model outputs, combining nugget geometry with geometrical and temporal attributes [1] extracted from process signals such as dynamic resistance and electrode force. These features provide physically interpretable information on process evolution while reducing redundancy and correlation relative to full time-series data, enabling more effective parameter identifiability than nugget-only calibration. The methodology is first assessed on an analytical model using artificial data to evaluate inference accuracy and computational efficiency. Bayesian inference is carried out using the Streamlined Bayesian Active Learning Cubature (SBALC) method [2], which infers posterior distributions of uncertain model parameters (such as contact thickness, contact resistance, and temperature-dependent material properties) while requiring only a limited number of forward model evaluations.

***Key words:*** Resistance spot welding, Finite element simulation, Uncertainty, Validation, Electro-thermomechnical model

## 1 Introduction

Numerical models of dynamical systems, ranging from heat transfer and coupled multi-physics processes to biological and ecological systems, often depend on parameters that are cannot be measured directly. Instead, these parameters must be estimated from experimental observations. Traditional calibration methods treat this a determinstic optimisation problem and tune the parameters until the model's output best matches the data in a least squares sense. While these methods are widely applied, they yield only point estimates, ignore epistemic uncertainty, and are sensitive to local minima. The aim is to approximate the system at hand as closely as possible, but a model will always be an imperfect representation of reality, and true parameter values may vary between experiments. Therefore, a calibration framework takes these facts into account should deliver probability distributions over the unknown parameters rather than single values.

Bayesian calibration offers a framework for this problem [3]. By combining prior knowledge with the likelihood of the observed data, Bayes' theorem yields a posterior distribution that quantifies the remaining uncertainty on the parameters given the available data. The main challenge to applying Bayesian calibration to high fidelity numerical models is its computational cost [4]. Classical sampling methods to estimate the posterior, such as Markov Chain Monte Carlo (MCMC) [5] or transitional MCMC [6] require thousands to millions of likelihood evaluations, each of which requiring a full forward model run. When a single simulation takes minutes to hours, this becomes infeasible. Surrogate-assisted methos adress this bottleneck. Recent research in this, the Streamlined Bayesian Active Learning Cubature (SBALC) method [2], fits a Gaussian process surrogate directly to the log-likelihood function and uses active learning to minimise the number of forward model evaluations required.

A separate challenge arises when the model output is a time series, rather than a small set of scalar metrics. Time series produced by dynamical models, such as temperature histories, foce signals, populations over time, and similar quantities, are typically highly autocorrelated. This means that successive datapoint almost carry the identical information on the underlying parameters. Treating such data as independent observations using a diagonal covariance matrix counts the same physical discrepancy many times, producing an artificially sharp likelihood and, consequently, overconfident posteriors. The solution, a full covariance matrix, is impractical for long signals due to its cubic inversion cost. Moreover, when surrogate-based methods such as SBALC are applied, the extreme dynamic range of the

resulting log-likelihood surface causes numerical difficulties in the Gaussian process regression (GPR), such as short kernel length scales and clustering of new training points that can trigger Cholesky factorisation failures.

This paper proposes a solution to this by compressing the time series outputs into small sets of physically motivated scalar features before calibration. This reduces the residual dimension from $m$ to $p \ll m$, which broadens the likelihood surface, avoids the autocorrelation-induced overconfidence of the posterior, and avoids the numerical fragility of surrogate-based calibration methods on high-dimensional outputs. The approach is demonstrated on an analytical benchmark of a copper wire heated by a time-varying electrical current, governed by an energy-balance ODE with convective and radiative heat losses. Because the ground truth is known, this setup allows controlled comparison against full-signal calibration.

Section 2 presents the Bayesian calibration framework, the SBALC method, and the proposed feature-based calibration target. Section 3 introduces the analytical benchmark and defines the extracted features. Section 4 compares four calibration scenarios that progressively test parameter identifiability and numerical robustness. Section 5 discusses the results and their implications for calibrating multi-physics models with time-series outputs.

## 2 Methodology

### *2.1 Bayesian model calibration*

Bayesian model calibration combines prior knowledge (informative or non-informative), measurement noise and model discrepancy together and solves the inverse problem in order to estimate posterior distributions over the unknown process parameters.

Let $\mathbf{x} \in \mathcal{D}_X \subseteq \mathbb{R}^d$ denote the vector of uncertain model parameters. Given observations $Y_{\text{obs}}$, Bayes' theorem updates the prior distribution $f(\mathbf{x})$ to the posterior:

$$f(\mathbf{x} \mid Y_{\text{obs}}) = \frac{L(Y_{\text{obs}} \mid \mathbf{x})\, f(\mathbf{x})}{c}, \tag{1}$$

where $L(Y_{\text{obs}} \mid \mathbf{x})$ is the likelihood function and $c$ is the model evidence (marginal likelihood):

$$c = \int_{\mathcal{D}_X} L(Y_{\text{obs}} \mid \mathbf{x})\, f(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{2}$$

Assuming Gaussian measurement error with zero mean and covariance $\mathbf{\Sigma}$, the log-likelihood for $N_{\text{obs}}$ independent observation replicates is:

$$\ln L(\mathbf{x}) = \mathscr{L}(\mathbf{x}) = -\frac{1}{2} \sum_{k=1}^{N_{\text{obs}}} \mathbf{r}_k(\mathbf{x})^\top\, \mathbf{\Sigma}^{-1}\, \mathbf{r}_k(\mathbf{x}) + \text{const}, \tag{3}$$

where $\mathbf{r}_k(\mathbf{x}) = Y_{\text{obs}}^{(k)} - \mathcal{M}(\mathbf{x})$ is the residual between the $k$-th observation and the model response $\mathcal{M}(\mathbf{x})$. The structure of $\Sigma$ depends on the choice of calibration target. When the target is a time series with strong autocorrelation, constructing and inverting $\Sigma$ becomes non-trivial. This issue is addressed in Section 3.2.

### *2.2 Streamlined Bayesian Active Learning Cubature (SBALC)*

The posterior distribution defined by Eq. (1) must be evaluated numerically. A broad family of methods exists for this purpose, including Markov Chain Monte Carlo [5], Transitional MCMC [6], Sequential Monte Carlo [7], structural reliability methods [8], and surrogate-assisted approaches based on Gaussian Process Regression (GPR) or Polynomial Chaos Expansions (PCE). All of these require a large number of likelihood evaluations, which becomes unfeasable when a single multi-physics RSW simulation takes hours of computing time.

The Streamlined Bayesian Active Learning Cubature (SBALC) method [2] addresses this bottleneck by fitting a Gaussian Process (GP) surrogate directly to the log-likelihood function and using active learning to minimise the number of forward model evaluations required. A GP prior is assigned over $\mathscr{L}(\mathbf{x})$, and the posterior mean function is used as a plug-in estimator for the model evidence:

$$\hat{c} = \mathbb{E}_{\mathbf{x}}[\exp(\hat{m}_{\mathscr{L}_n}(\mathbf{x}))]\,, \tag{4}$$

where $\hat{m}_{\mathscr{L}_n}$ is the GP posterior mean after $n$ evaluations and the expectation is approximated by Monte Carlo over $N_M = 2 \times 10^4$ prior samples, following the default setting recommended by Li et al. [2].

The initial design consists of $n_0 = 4$ Hammersley points mapped to parameter bounds via $\delta_0 = 10^{-2}$ quantile truncation. At each iteration the next evaluation point is selected by maximizing the learning function:

$$\mathcal{LF}(\mathbf{x}) = \sigma^2_{\mathscr{L}_n}(\mathbf{x}) \cdot \left( e^{m_{\mathscr{L}_n}(\mathbf{x}) + b\,\sigma_{L_n}(\mathbf{x})} - e^{m_{\mathscr{L}_n}(\mathbf{x}) - b\,\sigma_{L_n}(\mathbf{x})} \right), \tag{5}$$

with $b = 1$. Iterations stop when the convergence ratio $|c_{\text{upper}} - c_{\text{lower}}|/\hat{c} < b_{cr} = 0.1$ holds over two consecutive iterations. Posterior samples are generated via Sampling-Importance-Resampling (SIR) [9].

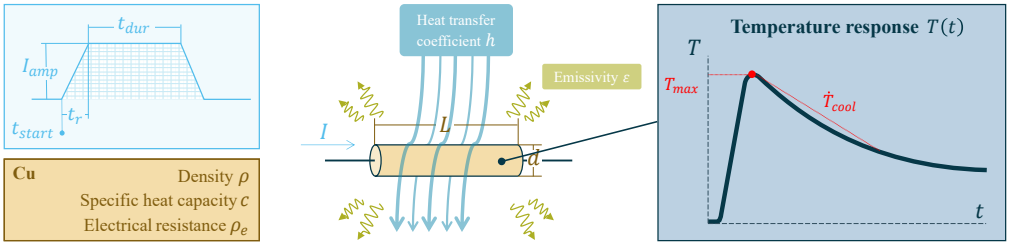


**Figure 1.** Schematic of the heated wire benchmark

### *2.3 Calibration target*

When the model output is a time series of $m$ datapoints, $\mathcal{M}(\mathbf{x}) \in \mathbb{R}^m$, the log-likelihood in Eq. (3) becomes a sum of $m$ squared residual terms. Temperature time series produced by dynamical models are highly autocorrelated. This means that successive samples carry nearly identical information about $\mathbf{x}$ as they are generated by an ordinary differential equation driven by a varying input. Treating them as independent under a diagonal $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_m$ amounts to counting the same physical discrepancy $m$ times, producing a posterior that is overconfident. Its width is controlled by the sample count rather than by the true or unique information on the signal. The remedy would be using a full covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ that captures the actual autocorrelation of the residuals. This is inefficient in practice as inverting it costs $\mathcal{O}(m^3)$ per likelihood evaluation. Using full time-series for calibration therefore proceeds with the diagonal approximation, accepting the pseudo-replication penalty.

Beyond the statistical issue, high-dimensional time-series outputs also create numerical difficulties for surrogate-based inference. The log-likelihood varies over a a very large range, with large negative values away from the optimum and almost zero close to it. This forces the Gaussian process to adopt a very short kernel length scale. At the same time, the learning function tends to select new evaluation points in high-likelihood regions, leading to clusters of training points that can cause failures in the Cholesky factorisation. This effect is visualised in Section 4 (Figure **??**).

The solution proposed in this study is to extract a small set of physically motivated scalar features from these time series data. This reduces the residual dimension from $m$ to $p \ll m$ which broadens the likelihood surface (Figure 2), mitigates autocorrelation-induced overconfidence and avoids the numerical fragility described above. For $d$ unknown parameters, at least $d$ features are needed, and their sensitivities to those parameters must be approximately linearly independent across the prior range. The following sections test this hypothesis on a controlled benchmark.

## 3 Analytical model

### *3.1 Physical model*

A copper wire of diameter $d = 0.5\,\text{mm}$ and length $L_0 = 1\,\text{m}$ is subjected to a time-varying current profile $I(t)$ with amplitude $I_{\text{amp}} = 1500\,\text{A}$ (Fig. 1). The wire temperature $T(t)$ evolves according to:

$$m_w c_p \frac{\mathrm{d}T}{\mathrm{d}t} = I(t)^2 R - hA\big(T - T_\infty\big) - \varepsilon\,\sigma_{\text{SB}} A\big(T^4 - T_\infty^4\big), \tag{6}$$

where $m_w = \rho\pi d^2 L_0/4$ is the wire mass, $A = \pi d L_0$ is the surface area, $R = \rho_e L_0/(\pi d^2/4)$ is the electrical resistance, and $\sigma_{\text{SB}} = 5.67 \times 10^{-8}\,\text{W/m}^2\text{K}^4$ is the Stefan–Boltzmann constant. Fixed material properties are: $\rho = 8960\,\text{kg/m}^3$, $c_p = 380\,\text{J/(kg K)}$, $\rho_e = 1.7 \times 10^{-8}\,\Omega\text{m}$, and $T_\infty = 25\,^\circ\text{C}$. Eq. (6) is integrated with with an explicit Runge–Kutta (4,5) method using event detection to locate $T_{\max}$ precisely. The unknown parameters are the convective heat transfer coefficient $h$ and the surface emissivity $\varepsilon$, with true values $h_{\text{true}} = 8000\,\text{W/m}^2\text{K}$ and $\varepsilon_{\text{true}} = 0.8$.

### *3.2 Feature extraction*

Two scalar features are extracted from the temperature time series:

- $T_{\max}$: the peak temperature during Joule heating (°C), reflecting the net energy balance and depending on both $h$ and $\varepsilon$.
- $\dot{T}_{\text{cool}}$: the post-peak cooling slope (°C/s), computed over a short interval following $T_{\max}$. As convective heat losses scale linearly with temperature excess while radiative losses scale as $T^4 - T_\infty^4$, the cooling rate responds differently to changes in $h$ and $\varepsilon$ compared to $T_{\max}$.

Two features constitute the minimum for a two-parameter calibration. The failure of a single feature to identify two parameters is a direct consequence of this identifiability condition (see Scenario 2 in Section **??**).

### *3.3 Synthetic observations and noise model*

Synthetic observations are generated by evaluating the forward model at the true parameter values and adding multivariate Gaussian noise. $N_{\text{obs}} = 15$ replicates are generated with noise standard deviations $\sigma_1 = 5\,°\text{C}$ (for $T_{\max}$) and $\sigma_2 = 50\,°\text{C/s}$ (for the cooling slope), with zero cross-correlation. The likelihood used in calibration adopts inflated values $\sigma_1 = 10$ and $\sigma_2 = 75$, providing conservatism about the true noise level and partial allowance for model inadequacy. After calibration, a Maximum Likelihood Estimation (MLE) of the effective noise level is obtained by evaluating the forward model at the posterior mean and computing the root mean squared residual against the observations.

## 4 Results

Before comparing four calibration scenarios, Figure‘2 illustrates the core problem motivating a feature-based approach. The left panel shows the log-likelihood $\mathscr{L}$ for the wire benchmark of Section 3 at fixed $m = 500$ for varying noise levels $\sigma$. As $\sigma$ decreases, the log-likelihood sharpens as expected. The right panel fixes $\sigma = 25$ and varies the amount of time-series datapoints $m$. The log-likelihood narrows in proportion to $m$, confirming that a diagonal covariance assumption treats each sample as an independent observation and artificially concentrates the likelihood around the true value.

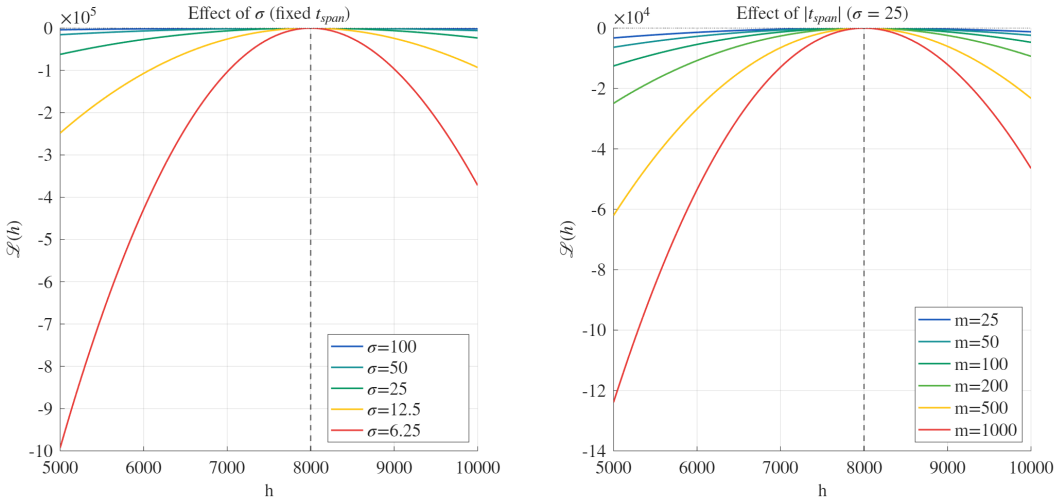


**Figure 2.** Log-likelihood $\mathscr{L}(h)$ of the wire benchmark ($h_{\text{true}} = 8000\,\text{W/m}^2\text{K}$).
*Left*: effect of $\sigma$ at fixed $m = 500$. *Right*: effect of $m$ at fixed $\sigma = 25$.

Four scenarios are designed to progressively test the proposed feature-based calibration. Scenario 1 calibrates a single parameter against a single feature. This forms the simplest possible case, intended to verify that SBALC recovers the correct posterior. Scenario 2 attempts to calibrate two parameters against only one feature, testing whether the expected identifiability failure occurs. Scenario 3 adds a second feature with different physical sensitivity, testing whether identifiability is restored. Finally, Scenario 4 replaces the scalar features with the full temperature time series to expose the numerical difficulties predicted in Section 2.3.

### *4.1 Scenario 1: single parameter, single feature*

Calibrating the convection coefficient $h$ alone against $T_{\max}$ is the simplest test case. An uninformative uniform prior $h \sim \mathcal{U}(1000,\ 10\,000)\,\text{W/(m}^2\,\text{K)}$ is adopted, spanning a wide range around the true value $h_{\text{true}} = 8000\,\text{W/(m}^2\,\text{K)}$. Figure 3 shows the prior and posterior distributions obtained with $N_{\text{obs}} = 4$ replicates. SBALC converges rapidly as only 6 forward model evaluations are required (including 4 initial design points), which equals the minimal amount of samples. The coefficient of variation of the model evidence estimate is 0.0068, indicating a reliable approximation.

The posterior mean is $8002.36\,\text{W/m}^2\text{K}$ (true: $8000\,\text{W/m}^2\text{K}$), the posterior standard deviation is $19.8\,\text{W/m}^2\text{K}$. Increasing the number of observations, decreases the posterior standard deviation. When $N_{\text{obs}} = 100$, the posterior standard deviation is reduced to $6.7\,\text{W/m}^2\text{K}$.

### *4.2 Scenario 2: two parameters, one feature*

When only $T_{\max}$ is observed, many combinations of $h$ and $\varepsilon$ produce the same peak temperature. This is because both convective and radiative losses reduce $T_{max}$, so an increase in $h$ can be compensated by a decrease in $\varepsilon$ and vice

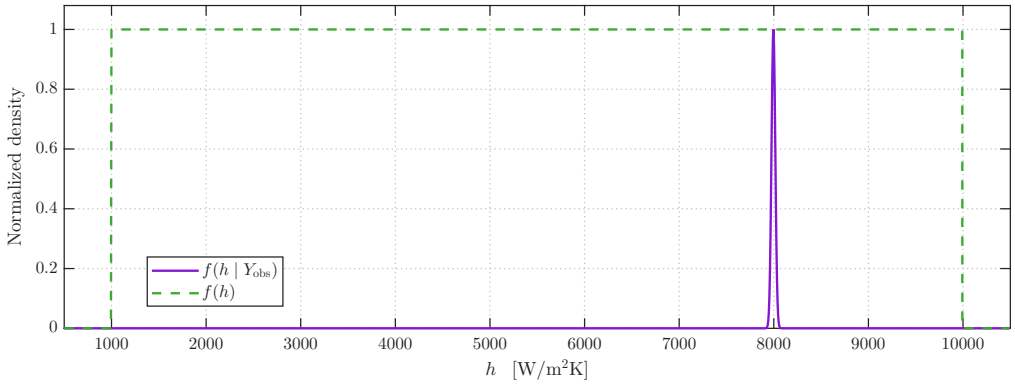


**Figure 3.** Scenario 1: prior (green) and posterior (purple) distribution of the convection coefficient $h$, calibrated against $T_{\max}$ alone

versa. Infinitely many parameter combinations therefore yield the same likelihood, and the data cannot distinguish between them. The posterior remains spread across all these combinations. This is an identifiability failure, independent of the inference algorithm.

The marginal posteriors $f_2(h \mid Y_{\text{obs}})$ and $f_2(\varepsilon \mid Y_{\text{obs}})$ reflect this: both remain broad, with posterior means $\hat{h} = 8063\,\text{W/m}^2\text{K}$ and $\hat{\varepsilon} = 0.791$, and posterior standard deviations $1092\,\text{W/m}^2\text{K}$ and $0.121$ respectively. 19 forward model evaluations were required to achieve this result.

### *4.3 Scenario 3: two parameters, two features*

Adding $\dot{T}_{\text{cool}}$ as a second feature resolves the identifiability problem. The posterior concentrates near $(h_{\text{true}}, \varepsilon_{\text{true}})$, as shown by the sharp peaks of the posteriors in Fig. 4.

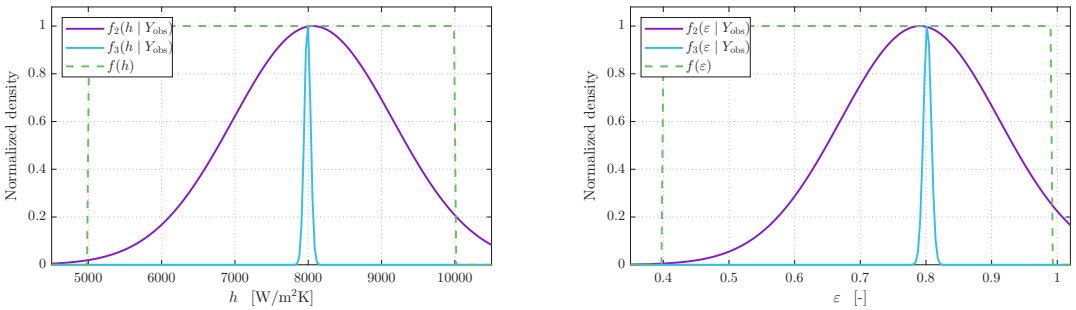


**Figure 4.** Marginal prior and posterior densities of $h$ (left) and $\varepsilon$ (right). The green dashed line is the shared prior. Purple: Scenario 2 posterior ($f_2$, one feature); blue: Scenario 3 posterior ($f_3$, two features).

The posterior means are $\hat{h} = 7990\,\text{W/m}^2\text{K}$ and $\hat{\varepsilon} = 0.799$, with posterior standard deviations $23.9\,\text{W/m}^2\text{K}$ and $0.0045$ respectively. 19 forward model evaluations were required to achieve this result.

The MLE of the noise standard deviations, evaluated at the posterior mean, yields $\hat{\sigma}_1 = 5.2$ and $\hat{\sigma}_2 = 43.9$, compared to the true generation values $\sigma_1 = 5$ and $\sigma_2 = 50$.

### *4.4 Scenario 4: full time-series calibration*

Replacing the scalar features with the full temperature time series ($m = 1000$ datapoints) is tested where only $h$ is unknown and $\varepsilon$ is fixed at $\varepsilon_{\text{true}}$. The SBALC active learning loop terminates prematurely with a Cholesky factorization failure when the GP surrogate of the log-likelihood is updated, typically after four to five active-learning iterations. Diagnostic logging of the training set at every iteration, summarized in Figure 5, reveals two reasons for this phenomenon.

First, the dynamic range of the log-likelihood is extreme. Training points placed near the prior boundaries yield $\mathscr{L}$-values on the order of $-10^7$, while points near $h_{\text{true}}$ yield values closer to zero. Even with output standardization applied during the Gaussian process hyperparameter optimisation, this spread forces the GP hyperparameter optimization to fit a very short kernel length scale (Figure 5, bottom-left). Second, the learning function of Eq. (5) preferentially selects new evaluation points in the neighbourhood of the high-likelihood region: at iteration 1 the minimum pairwise distance between training points is on the order of $10^3$ in $h$, but by iteration 4 it has collapsed to less than one unit (Figure 5, top-left). This effect can also be seen on the right panel as from the initial Hammersley design that spans the prior range, every acquired point is placed within a narrowing neighbourhood of $h_{\text{true}}$.

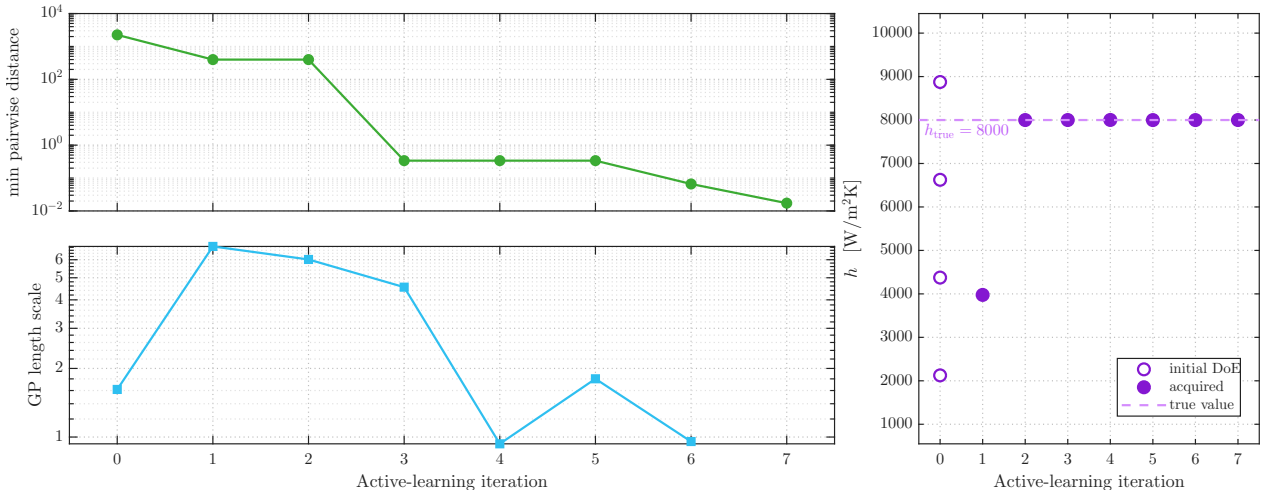


**Figure 5.** Diagnostic of SBALC applied to the full temperature time series ($m = 1000$, 1D calibration of $h$)
*Top-left*: minimum pairwise distance between training points *Bottom-left*: GP kernel length scale fitted by `fitrgp` at each iteration *Right*: training points at each iteration

However, when reducing the number of datapoints to a smaller $m$ (for example $m = 50$) the issue is resolved and the unknown parameter $h$ can be calibrated. As the dynamic range of the log-likelihood is roughly $20\times$ smaller, the GP fits a moderate length scale, and the active learning loop converges before clustering becomes an issue. In contrast, Scenarios 1 and 3 with feature-based outputs converge in a small number of forward model evaluations without any conditioning issues. Feature extraction therefore not only addresses the statistical pseudo-replication penalty discussed in Section 2.3 but also avoids the numerical fragility that affects active-learning-based GP surrogates on high-dimensional outputs.

## 5 Discussion and conclusions

The four scenarios test the proposed method of Section 2.3. Scenarios 1 to 3 follow the predicted pattern: one feature suffices for one parameter, fails for two, and adding a second feature with a different physical sensitivity restores identifiability. Scenario 4 confirms that full-signal calibration is numerically fragile. The diagnostic in Figure 5 shows that the extreme dynamic range of the log-likelihood forces the GP to fit a short kernel length scale and that the active learning loop clusters training points near the high-likelihood region. Reducing $m$ mitigates both, which is why $m = 50$ converges and $m = 1000$ does not.

The number and choice of features matters independently of the inference algorithm. For $d$ unknown parameters, at least $d$ features are needed, and their sensitivities to those parameters must be approximately linearly independent across the prior range. In the wire benchmark, $T_{\max}$ and the cooling slope $\dot{T}_{\text{cool}}$ satisfy this because convection scales with $T - T_\infty$ while radiation scales with $T^4 - T_\infty^4$.

Two physically motivated features recovered both unknown parameters of the wire benchmark accurately from a small number of forward model evaluations, while full-signal calibration with the same algorithm failed numerically. Feature-based outputs, that facilitate a dimensional reduction of time series, are therefore a necessity for surrogate-based Bayesian calibration of dynamic numercial models.

The next step is to apply this framework to a multi-physics finite element model of resistance spot welding. Dynamic resistance and electrode displacement are autocorrelated time signals for which full-signal calibration would inherit the same difficulties shown here. Candidate features, such as the dynamic resistance peak, the time and rate of the resistance drop at nugget formation, the electrode force plateau, and the electrode displacement at the end of welding [1, 10], each target a distinct physical mechanism and a distinct subset of the uncertain model parameters.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Research Foundation – Flanders (FWO) under grant 1SE1423N, the internal funding of KU Leuven grant C24E/21/026 and the support of the Flemish Government through the Flanders Make

## REFERENCES

[1] A. El Ouafi, R. Belanger, and M. Guillot, “Dynamic Resistance Based Model for On-Line Resistance Spot Welding Quality Assessment,” *Materials Science Forum*, vol. 706-709, pp. 2925–2930, Jan. 2012.

[2] P.-P. Li, C. Dang, C. H. Acevedo, M. A. Valdebenito, and M. G. R. Faes, "Bayesian model updating via streamlined Bayesian active learning cubature," Sept. 2025. arXiv:2509.11204 [stat].

[3] M. C. Kennedy and A. O'Hagan, "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 63, pp. 425–464, Sept. 2001.

[4] L. Bogaerts, M. G. Faes, and D. Moens, "A Data Driven Black Box Approach for the Inverse Quantification of Set-Theoretical Uncertainty," *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, vol. 11, Nov. 2024.

[5] J. L. Beck and S.-K. Au, "Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation," *Journal of Engineering Mechanics*, vol. 128, pp. 380–391, Apr. 2002. Publisher: American Society of Civil Engineers.

[6] J. Ching and Y.-C. Chen, "Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging," *Journal of Engineering Mechanics*, vol. 133, pp. 816–832, July 2007. Publisher: American Society of Civil Engineers.

[7] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, p. 107, 1993.

[8] D. Straub and I. Papaioannou, "Bayesian Updating with Structural Reliability Methods," *Journal of Engineering Mechanics*, vol. 141, p. 04014134, Mar. 2015.

[9] D. Rubin and D. Rubin, "Using the SIR algorithm to simulate posterior distributions," 1988.

[10] American Welding Society, Y.-J. Xia, T.-L. Lv, Y.-B. Li, H. Ghassemi-Armaki, and B. E. Carlson, "Quantitative Interpretation of Dynamic Resistance Signal in Resistance Spot Welding," *Welding Journal*, vol. 102, pp. 69–87, Apr. 2023.