HOKKAIDO UNIVERSITY

Title	Adaptive k-Nearest-Response Calibration for Likelihood-Free Stochastic Model Updating
Author(s)	Rocchetta, Roberto; Lorenzo, Nespoli; Medici, Vasco
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/2115/100431
Type	conference paper
File Information	C000038_1.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026),
28 June-01 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

Adaptive k-Nearest-Response Calibration for Likelihood-Free Stochastic Model Updating

Roberto Rocchetta[1], Nespoli Lorenzo[1], Vasco Medici[1, *]

[1]*Scuola Universitaria Professionale della Svizzera Italiana, SUPSI, Mendrisio, Via Ronchati, Switzerland*
E-mail: roberto.rocchetta@supsi.ch

Stochastic model updating methods seek identification of uncertain input parameters from empirical system responses. Often, likelihood functions are unavailable or computationally prohibitive, hence, simulation-based likelihood-free approaches such as Approximate Bayesian Computation (ABC) became growingly attractive in this setting. However, ABC application remains limited in practice due to high computational cost and weak parameter identifiability in complex, high-dimensional models. This paper introduces a likelihood-free calibration method that exploits pre-existing simulation archives to reduce computational burden. A k-nearest-neighbor search in response space allows retrieving the K closest simulated responses and mapping their associated archive realizations on the parameter space. This approach combined with a prior waiting of the archive samples yields a fully distribution-free posterior surrogate. To address data sparsity and non-identifiability, an adaptive resampling strategy selectively enriches the archive in regions of high posterior mass, and surrogate model support can further reduce computational burden. The approach is demonstrated on three representative benchmarks: (i) a synthetic quadratic model, (ii) a mechanical application based on the DLR-AIRMOD surrogate, and (iii) data from a building energy simulation model. A preliminary comparison with MCMC-based solutions demonstrates the method has potential to reduce the computational costs while yielding similar posterior reconstruction accuracy.

Keywords: Stochastic model updating, inverse problem, likelihood-free inference, model calibration, k-nearest

1 Introduction

Deterministic inverse problems estimate unknown parameters by minimizing the discrepancy between observed system responses and simulation outputs. However, such approaches do not explicitly represent uncertainty in data or models, and even in ideal conditions inverse problems are often ill-posed because non-injective models may map multiple input vectors to the same response, Tarantola (2005), Mottershead (1993), A. Simon et al. (2019). To address this, stochastic calibration and Bayesian model updating infer probability distributions over uncertain inputs instead of single parameter values. This framework is increasingly important in structural and reliability engineering, Kitahara et al (2022), prognostic and health management, e.g., Rocchetta et al. (2019), and for input uncertainty calibration for black-box models, e.g. see Gray, et al. (2022). In Bayesian updating, prior distributions are combined with empirical data and model responses (via a likelihood) into posterior distributions. In practice, likelihood evaluation is often unavailable or computationally expensive, and standard sampling strategies based on Markov Chaing sampling may require a very large number of forward model evaluations, see A. Lye (2021) for a tutorial on sampling methods in this context.

Although ABC methods avoid explicit likelihood specification, they remain costly when many simulations are needed. Furthermore, many practical use cases have existing archives of simulations, i.e., data sets of input-output model responses that remain underexploited in calibration and updating tasks. Such archives offer an opportunity to reduce computational effort while preserving the ability to perform uncertainty-aware inference. This work proposes a likelihood-free stochastic model updating approach that leverages these archives to speed up posterior identification. The method proposed adopts a K-Nearest (KNN) algorithm for finding a set of low-discrepancy responses and corresponding simulated inputs, kernel density estimators (KDE) in parameter space, and adaptive archive enrichment in sparsely supported but informative regions. In this way, the proposed framework reuses available simulations, reduces the number of additional model calls, and preserves the ability to represent non-Gaussian and multi-modal posterior structure, making it particularly suitable for large-scale reliability and risk-informed applications.

2 Problem statement

Consider a model $\tilde{y} = \mathcal{M}(\theta, \xi)$ that emulates the behavior of a physical system by mapping uncertain model parameters $\theta \in \mathbb{R}^{n_\theta}$ and controllable design variables $\xi \in \mathbb{R}^{n_\xi}$ to a response $\tilde{y} \in \mathbb{R}^{T \times n_y}$, where n_θ, n_ξ, and

$T \times n_y$ denote the dimensions of the parameter space, design space, and output space, respectively. A dataset of empirical observations is collected by repeating physical experiments under different design settings:

$$\mathcal{D} = \{\xi^e, Z^e\}_{e=1}^{n_e}, \quad Z^e = \{Z(y_i^e)\}_{i=1}^{N} \tag{1}$$

where N are the number of observations for each experiment and n_e is the total number of experiments. A summary statistic $Z: \mathrm{Y} \to \mathbb{R}^{n_z}$ compresses the high-dimensional $T \times n_y$ features to a lower dimension $n_z \ll T \times n_y$, for instance by computing moments over time or via dedicated performance or reliability indicators. We denote by, $\mathrm{z} = Z(\mathrm{y})$ the compression of a sample and by $Z^e = \{Z(y_i^e)\}_{i=1}^{N}$ the compression the dataset. A database, i.e., an archive, of simulated responses is defined as, $\mathcal{D}^{\text{archive}} = \left\{\xi^e, \{\theta_i^e, Z(\tilde{y}_i^e)\}_{i=1}^{N_{sim}}\right\}_{e=1}^{n_e}$, where each response $\tilde{y}_i^e = \mathcal{M}(\theta_i^e, \xi^e)$, is simulated for θ_i^e and the design ξ^e. This database may represent data from previous analysis, like surrogate modelling, sensitivity analysis, or uncertainty propagation tasks. The main goal of Bayesian model updating is to calibrate and update a posterior distribution for the unknown parameters, θ. This is done by updating prior knowledge using new experimental data, $\mathcal{D}$, while accounting for uncertainty like model errors, response variability and data scarcity. In this work, we will explore a Bayesian KNN-KDE updating approach that exploits this archive to speed up the procedure. The key objective here is, when possible, to exploit existing data to reduce computational burden, speeding up the updating.

2.1 *Approximate Bayesian computation*

In Bayesian model updating, uncertainty in the model input parameters θ is represented probabilistically through Bayes' theorem:

$$p_{\theta|\mathcal{D},\mathcal{M}} = \frac{p_{\mathcal{D}|\theta,\mathcal{M}}\, p_{\theta|\mathcal{M}}}{p_{\mathcal{D}|\mathcal{M}}}, \qquad p_{\mathcal{D}|\mathcal{M}} = \int p(\mathcal{D} \mid \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})\, d\theta \tag{2}$$

where $p_{\theta|\mathcal{M}}$ is the prior distribution encoding the state of knowledge about θ given the model $\mathcal{M}$, $p_{\mathcal{D}|\theta,\mathcal{M}}$ is the likelihood function, $p_{\theta|\mathcal{D},\mathcal{M}}$ is the posterior and $p_{\mathcal{D}|\mathcal{M}}$ the normalization constant (evidence). We will be dropping the conditioning on the model class for simplicity. Assuming independence across experiments, a log-likelihood function can be defined by summing up the individual experiment log-likelihoods, by further assuming conditional independence across samples the log-likelihood simplifies to, $p_{\mathcal{D}|\theta} = \sum_{e=1}^{n_e}\sum_{i=1}^{N} \log p(y_i^e|\theta, \xi^e)$. Note that for complex likelihood functions, estimation of the log-likelihood can be costly and not tractable. This motivates the adoption of approximate and likelihood-free approaches. When the likelihood is unavailable or costly to evaluate, ABC replaces it with a discrepancy-based comparison between observed and simulated data, this enables Bayesian inference without explicit likelihood evaluation. Let us denote a candidate distribution for the uncertain model inputs, parameterized by hyperparameters η as $\hat{p}_\eta(\theta)$. A simulation set $\widetilde{\mathcal{D}}(\eta)$, is a simulated dataset generated using model $\mathcal{M}$ and forward pass of samples from $\hat{p}_\eta(\theta)$ for the n_e designs. An parametric ABC approximation of the posterior over η is $\hat{p}_\eta(\theta|\mathcal{D}) \propto \hat{p}_\eta \widehat{LL}_e(\eta)$, where the log-likelihood is $\widehat{LL}_e(\eta) = -\rho\left(\mathcal{D}, \widetilde{\mathcal{D}}(\eta)\right)/\epsilon$ with $\epsilon > 0$ a scale parameter and ρ a discrepancy score. Note that smaller ϵ leads to peaked (lower variance) posteriors but can lack convergence if the model is inadequate. Depending on the available data, $\rho(\cdot,\cdot)$ may be computed on individual trajectories, summary statistics, or represent distributional summaries such as KL, Wasserstein, or Bhattacharyya distances, see e.g., Bi, S. et al. (2019, 2021) or Gray A., et al. (2022).

3 The proposed approach

In this paper, an adaptive ABC method based on KDE-KNN is proposed aiming to exploit a pre-computed archive of simulated input-design-output triplets. For each empirical summary z_i^e, this method identifies the K closest simulated response in the archive conditioning on ξ^e. Then, it maps them back to the associated input samples, which are then used to construct local densities for θ. The final posterior surrogate is obtained by combining these local factors across observations and designs, with important correction when the archive has been enriched using proposal distributions different from the prior. Large K-nearest-neighbour distances in response space are interpreted as indicators of poor local archive coverage. The corresponding response-space neighborhoods, mapped-back input sets, and local kernels are therefore flagged, and the archive is enriched by generating new simulations from a mixture of these flagged kernels together with a prior component. The procedure is then iterated until a stopping criterion is reached.

Step 1 KNN in the summary space and input reconstruction. Let $\tilde{z}_k^e = Z(M(\theta_k, \xi^e)), k = 1, \dots, N_{\text{sim}}$, for each empirical observation z_i^e the distances $\rho(z_i^e, \tilde{z}_k^e)$ are computed in summary space and the K-nearest simulated responses identified. These responses define an index set I_i^e, and a radius ε_i^e is computed as the largest distance for the responses in $k \in I_i^e$ which will be used as an adaptive local tolerance. The corresponding parameter-space neighborhood is then reconstructed by mapping the responses to the input yielding $\Theta_i^e = \{\theta_k : k \in I_i^e\}$, which identifies the archive samples that are close (low discrepancy) with the empirical observation for the design ξ^e. Note that, K is user-specified, whereas ε_i^e is data-adaptive, i.e., depends on the sparsity of the archive around z_i^e.

Step 2 Local conditional KED in parameter space. The retrieved set Θ_i^e is not itself a posterior; rather, it is used to construct a local approximation of a plausible set around z_i^e. Following the KNN-KDE in Olsen, K. et al (2024), local kernel weights are not assigned uniformly, but instead combine response-space proximity with archive weight correction terms $\widetilde{\omega}_{ik}^e = \frac{p(\theta_k)}{q(\theta_k)} \exp\left[-\frac{1}{2}\left(\frac{\rho(z_i^e, \tilde{z}_k^e)}{\varepsilon_i^e}\right)^2\right]$ and $\omega_{ik}^e = \frac{\widetilde{\omega}_{ik}^e}{\sum_{j \in I_i^e} \widetilde{\omega}_{ij}^e}$, for $k \in I_i^e$ and with $p(\theta)/\mathrm{q}(\theta)$ is the ratio between prior density and a proposal density from which the archive sample was generated. if the proposal density is unknown, we set the ratio to 1. The weights $\widetilde{\omega}_{ik}^e$and ω_{ik}^e are unnormalized and normalized local weights and quantify the relevance of archive sample θ_k for reconstructing the local density around z_i^e. The local mean vector and covariance matrix are then estimated as follows:

$$\hat{\mu}_i^e = \frac{1}{K} \sum_{k \in \mathcal{I}_i^e} \omega_{ik}^e \theta_k ; \qquad \hat{\Sigma}_{\mathrm{i}}^e = \sum_{k \in I_i^e} \omega_{ik}^e \ (\theta_k - \hat{\mu}_i^e)(\theta_k - \hat{\mu}_i^e)^{\top} + \lambda I, \tag{3}$$

where $\lambda > 0$ is a small ridge term introduced for numerical stability. The corresponding local conditional density is then approximated by the Gaussian kernel mixture $\hat{p}(\theta | z_i^e) = \sum_{k \in I_i^e} \omega_{ik}^e \ \mathcal{N}\left(\theta; \theta_k, \hat{\Sigma}_i^e\right)$.

Step 3 ABC posterior surrogate and adaptive re-sampling. For each design e, the local conditional densities are averaged over the empirical observations to define a design-wise pseudo-likelihood factor, $\hat{L}_e(\theta) = \frac{1}{N} \sum_{i=1}^{N} \hat{p}(\theta | z_i^e) \ \ z_i^e \in \mathcal{D}$. hese factors are then evaluated at each archive particle θ_k, and combined across all experiments to form the posterior surrogate, $\hat{p}(\theta \mid \mathcal{D}) \propto p(\theta) \prod_{e=1}^{n_e} \hat{L}_e(\theta)$. In practice, this posterior is represented on the archive support through normalized weights. If the archive has been generated directly from the prior, the unnormalized archive weights satisfy, $\mathrm{w_k} \propto \prod_{e=1}^{n_e} \hat{L}_e(\theta)$. If instead the archive includes samples drawn from adaptive proposal densities $q(\theta_k)$, then an importance correction is applied. Large values of the local response-space radius ε_i^e, diffuse local covariance matrices $\hat{\Sigma}_i^e$, or low local effective sample size are interpreted as indicators of poor local support in the archive. Adaptive re-sampling is performed by identifying the local model with the largest response-space radius and constructing a Gaussian mixture proposal from its mapped-back parameter samples. Specifically, if $\{\theta_j\}_{j=1}^K$are the mapped-back neighbors of the worst local KNN ball, with normalized local weights $\{w_j\}_{j=1}^K$and component-specific covariance matrices $\{\Sigma_j\}_{j=1}^K$, then new parameters are sampled from the proposal $q_{\text{new}}(\theta) = \sum_{j=1}^{K} w_j \ \mathcal{N}(\theta; \theta_j, \alpha \Sigma_j)$, where $\alpha > 0$is an inflation factor. These new samples are propagated through the forward model and appended to the archive. Here, to assess the quality of the calibration, sliced Wasserstein scores are computed for both the parameter and response distributions. In the parameter space, the Wasserstein score $W_\theta(f_\theta, \hat{p}_{\theta|\mathcal{D}})$ is estimated using samples from the calibrated posterior $\hat{p}_{\theta|\mathcal{D}}$ and a target distribution f_θ. Note that the latter is generally unavailable and only known in controlled simulated experiments. In the response space, the Wasserstein score $W_y(\mathcal{D}, \widetilde{\mathcal{D}})$ is estimated using the simulated responses in $\widetilde{\mathcal{D}}$ from samples of the calibrated posterior $\hat{p}_{\theta|\mathcal{D}}$ and the empirical outputs.

4 Test cases and results

Use Case 1 - Synthetic data and quadratic low-dimensional model: Use Case 1 was designed as a controlled synthetic benchmark based on the quadratic model under three data-generating mechanisms with increasing informativeness. This is a low-dimensional benchmark designed to verify key score indicators and posterior identification for a known ground truth. The deterministic simulator $\mathcal{M}(\theta, \xi)$ has a simple quadratic form, $y = \theta_1^2 + \frac{1}{2}\theta_1\theta_2(1 + \xi) + \frac{3}{2}(\theta_2 + \xi)^2$, where $\theta = (\theta_1, \theta_2)$ are the unknown parameters, ξ is a scalar that varies across experimental designs, and y a 1-dimensional output. Despite its simplicity, this map is non-injective, i.e., multiple θ, ξ vectors produce identical outputs, generating the multi-modal posteriors. We consider different empirical observations obtained from three distinct data-generating mechanisms (DGM): (1) Single design, single sample

and target point-vector $\theta^\star = (2,3)$; (2) Single design, multi-samples, and target bivariate Gaussian PDF, $\mathcal{N}(\mu, \mathrm{I}\sigma)$; (3) Multi-design, multi-samples, and target bivariate Gaussian PDF.

Table 1. Data availability and DGM for use case 1, including the target density f_θ and associated parameters.

	Sample size		Design	Target	
	n_e	N	$\xi^e \in$	f_θ	Parameters
DGM-1	1	1	$\{0\}$	$\delta(\theta-\theta^\star)$	$\theta^\star = (2,3)$
DGM-2	1	200	$\{0\}$	$\mathcal{N}(\mu, \mathrm{I}\sigma)$	$\mu = (2,3), \sigma = (0.5,0.5)$
DGM-3	6	200	$\{-2,-1,0,0.2,1.3,2\}$	$\mathcal{N}(\mu, \mathrm{I}\sigma)$	$\mu = (2,3), \sigma = (0.5,0.5)$

Table 1 summarizes: empirical observations, number of designs, n_e, the target, f_θ, used to generate N samples per design. For DGM-1 a Dirac $\delta(\theta-\theta^\star)$ at the true target $\theta^\star$ is considered, whilst uncorrelated $\mathcal{N}(\mu, \mathrm{I}\sigma)$ are used in the other cases.

As a figure of merit, we present a visualization of the input reconstruction from an archive $\mathcal{D}^{\text{archive}}$ with $N_{sim} = 10^4$ simulated pairs for each ξ^e, generated by uniform $\theta \sim U(-10, +10)$ and evaluating the quadratic model. Fig 1 illustrates the reconstruction with $K = 100$ nearest responses, where the union set $\bigcup_{i=1}^{N} \Theta_i^e$ are displayed by blue markers. For DGM -1, $N = 1$ obtained from $\theta^\star$ (red marker), the reconstructed inputs clearly reveal the model structure and non-uniqueness. DGM-2 remains weakly identifiable despite the availability of multiple samples (red markers) under a single design. DGM-3 shows interesting localization of the target region in red, because multiple designs provide more informative and partially amend for the non-invertibility of the model.

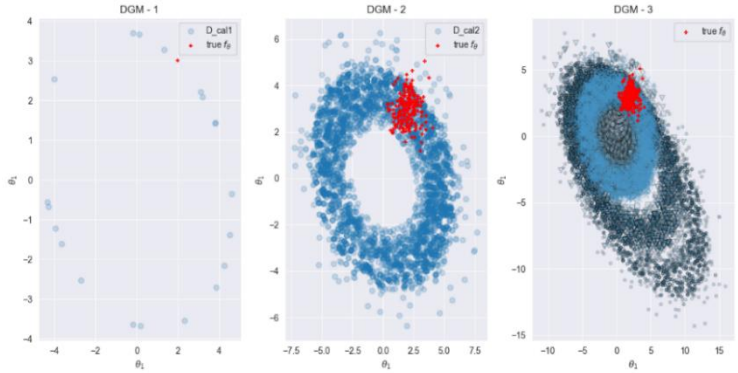

Figure 1: Reconstructed inputs via KNN for the three DGMs

Use Case 1 – Experimental setup and results: The proposed method was evaluated on Use Case 1 under three DGM scenarios using the same adaptive KNN-KDE settings. The initial archive contained 2000 simulations with $K = 100$; each refinement added 500 new simulations in weakly supported response regions, identified by large KNN radii. The refinement used $\lambda = 0.01$, $\alpha = 0.95$, and early stopping when the maximum-radius improvement fell below 0.001. Table 2 reports the archive size, target-space W_θ, response-space W_z, maximum radius $\overline{\varepsilon_i^e} = \max_{i,e} \varepsilon_i^e$, and mean radius $\mu_{\varepsilon_i^e}$. Figure 2 shows the resulting posterior distributions for DGM-2 and DGM-3, indicating overall good identification of the target distributions. Overall, adaptive enrichment reduced both mean and maximum radii in all cases, indicating improved local archive support. Note that W_θ, and W_z, changed only slightly. This is likely due low dimensionality of the problem limiting effectiveness of the proposed adaptive approach.

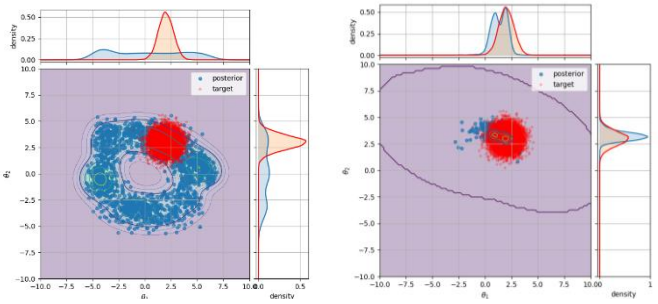

Figure 2 Posterior (blue) vs target density (red) for the DGM-2 (right) and DGM-3 (left).

Table 2: Results of the updating procedure for the 3 data generating mechanisms and use case 1.

USE CASE 1			RESPONSE KNN		WASSERSTEIN SCORES	
DGM	**Iter**	$\lvert\mathcal{D}^{\text{archive}}\rvert$	$\mu_{\varepsilon_i^e}$	$\overline{\varepsilon_i^e}$	W_θ **(target)**	W_z **(out)**
1	1	2000	0.068	0.068	3.95	0.44
1	2	2500	0.025	0.025	3.89	0.54
1	3	3000	0.016	0.016	3.68	5.34
1	4	3500	0.013	0.013	3.73	6.33
1	5	4000	0.010	0.010	1.64	1.11
2	1	2000	0.0832	0.299	3.29	2.71
			…			

2	7	5000	0.032	0.061	3.36	2.31
2	8	5500	0.030	0.044	3.32	1.70
3	1	2000	0.107	1.315	0.54	4.99
				…		
3	6	4000	0.043	0.081	0.48	5.15
3	7	4500	0.036	0.064	0.64	5.20

Use Case 2 – DLR-AIRMOD surrogate. This case study evaluates the proposed method on a higher-dimensional engineering benchmark based on a surrogate of the DLR-AIRMOD finite-element model. Empirical data is collected from the repository in Grashorn et al. (2024). The calibration involves 12 uncertain inputs: 11 normalized mass and stiffness parameters and one dispersion parameter. The model outputs 10 modal response features, including roll, pitch, heave, bending, and torsional modes. The target posterior mode is set to $\theta^{\star} = (0.3, \ldots, 0.3, -0.6)$. Unlike Use Case 1, no design variable is varied, so the analysis considers a single calibration setting. This benchmark is used to assess the method in a realistic surrogate-based stochastic updating problem, where repeated evaluations of the original simulator would be computationally expensive.

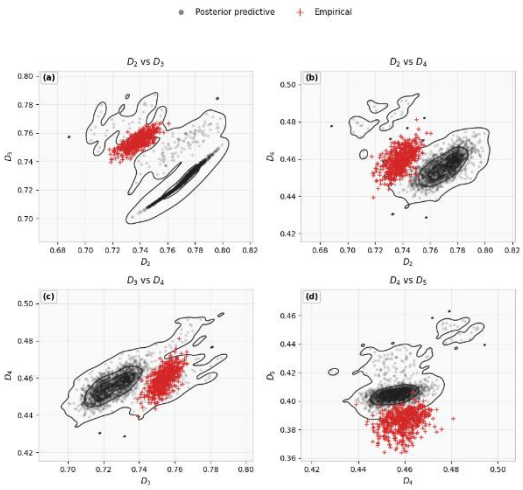

Figure 3. Selected pairwise posterior predictive distributions in output space. Posterior predictive samples and KDE contours are shown in blue, while empirical observations are shown in red.

Use Case 2 – Experimental setup and results: An empirical dataset with 500 observations is used, together with an initial archive of 5000 surrogate-generated input-output pairs. The adaptive KNN-KDE procedure enriches the archive by 500 simulations per iteration, targeting weakly supported regions in response space. As reported in Table 3, the mean KNN radius decreases from 1.699 to 0.353 and the maximum radius from 1.734 to 0.443 as the archive grows to 11000 samples. Over the same iterations, the response-space Wasserstein distance W_Z decreases from 0.0957 to approximately 0.019, showing a clear improvement in posterior-predictive agreement. Figure 3 confirms that the posterior predictive distribution captures the main empirical modal-response patterns, although some local offsets remain. Figure 4 shows that the inferred posterior marginals are in good qualitative agreement with the TMCMC and TM reference estimates of Grashorn et al. (2024), with some discrepancies in weakly constrained parameter directions. Overall, the results indicate that the proposed adaptive procedure can recover the main posterior structure while reducing the computational burden. A systematic assessment of computational cost and efficiency gains is left for a future extension of this work.

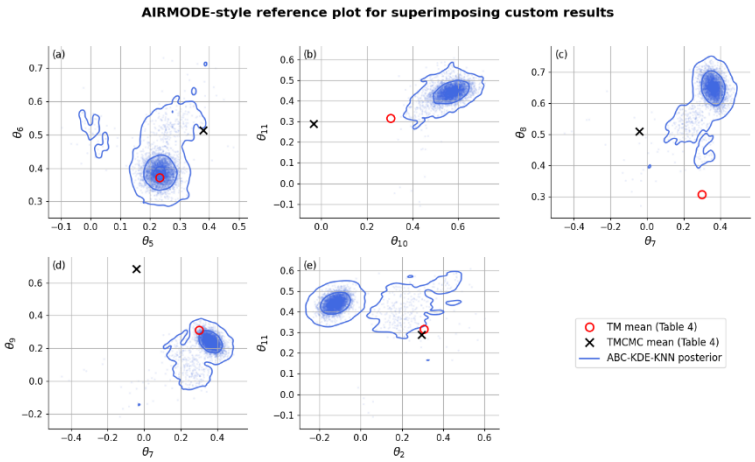

Figure 4. Selected bivariate marginals of the AIRMODE posterior distribution. Posterior samples and KDE contours from the proposed method are compared with the reference, Grashorn et al. (2024)

Table 3: Results of the ABC-KNN-KDE procedure on the 2nd use case AIRMODE

ITER	$\lvert\mathcal{D}^{ARCHIVE}\rvert$	$\mu_{\varepsilon_i^e}$	$\bar{\varepsilon}_i^e$	W_Z *(OUT)*
1	5000	1.69	1.73	0.0959
2	5500	1.31	1.35	0.0639
3	6000	1.06	1.09	0.0356
		…		
12	10500	0.371	0.467	0.0197
13	11000	0.352	0.442	0.0190

Case 3 – Energy plus simulator. As a third case study, we considered the Energy Plus-based dataset of Altieri et al. (2025), which provides a higher-dimensional building-performance benchmark generated from controlled simulation experiments. Due to space limitations, the full results are not reported in this work. Preliminary analyses are consistent with the trends observed in the previous cases, showing improved local archive support and posterior-predictive agreement after adaptive enrichment. A more detailed presentation of this benchmark, including posterior diagnostics and accuracy metrics, will be included in a future extension of this work.

5 Discussion, preliminary conclusions, and future work

The results support the main premise of the proposed approach: a simulation archive can be reused as a likelihood-free posterior surrogate and improved through targeted adaptive enrichment. In Use Case 1, the mean and maximum KNN radii decreased across the three data-generating mechanisms, indicating improved local support around the empirical responses. The benefits were strongest in DGM-3, where multiple designs helped mitigate the non-injectivity of the forward map and led to better posterior localization. In contrast, DGM-2 remains less identifiable, showing that improved response-space coverage does not always guarantee accurate recovery of the latent parameter distribution. The AIRMOD case 2 further confirms the relevance of the method in a higher-dimensional engineering setting. Adaptive enrichment improved posterior-predictive agreement and reduced response-space discrepancies, while the inferred posterior showed reasonable agreement with reference TMCMC and transport-map estimates for several parameter projections. Some inconsistencies remain, especially along weakly constrained directions, highlighting the dependence of the method on the informativeness of the selected summaries and the local density of the archive. Overall, the method appears promising as an adaptive kernel-ABC surrogate that reduces the need for repeated full-scale simulation. Future work will provide a more systematic comparison against TMCMC, transport-map methods, and ABC-MCMC in terms of posterior accuracy, simulator calls, and computational cost.

Acknowledgments

This work was funded by the Driving Urban Transitions (DUT) Partnership (Call 2024) and co-funded by the European Commission within the project PEDAID3, *Positive Energy District supported by AI and Data-Driven Decisions*

References

Tarantola, A., Inverse problem theory and methods for model parameter estimation, *Society for industrial and applied mathematics*, SIAM, 2005.

Mottershead, John E., and M. I. Friswell, Model updating in structural dynamics: a survey*, Journal of sound and vibration*, Vol 167.2, 347-375, 1993.

Simon A., Maass P., Öktem O. and Carola-Bibiane Schönlieb, Solving inverse problems using data-driven models, *Acta numerica*, Vol 28, 1-174, 2019.

Lye A., Cicirello A. and Patelli E. Sampling methods for solving Bayesian model updating problems: A tutorial, *Mechanical Systems and Signal Processing,* Vol 159, 107760, 2021.

Gray, A., Wimbush A., de Angelis M., et al., From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information, *Mechanical Systems and Signal Processing*, Vol 165, 108210, 2022.

Kitahara, M., Bi, S., Broggi, M., and Beer M., Nonparametric Bayesian stochastic model updating with hybrid uncertainties, *Mechanical Systems and Signal Processing*, Vol 163, 108195, 2022.

Grashorn J., Broggi M., Chamoin L, Beer M., Efficiency comparison of MCMC and Transport Map Bayesian posterior estimation for structural health monitoring, *Mechanical Systems and Signal Processing*, Vol 216, 2024.

Altieri D., Saez R., Perez M. and Branca G., Physics-Informed Neural Network for Building Energy Demand Prediction, *Journal of Physics: Conference Series*, Vol 3140, 5, 2025.

Olsen, K., et al., Think Global, Adapt Local: Learning Locally Adaptive K-Nearest Neighbor Kernel Density Estimators, *Proceedings of Machine Learning Research*, 238:4114 - 4122, 2024.

Bi, S., Broggi, M., Beer, M., The role of the Bhattacharyya distance in stochastic model updating, *Mechanical Systems and Signal Processing*, 117, 437-452, 2019

Bi, S., et al., Optimization or Bayesian strategy? Performance of the Bhattacharyya distance in different algorithms of stochastic model updating. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems,* 7(2), 020903, 2021.

Frogner, C., et al., A. Learning with a Wasserstein loss. *Advances in neural information processing systems*, 28, 2015.

Rocchetta, R., et al., On-line Bayesian model updating for structural health monitoring, *Mechanical Systems and Signal Processing*, 103, 174-195, 2018.