MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Sequential Model Calibration of Vapor Compression Cycles using Approximate Bayesian Computation

Zinage, Shrenik; Dixit, Vaibhav; Chakrabarty, Ankush; Qiao, Hongtao; Laughman, Christopher R.; Bilionis, Ilias; Deshpande, Vedang M.

TR2026-061 May 28, 2026

Abstract

Physics-based simulation models are essential for the development of vapor compression cycles (VCCs), the core technology behind most modern refrigeration and air conditioning systems. These models enable robust control, monitoring, fault detection and diagnostics, and digital twin technologies. However, nonlinear dynamics, high-dimensional parameter and state spaces, numerical stiffness, and the limited integration of conventional modeling environments with scientific machine learning workflows create significant challenges for efficient, transferable, and automated calibration. Existing approaches typically rely on deterministic methods and lack mechanisms for principled knowledge transfer across calibration tasks, while also failing to quantify epistemic and aleatoric uncertainty. To address these limitations, we propose a Bayesian calibration framework for VCC systems that explicitly quantifies various sources of uncertainty in model predictions. The framework supports transferability across datasets by sequentially up- dating informative priors from previously inferred posteriors. We implement and evaluate this approach on a commercially available high-fidelity Julia based dynamic VCC model, demonstrating its ability to successfully estimate key system parameters.

American Control Conference (ACC) 2026

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Sequential Model Calibration of Vapor Compression Cycles using Approximate Bayesian Computation

Shrenik Zinage[1], Vaibhav Dixit[2], Ankush Chakrabarty[3], Hongtao Qiao[4], Christopher R. Laughman[4], Ilias Bilionis[5], and Vedang M. Deshpande[†,4]

***Abstract*— Physics-based simulation models are essential for the development of vapor compression cycles (VCCs), the core technology behind most modern refrigeration and air conditioning systems. These models enable robust control, monitoring, fault detection and diagnostics, and digital twin technologies. However, nonlinear dynamics, high-dimensional parameter and state spaces, numerical stiffness, and the limited integration of conventional modeling environments with scientific machine learning workflows create significant challenges for efficient, transferable, and automated calibration. Existing approaches typically rely on deterministic methods and lack mechanisms for principled knowledge transfer across calibration tasks, while also failing to quantify epistemic and aleatoric uncertainty. To address these limitations, we propose a Bayesian calibration framework for VCC systems that explicitly quantifies various sources of uncertainty in model predictions. The framework supports transferability across datasets by sequentially updating informative priors from previously inferred posteriors. We implement and evaluate this approach on a commercially available high-fidelity Julia based dynamic VCC model, demonstrating its ability to successfully estimate key system parameters.**

I. Introduction

Vapor compression cycles (VCCs) are the core thermodynamic technologies underlying modern refrigeration, heat pump, and air conditioning systems. Beyond providing essential thermal comfort and refrigeration, VCCs also act as flexible demand side resources, capable of interacting with variable renewable generation to improve grid stability and accelerate decarbonization of energy networks. Therefore, they are positioned as a key technology to mitigate the impacts of climate change on human health and infrastructure resilience.

Simulation models of VCCs that capture their dynamic behavior allow a detailed analysis of transient responses and component interactions that cannot be easily observed experimentally. Dynamic simulations of VCCs have supported the development of advanced control strategies [1], state estimation methods [2], [3], [4], and digital twin technologies for monitoring and optimization [5]. However, the accuracy and performance of these technologies depend critically on the fidelity of the model. Without proper calibration to experimental data and system specifications, models risk deviating from physical reality.

Physics-based models of VCCs are typically formulated as systems of differential-algebraic equations (DAEs). These models are highly valued for their physical fidelity but are also characterized by nonlinearities, high numerical stiffness, and large parameter spaces. The discretization of governing conservation laws into finite volumes yields models with very large state dimensions, leading to substantial computational costs that are further exacerbated by stiffness across widely varying time scales. Moreover, because VCCs involve tightly coupled multiphysical phenomena, fluid flow, heat transfer, and phase change, such models are usually developed in equation oriented programming environments. While effective for simulation, traditional modeling tools are not always well suited to modern calibration workflows that require advanced optimization and statistical techniques. Collectively, these factors make the calibration of physics based VCC models an inherently difficult and computationally demanding problem.

Traditional VCC calibration typically uses least squares error minimization or evolutionary algorithms to fit model response to experimental data [5]. While straightforward, these approaches often become inefficient for dynamic models where nonlinearities, stiffness, and large parameter spaces worsen convergence difficulties. Bayesian optimization [6], [7], [8] has become a widely used method for calibrating HVAC and building system models, offering global optimization without the need for gradient information. Recent advances have extended these methods to operate in reduced latent dimensions, thereby lowering computational costs [6]. While effective in many calibration tasks, existing approaches exhibit important limitations. They lack a principled way to transfer knowledge across successive calibration problems and are generally restricted to deterministic parameter estimation. As a result, they fail to quantify epistemic uncertainty introduced by uncertain parameters, limiting their utility for uncertainty-aware control and diagnostics. These limitations motivate the development of a statistically rigorous, computationally tractable framework that supports sequential calibration, uncertainty quantification, and efficient transferability across datasets.

Contributions: First, we propose a sequential Bayesian calibration framework for VCC models that transfers infor-

[1]Shrenik Zinage is a Graduate Student in School of Mechanical Engineering, Purdue University, West Lafayette, IN. This work was completed during an internship at Mitsubishi Electric Research Laboratories.

[2]Vaibhav Dixit is with Meta, Menlo Park, CA. This work was completed during an internship at Mitsubishi Electric Research Laboratories.

[3]Ankush Chakrabarty is with Insulet Corporation, Acton, MA. He was with Mitsubishi Electric Research Laboratories at the time of this research.

[4]Authors are with Mitsubishi Electric Research Laboratories, Cambridge, MA.

[5]Ilias Bilionis is with the Faculty of School of Mechanical Engineering, Purdue University, West Lafayette, IN.

[†]Corresponding author. vmdeshpande@ieee.org

mation across datasets by using posteriors from one calibration task as informative priors for the next. Unlike existing methods that yield only deterministic estimates, our probabilistic approach produces posterior distributions over parameters, explicitly quantifying parametric uncertainty. Second, to illustrate the practical utility of our Bayesian calibration framework, we present a case study on detecting refrigerant mass leakage, a common degradation mode in aging VCC systems. To address numerical drift, we introduce a feedback controller that exponentially stabilizes the otherwise neutrally stable manifold defined by the total refrigerant mass. This ensures that mass remains conserved in the simulation in the absence of leakage, making it immune to numerical artifacts and allowing a clear distinction between physical faults and solver-induced errors. Consequently, refrigerant mass can be treated as a single calibration parameter, reducing the search space from 2N initialization variables to 1, where N is the number of discretization volumes in the DAEs, thereby significantly lowering the computational cost of calibration tasks. Finally, we demonstrate our approach on a high-fidelity, Julia-based dynamic model of a commercially available VCC system.

The rest of the paper is organized as follows. In Section II, we provide an overview of the VCC model and its implementation, followed by our calibration methodology in Section III. Numerical results are discussed in Section IV, with concluding remarks in Section V. The appendix provides a brief discussion on sensitivity analysis.

II. Dynamic Models of Vapor Compression Cycles

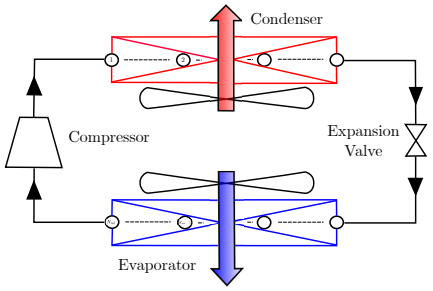

Fig. 1. Structure of a vapor compression cycle.

A variable-capacity VCC, shown in Fig. 1, includes a variable speed compressor, a modulating expansion valve, and two refrigerant-to-air heat exchangers (HEXs) with variable speed fans. Thermal energy is transferred from the evaporating HEX airflow to the condensing HEX airflow through the refrigerant by exploiting the latent heat of condensation and evaporation. As the HEXs dominate the cycle dynamics, they are modeled using systems of DAEs, whereas the compressor and expansion valve are represented with simpler algebraic relations. The HEX models rely on simplifying assumptions: one-dimensional refrigerant flow, instantaneous thermodynamic equilibrium within each discretized control volume, negligible gravity, and equal liquid–vapor velocities in the two-phase region. Under these assumptions, the HEX model forms an index-1 DAE system describing refrigerant flow, tube wall thermal behavior, and air-side convection.

Assuming ρ is density, A is the flow cross-sectional area, v is velocity, P is pressure, F_f is frictional pressure drop, u is internal energy, h is enthalpy, and Q is heat transfer rate, the refrigerant pipe model enforces conservation of mass, momentum, and energy using spline-based refrigerant property representations [9] as shown below:

$$\begin{aligned} &\frac{\partial(\rho A)}{\partial t} + \frac{\partial(\rho A v)}{\partial x} = 0, \\ &\frac{\partial(\rho v A)}{\partial t} + \frac{\partial(\rho v^2 A)}{\partial x} = -A\frac{\partial P}{\partial x} - F_f, \\ &\frac{\partial(\rho u A)}{\partial t} + \frac{\partial(\rho v h A)}{\partial x} = vA\frac{\partial P}{\partial x} + vF_f + \frac{\partial Q}{\partial x} \end{aligned} \tag{1}$$

The pipe wall was modeled with an ODE with a single thermal capacitance, while airflow was represented algebraically due to negligible energy storage. The equations were discretized into finite control volumes with a staggered grid scheme. Although not explicitly depicted in Fig. 1, the system model also incorporates a suction tank, represented as a single volume with fluid storage and wall, as well as pressure-drop devices modeled as simple pressure losses along the fluid flow. Further details on component models are available in [10], [11].

A. Parameterization

The simulation models are defined by several categories of parameters: geometric (e.g., pipe lengths, wall thicknesses, and cross-sectional areas), thermodynamic (e.g., material densities and heat transfer coefficients), model-specific coefficients (e.g., efficiency and correction factors), and computational (e.g., number of finite volumes and solver tolerances). Computational parameters were fixed based on a preliminary study that evaluated the trade-off between numerical accuracy and computational efficiency for a nominal model; once selected, these values were not modified. System parameters such as certain geometric dimensions and material densities, which can be directly obtained or reliably estimated from the system's technical specification sheet, also offer little opportunity for calibration. Consequently, these parameters were likewise fixed to ensure consistency between the physical system and the model representation. The parameters chosen for calibration and sensitivity analysis of model outputs are discussed in the appendix.

B. Regulation of refrigerant mass using PI feedback

Prior studies have shown that VCC models derived under standard assumptions, such as the mass conservation equation (1), do not strictly conserve refrigerant mass, leading to significant practical issues [12]. This artificial drift of the total system charge, although no real change occurs, arises purely from numerical errors and manifests as an increase or decrease in the refrigerant inventory [13]. This phenomenon is particularly problematic for long duration simulations, such as seasonal building performance studies or automotive drive cycles, where even small drift compounds into significant errors. In addition, calibration becomes unreliable, as

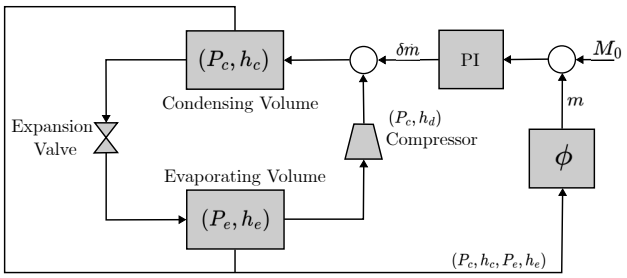

Fig. 2. Dynamic VCC model augmented with a PI controller to enforce conservation of total refrigerant mass.

parameter identification routines attempt to compensate for artificial mass variations rather than true physical behavior, resulting in non-physical parameter sets and misleading performance predictions.

Recent work [13] rigorously analyzed this issue and traced its origin to the mathematical structure of standard formulations. These models enforce mass conservation only through the derivative condition $\dot{M}_{\text{Total}} = 0$, without explicitly imposing the constraint $M_{\text{Total}} = \phi(x) = \text{const}$. Analytically, trajectories initialized on the manifold defined by M_{Total} remain confined to it and conserve mass. In practice, however, numerical solver errors perturb the dynamics along the neutrally stable direction that traverses this manifold, causing the solution to drift rather than return to equilibrium. Over time, these accumulated errors manifest as the apparent drift or "leakage" of refrigerant mass.

To address this issue, the model is augmented with a virtual proportional–integral (PI) feedback loop, shown in Fig. 2, which regulates the total mass $M_{\text{Total}} = \phi(x)$ to its desired nominal value M_0, adopting the approach proposed in [13]. Implemented as a synthetic mass flow source at the compressor discharge, this controller drives the error $e(t) = M_0 - \phi(x)$ to zero, thereby stabilizing the total refrigerant charge. The PI feedback shifts the zero eigenvalue associated with mass drift into the left-half plane, rendering the equilibrium exponentially stable. For sufficiently small PI gains, this correction leaves the dominant cycle dynamics on the invariant manifold nearly unchanged [13]. This approach offers two key advantages. First, it guarantees conservation of total mass and eliminates drift, thereby improving the reliability of calibration methods. Second, the refrigerant mass can be initialized with a single parameter M_0, rather than specifying $2N$ thermodynamic variables (pressure and enthalpy) across N discretized fluid volumes, which substantially reduces the computational cost of calibration.

C. Implementation

The model was implemented in Julia using `ModelingToolkit.jl` [14], a symbolic computational algebra framework that facilitates the construction of large scale acausal system models. This is achieved by composing and inheriting smaller component models, allowing flexible and modular development. Such model representation can then be symbolically transformed and compiled into efficient imperative Julia code, making the models directly compatible with high performance numerical tools, including state of the art differential equation solvers and optimization libraries.

Assuming f_d represents the governing differential equations and f_a are the algebraic constraints, this model results in a set of index-1 DAEs which are represented as follows

$$\begin{aligned}\dot{x}(t) &= f_d(\mathbf{x}(t), u(t), w(t)),\\ 0 &= f_a(\mathbf{x}(t), u(t), w(t)),\end{aligned}$$

where $u(t)$ and $w(t)$ denote control inputs and external disturbances acting on the system. The state vector $\mathbf{x}$ includes both differential and algebraic variables. Due to the inherent dynamic properties of VCC models, this system exhibits high numerical stiffness [2]. We use standard stiff DAE solvers to integrate the dynamics forward in time.

III. Methodology

In this section, we first review the basics of approximate Bayesian computation (ABC) and then present our ABC rejection based calibration methodology.

A. Approximate Bayesian Computation (ABC)

ABC [15] is a family of methods for Bayesian inference in scenarios where the likelihood function cannot be evaluated in closed form, but forward simulation of the model is feasible. Consider a probabilistic model M parameterized by $\gamma \in \Theta$, where Θ denotes the parameter space. The observed data D_{obs} originates from an unknown true process. Assuming $p(D_{\text{obs}}|\gamma)$ is the likelihood and $p(\gamma)$ is the prior, traditional Bayesian inference evaluates the posterior distribution as:

$$p(\gamma|D_{\text{obs}}) \propto p(D_{\text{obs}}|\gamma)p(\gamma).$$

For some scenarios, the true likelihood function $p(D_{\text{obs}}|\gamma)$ is often unavailable. ABC is a likelihood free inference method that bypasses the direct evaluation of the likelihood function. Instead, ABC uses repeated simulations from the model to compare observed data with simulated data based on a chosen distance metric or a set of summary statistics.

Fig. 3 illustrates the core idea behind ABC. Initially, a parameter value γ is sampled from the prior distribution $p(\gamma)$. Using this parameter, simulated data D_{sim} is generated from the probabilistic model M, which defines $p(D|\gamma)$. Subsequently, the similarity between the simulated data D_{sim} and the observed data D_{obs} is evaluated by computing a distance metric $\Delta(D_{\text{sim}}, D_{\text{obs}})$. If the computed distance is less than or equal to a pre-specified tolerance threshold ϵ, the sampled parameter value γ is retained; otherwise, it is rejected. This process is repeated iteratively to approximate the posterior parameter distribution. Assuming $\mathbb{I}[\cdot]$ represents the indicator function, the posterior is approximated as:

$$p_{\text{ABC}}(\gamma|D_{\text{obs}}, \epsilon) \propto \int \mathbb{I}[\Delta(D_{\text{sim}}, D_{\text{obs}}) \leq \epsilon] p(D_{\text{sim}}|\gamma)p(\gamma)\, dD_{\text{sim}}.$$

The accuracy of ABC depends critically on the tolerance ϵ and the distance metric Δ. As $\epsilon \to 0$, and provided that Δ captures all relevant information in the data, the

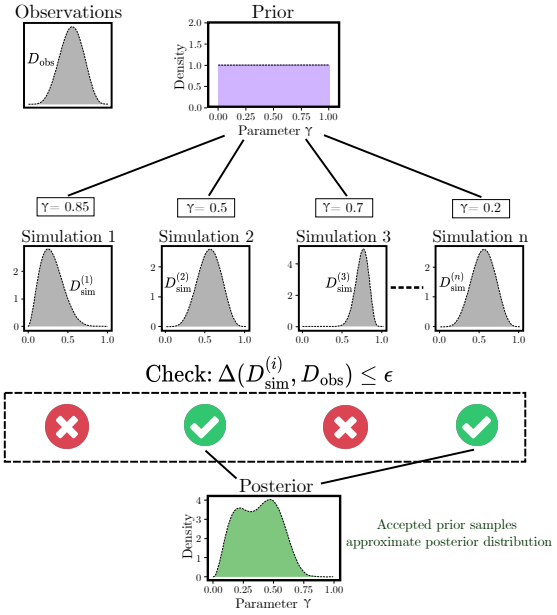

Fig. 3. Parameter estimation using Approximate Bayesian Computation.

ABC posterior $p_{\text{ABC}}(\gamma|D_{\text{obs}}, \epsilon)$ converges to the exact posterior $p(\gamma|D_{\text{obs}})$. This is rigorously proved mathematically in [16]. In practice, however, the choice of ϵ introduces a fundamental tradeoff: smaller thresholds yield posterior approximations that are closer to the true distribution but require a substantially larger number of simulations to obtain sufficient accepted samples, whereas larger thresholds reduce computational demands but admit parameter values that generate poor matches to the observed data, thereby degrading the fidelity of the posterior approximation.

B. VCC calibration using ABC rejection

Recognizing the limitations of gradient-based methods for models with ill-conditioned gradients, which can lead to computational difficulties, we use an ABC rejection [17] approach, a well known variant of ABC for calibration, as outlined in Algorithm 1. The calibration process begins with a pilot sampling phase to determine an appropriate acceptance threshold. We draw multiple samples from the prior and simulate the model outputs for each parameter set. The model outputs consist of temperatures and pressures at the compressor discharge (TD, PD) and suction (TS, PS) ports, in the condenser (TC, PC) and the evaporator (TE, PE), along with the compressor mass flow rate ($\dot{m}_{\text{comp}}$). This choice of outputs is motivated by the in-house experimental setup that will be used to validate the proposed approach in future.

To evaluate the discrepancy between simulation and observed data, we use the Kolmogorov–Smirnov (KS) statistic as the distance metric. In the context of this work, it provides a robust method for comparing the time series data generated by a simulation against the observed (or ground truth) data. Unlike metrics that only compare point wise errors, the KS statistic compares the overall distributional properties of the two time series. The statistic is based on the empirical cumulative distribution function (ECDF). For a given one dimensional dataset $X = \{x_1, x_2, \ldots, x_n\}$, the

Algorithm 1 Sequential model calibration using ABC

Require: Model M, datasets $\{D_{\text{obs}}^{(k)}\}_{k=1}^{K}$, inputs $\{U^{(k)}\}_{k=1}^{K}$, initial uniform prior sampler $\texttt{Prior}^{(1)}$ defined based on physics informed bounds, pilot draws N_{pilot}, main draws N_{main}, desired accepted samples N_{desired}, Quantile $\zeta \in (0,1)$, output index set $J = \{1, \ldots, n_y\}$, number of calibration parameters for VCC model n_{model}, parameter ranges $\mathcal{R} = \{[a_j, b_j]\}_{j=1}^{n_{\text{model}}+n_y}$, distance metric Δ

Ensure: Posterior sample sets $\{\mathcal{S}^{(k)}\}_{k=1}^{K}$

for $k \leftarrow 1$ **to** K **do**

$Y_{\text{obs}} \leftarrow D_{\text{obs}}^{(k)}$ and set $n_y \leftarrow |J|$ and $n_t \leftarrow$ number of columns of Y_{obs}

if $k = 1$ **then**

Set $\texttt{Prior}^{(k)} \leftarrow \texttt{Prior}^{(1)}$

else

$\texttt{Prior}^{(k)} \leftarrow \mathcal{S}^{(k-1)}$ ▷ posterior from dataset $k-1$ becomes prior for dataset k

end if

Pilot sampling phase to select the tolerance

$\mathcal{D}_{\text{pilot}} \leftarrow \varnothing$

for $i \leftarrow 1$ **to** N_{pilot} **do**

$\theta^{(i)} = [\gamma^{(i)}; \nu^{(i)}] \sim \texttt{Prior}^{(k)}$

$Y_{\text{sim}}^{(i)} \leftarrow M(\gamma^{(i)}, U^{(k)})$

$S_{j,t}^{(i)} = \nu_j^{(i)} \cdot |Y_{\text{sim}}(j,t)|$ where $S \in \mathbb{R}^{n_y \times n_t}$

Draw $Z \sim \mathcal{N}(0,1)$ where $Z \in \mathbb{R}^{n_y \times n_t}$

$Y_{\text{sim}}^{(i)} \leftarrow Y_{\text{sim}}^{(i)} + S \odot Z$

$d_i \leftarrow \Delta(Y_{\text{sim}}^{(i)}, Y_{\text{obs}}, J)$

append d_i to $\mathcal{D}_{\text{pilot}}$

end for

$\epsilon^{(k)} \leftarrow \texttt{Quantile}(\mathcal{D}_{\text{pilot}}, \zeta)$

Main sampling phase with the fixed tolerance

$\mathcal{S}^{(k)} \leftarrow \varnothing$

for $i \leftarrow 1$ **to** N_{main} **do**

Draw $\theta^{(i)} = [\gamma^{(i)}; \nu^{(i)}] \sim \texttt{Prior}^{(k)}$

$Y_{\text{sim}}^{(i)} \leftarrow M(\gamma^{(i)}, U^{(k)})$

$S_{j,t}^{(i)} = \nu_j^{(i)} \cdot |Y_{\text{sim}}(j,t)|$ where $S \in \mathbb{R}^{n_y \times n_t}$

Draw $Z \sim \mathcal{N}(0,1)$ where $Z \in \mathbb{R}^{n_y \times n_t}$

$Y_{\text{sim}}^{(i)} \leftarrow Y_{\text{sim}}^{(i)} + S \odot Z$

$d_i \leftarrow \Delta(Y_{\text{sim}}^{(i)}, Y_{\text{obs}}, J)$

if $d_i \leq \epsilon^{(k)}$ **then**

$\mathcal{S}^{(k)} \leftarrow \mathcal{S}^{(k)} \cup \{\theta^{(i)}\}$

if $|\mathcal{S}^{(k)}| = N_{\text{desired}}$ **then**

break

end if

end if

end for

end for

ECDF, denoted by $\hat{F}_n(t)$, is defined as the fraction of data points in the sample that are less than or equal to a value t. Formally, it is given by:

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[x_i \leq t],$$

where $\mathbb{I}[\cdot]$ is the indicator function. Given two datasets, the observed data $D_{\text{obs}} = \{y_1, y_2, \ldots, y_n\}$ and the simulated data $D_{\text{sim}} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m\}$, we can compute their respective ECDFs, $\hat{F}_{D_{\text{obs}}}(t)$ and $\hat{F}_{D_{\text{sim}}}(t)$. The two sample KS statistic, $D_{n,m}$, is then defined as the maximum absolute difference between these two ECDFs over all values of t:

$$D_{n,m} = \sup_t |\hat{F}_{D_{\text{obs}}}(t) - \hat{F}_{D_{\text{sim}}}(t)| \tag{2}$$

The KS statistic offers several advantages over commonly used point wise error metrics such as the normalized RMSE. It is a nonparametric distance metric that makes no assumptions about the underlying data distributions. It is sensitive to differences in both the location (e.g., mean, median) and shape (e.g., variance, skewness) of the distributions, making it a comprehensive metric for goodness of fit.

The distance metric (2) is calculated for all outputs and then averaged to compute a scalar distance metric. Since the KS statistic is normalized in the range $[0, 1]$, averaging preserves comparability across outputs and provides a balanced way to combine the information from multiple outputs. Based on these distances, we determine the acceptance threshold as the top ζth percentile of the best fitting samples. This threshold sets the baseline for which simulations count as acceptable in the main sampling phase.

In the main phase, we continue drawing samples from the prior, simulate the outputs, and compute the KS distance to the observed data. We retain only those parameter sets whose distances fall below the threshold, up to a target number of accepted samples. The resulting accepted samples define the posterior distribution for that dataset.

For the first dataset, we define a uniform prior based on physics informed bounds derived from expert knowledge. For subsequent datasets, we use the posterior distribution from the previous calibration step as the prior. To ensure a robust prior approximation, especially when the posterior sample set is sparse, we fit a Gaussian mixture model (GMM) [18]. The GMM approximates the posterior by fitting components that capture its overall structure. Component means are not tied to individual samples, allowing flexible support and representation of multimodal distributions for downstream calibration. Additionally, to account for potential discrepancies between model predictions and observations, we allow the output noise variances associated with our 9 outputs to be learnable. Although unnecessary for simulation datasets with known ground truth, this mechanism prepares our methodology for future application to experimental data calibration, where model form error and measurement noise are expected to be present.

In practice, effective sequential updating typically requires gradual parameter variation across datasets, allowing the posterior from one stage to serve as an informative prior for the next. If datasets are incorrectly ordered or separated by large gaps, the resulting prior may be less consistent with the subsequent dataset, potentially slowing convergence or producing broader posterior distributions. In addition, the efficiency of ABC rejection generally decreases as the dimensionality of the parameter space increases. For higher-dimensional parameter spaces, more scalable variants such as sequential Monte Carlo ABC (ABC-SMC) can be employed to improve sampling efficiency by progressively adapting the particle population and tolerance thresholds.

TABLE I
HYPERPARAMETERS USED FOR ABC REJECTION SAMPLING

Distance metric	Kolmogorov–Smirnov Statistic
Pilot sample size (N_{pilot})	5000
Desired accepted samples (N_{desired})	4000
Main sample size (N_{main})	80000
Quantile ζ	0.05 (5 %)

IV. RESULTS

Table I summarizes the hyperparameters chosen for ABC rejection sampling. All computations were performed on a desktop equipped with an Intel® Core™ i9-10900K CPU operating at 3.70 GHz. For each dataset, the ABC rejection algorithm requires at most $N_{\text{pilot}} + N_{\text{main}}$ forward simulations. A single forward simulation of the VCC model requires approximately 60 s. To ensure computational tractability, multiple simulations were executed in parallel.

A. Estimation of parameters

To validate the sequential calibration approach, we generated two reference datasets using parameter values sampled from prior uniform distributions based on domain knowledge. The parameter sets were intentionally chosen to differ only slightly, reflecting either system-to-system variability or gradual parameter drift over the equipment lifetime. These datasets were generated by a sequence of step response of the VCC model to the following external inputs: linear expansion valve (LEV) position, compressor fan speed (CF), ambient temperature (T_{ambient}), and indoor temperature (T_{indoor}) over a period of 2120 seconds as shown in Fig. 4. The first 1060 seconds of simulation data were used for calibration to infer posterior parameter distributions, while the posterior predictive distribution (PPD) was evaluated on the full 2120 seconds dataset to assess both fit quality and generalization to previously unseen model excitations.

Fig. 5 shows the kernel density estimations (KDEs) of the prior and posterior samples for a representative subset of parameters after calibration using the first dataset. Some parameters (i.e., compressor c.f.: volumetric efficiency, compressor c.f.: power, compressor c.f.: power transferred to refrigerant, LEV c.f.: flow coefficient and total refrigerant mass) which contribute more to the model output variability showed significant posterior contraction toward their true values. Similarly, parameters with relatively smaller effect on model outputs exhibited minimal updates to their distribu-

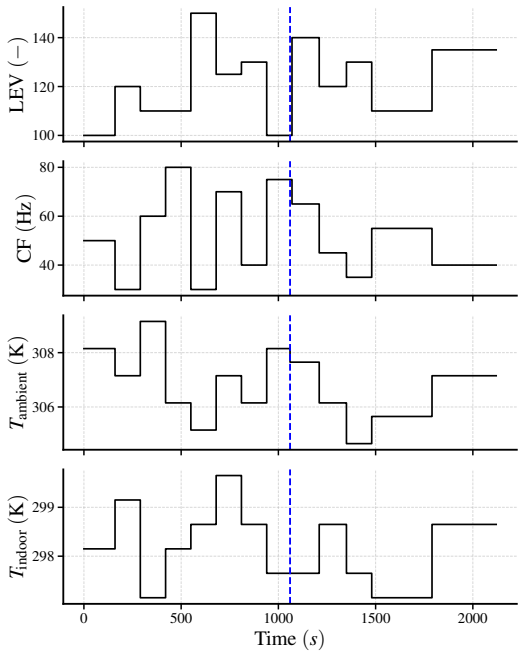

Fig. 4. Model inputs used for exciting the VCC model

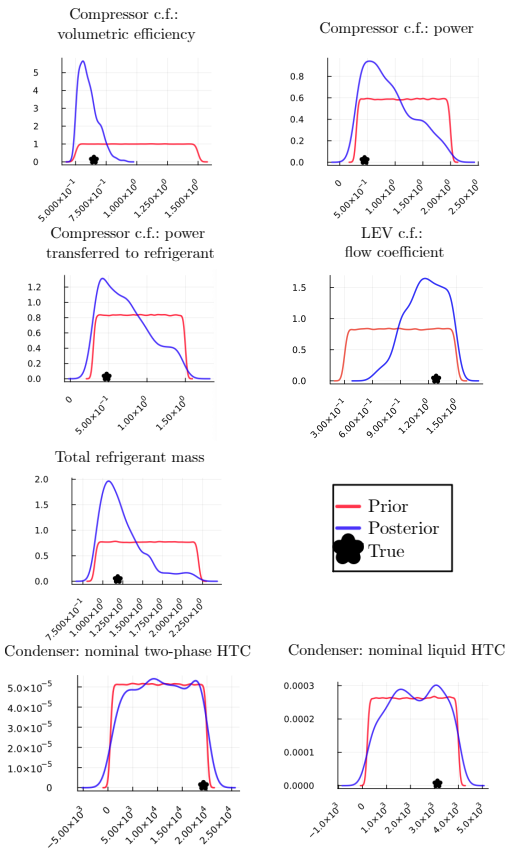

Fig. 5. KDE plots of posterior and prior samples of a few representative calibration parameters for first dataset.

tions. These observations follow from Fig. 5 when considered together with the sensitivity analysis shown in Figs 9 and 10.

Fig. 6 presents PPDs of outputs for the first dataset. We can see that the predictive medians track the ground truth well across most outputs. The discharge temperature (TD) shows notable deviation likely because of the large parametric uncertainties and relatively higher sensitivity of TD to the parameters. We can see that the parametric uncertainty dominates due to use of noise free simulation data to calibrate the parameters. We report predictive medians rather than means in this paper to minimize the influence of potential skewness and outliers. We can also see that the predictions generalize well to unseen model inputs, although slight deviations remain due to the threshold chosen at the 5th percentile of pilot phase distances. This threshold trades off calibration fidelity against posterior sample diversity. In the limit $\epsilon \to 0$, the predictive median would converge to the ground truth. However, this would result in prohibitively few accepted samples.

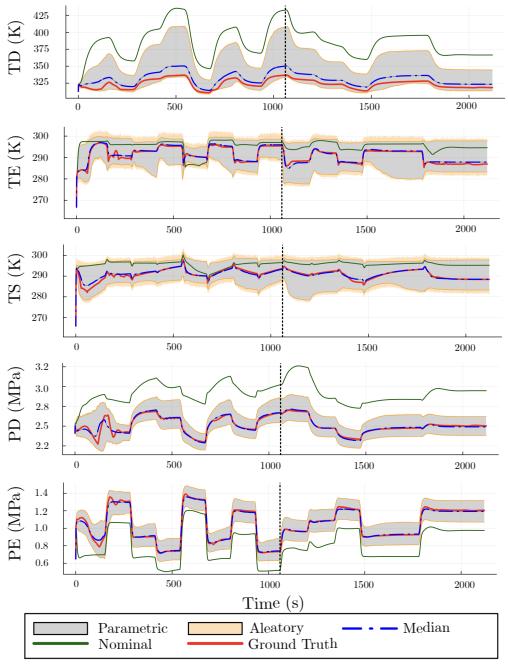

Fig. 6. Posterior predictive distribution of a few outputs for first dataset.

Fig. 7 shows KDE plots of the prior and posterior samples of calibration parameters for the second dataset, demonstrating effective updating and alignment with the true parameter changes. The corresponding PPDs (Fig. 8) indicate improved agreement between calibrated predictive median and ground truth output of second dataset compared to an uncalibrated scenario whose output was generated by using ground truth calibration parameters from the previous dataset (i.e the first dataset). In addition, the posterior predictive intervals in Fig. 8 are noticeably narrower than those for the first dataset, reflecting reduced parametric uncertainty achieved by incorporating prior information from the earlier calibration. Note that parameter combinations that caused

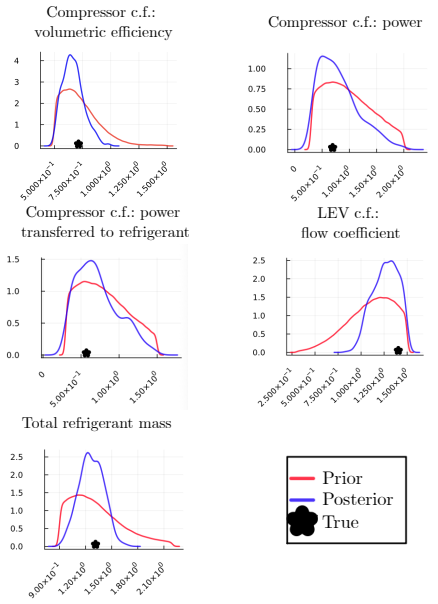

Fig. 7. KDE plots of posterior and prior samples of a few well calibrated parameters for second dataset.

simulation failures were excluded from posterior sampling to ensure reliability.

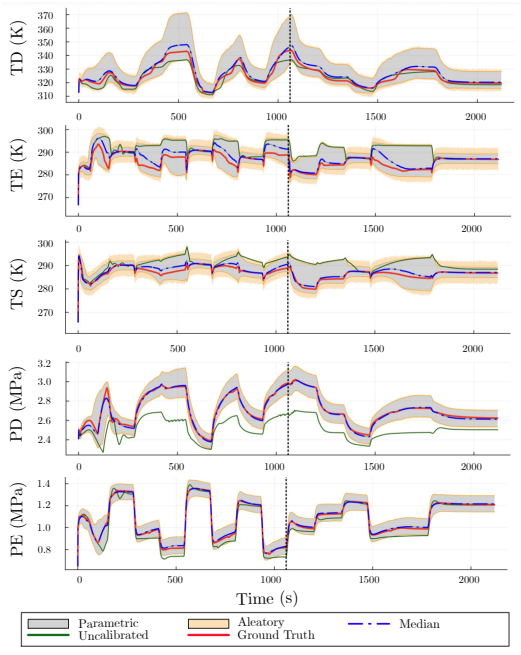

Fig. 8. Posterior predictive distribution of a few outputs for second dataset.

B. Case study: Identification of refrigerant mass leakage

Refrigerant mass leakage is a common fault in aging VCC systems, with severe consequences for both environmental impact and system efficiency and performance. To evaluate the diagnostic capability of our framework, we conducted a case study focused on identifying the total refrigerant charge under progressive leakage conditions. For the purpose of this study, we created three simulated datasets by varying only the refrigerant mass (M_0) to emulate leakage conditions while other parameters were fixed. During calibration, we assumed the refrigerant mass to remain constant within each dataset, a reasonable approximation since leakage typically occurs slowly over longer horizons than the data collection window. Since the problem involves only a single parameter with high sensitivity, we used a uniform prior at each calibration stage instead of sequential posterior updates to prevent overly concentrated priors from biasing the inference when leakage levels differ across datasets.

Table II compares the inferred mean and standard deviation of the posterior distributions to the reference values. Across all three leakage levels, the inferred means are within 0.02–0.03 kg of the true values, and the standard deviations remain tight. These results demonstrate that the proposed calibration framework can reliably estimate refrigerant mass, providing a valuable tool for system monitoring and predictive maintenance.

TABLE II
IDENTIFICATION OF REFRIGERANT MASS (KG)

Mean (kg)	Standard deviation (kg)	Reference (kg)
1.98	0.03	2.0
1.51	0.02	1.5
0.97	0.03	1.0

V. CONCLUSIONS

This paper presented a sequential Bayesian calibration framework for dynamic VCC models using ABC. By reducing the parameter space to a physically meaningful subset and leveraging simulation driven inference, the methodology demonstrated efficient and accurate Bayesian calibration of parameters. We further investigated refrigerant mass leakage as a representative case study. By isolating refrigerant mass as the primary calibration parameter, the proposed framework accurately tracked progressive leakage scenarios, yielding posterior distributions tightly centered on the true values. This result highlights the potential of the framework for fault detection and diagnostics in aging VCC systems. Future work will focus on calibrating against experimental datasets, where sensor noise and model form discrepancies must be addressed. The ability to jointly learn parameters and output noise variances positions the framework well for such settings.

APPENDIX

We performed Sobol sensitivity analysis to quantify the influence of the calibration parameters on the VCC model outputs. Figures 9 and 10 shows the resulting first order and total order Sobol indices as heatmaps. These figures illustrate the sensitivity of model outputs with respect to the 26 calibration parameters, which are ordered as listed in Table III.

The analysis was carried out using the `GlobalSensitivity.jl` [19] package, which is compatible with VCC models implemented in Julia. We excited the model using the first 1060 seconds of the

TABLE III

SET OF CALIBRATION PARAMETERS

ID	Name	Unit	Range
1	Total refrigerant mass	kg	[0.9, 2.2]
2	Compressor c.f.: volumetric efficiency	-	[0.5, 1.5]
3	Compressor c.f.: power	-	[0.3, 2.0]
4	Compressor c.f.: power transferred to refrigerant	-	[0.3, 1.5]
5	Compressor c.f.: power dissipated through shell	-	[0.3, 1.5]
6	LEV c.f.: flow coefficient	-	[0.3, 1.5]
7	Condenser c.f.: pressure drop	-	[0.59, 2.36]
8	Evaporator c.f.: pressure drop	-	[0.59, 2.36]
9	Condenser: nominal vapor HTC	W/(m^{2}K)	[200, 2000]
10	Condenser: nominal two-phase HTC	W/(m^{2}K)	[500, 20000]
11	Condenser: nominal liquid HTC	W/(m^{2}K)	[200, 4000]
12	Evaporator: nominal vapor HTC	W/(m^{2}K)	[200, 2000]
13	Evaporator: nominal two-phase HTC	W/(m^{2}K)	[500, 20000]
14	Evaporator: nominal liquid HTC	W/(m^{2}K)	[200, 4000]
15	Outdoor fan: nominal volumetric flow rate	m^3/s	[0.5, 1.5]
16	Outdoor fan: nominal power	W	[50, 500]
17	Outdoor fan: nominal speed	Hz	[500, 5000]
18	Indoor fan: nominal volumetric flow rate	m^3/s	[0.1, 0.8]
19	Indoor fan: nominal power	W	[50, 500]
20	Indoor fan: nominal speed	Hz	[50, 500]
21	Condenser: air side HTC	W/(m^{2}K)	[10, 150]
22	Evaporator: air side HTC	W/(m^{2}K)	[10, 150]
23	Condenser coil length	m	[18, 19.8]
24	Evaporator coil length	m	[7.3, 8]
25	Vapor pipe length	m	[14, 30]
26	Liquid pipe length	m	[14, 30]

Abbreviations: correction factor (c.f.), heat transfer coefficient (HTC)

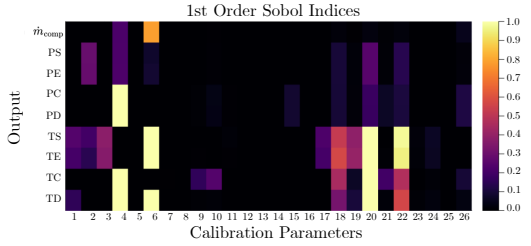

Fig. 9. Heatmap of first order sobol indices

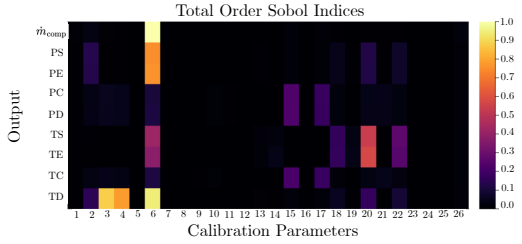

Fig. 10. Heatmap of total order sobol indices

control inputs shown in Fig. 4, as these inputs correspond to the dataset used for parameter estimation. To mitigate the effects of initial sharp transients while the model stabilizes its refrigerant charge, we discarded the first 20 seconds of the simulation. We defined the model response for the analysis as the time average of the output values from 20 s to 1060 s. We used 10,000 Sobol samples in our analysis, with domain-informed bounds for calibration parameters as listed in Table III. Due to the model's complexity, the DAE solver occasionally failed to converge for certain parameter combinations; to ensure reliable sensitivity indices, these nonconvergent cases were systematically excluded prior to analysis. Figures 9 and 10 show that only a subset of parameters substantially affect the model outputs. Such behavior is typical in complex models, where certain parameters may become non-identifiable under the selected parameter bounds and input excitations.

REFERENCES

[1] S.A. Bortoff, P. Schwerdtner, C. Danielson, and S. Di Cairano. H-infinity loop-shaped model predictive control with heat pump application. In *18th European Control Conference*, pages 2386–2393, 2019.

[2] Vedang M. Deshpande, Christopher R. Laughman, Yingbo Ma, and Chris Rackauckas. Constrained smoothers for state estimation of vapor compression cycles. In *2022 American Control Conference (ACC)*, pages 2333–2340, 2022.

[3] Ankush Chakrabarty, Gordon Wichern, Vedang M. Deshpande, Abraham P. Vinod, Karl Berntorp, and Christopher R. Laughman. Meta-learning for physically-constrained neural system identification. *Neurocomputing*, 651:130945, 2025.

[4] Vedang M. Deshpande, Ankush Chakrabarty, Abraham P. Vinod, and Christopher R. Laughman. Physics-constrained deep autoencoded kalman filters for estimating vapor compression system states. *IEEE Control Systems Letters*, 7:3483–3488, 2023.

[5] Christian Vering, Sebastian Borges, Daniel Coakley, Hannah Kruetzfeldt, Philipp Mehrfeld, and Dirk Müller. Digital twin design with on-line calibration for HVAC systems in buildings. In *Proceedings of Building Simulation 2021: 17th Conference of IBPSA*, volume 17 of *Building Simulation*, pages 2938–2945, Bruges, Belgium, September 2021. IBPSA.

[6] Jiacheng Ma, Donghun Kim, and James E Braun. Reduced-dimension bayesian optimization for model calibration of transient vapor compression cycles. *International Journal of Refrigeration*, 168:246–258, 2024.

[7] Ankush Chakrabarty, Emilio Maddalena, Hongtao Qiao, and Christopher Laughman. Scalable bayesian optimization for model calibration: Case study on coupled building and hvac dynamics. *Energy and Buildings*, 253:111460, 2021.

[8] Victor Martinez-Viol, Eva M. Urbano, Miguel Delgado-Prieto, and Luis Romeral. Automatic model calibration for coupled hvac and building dynamics using modelica and bayesian optimization. *Building and Environment*, 226:109693, 2022.

[9] C. Laughman and H. Qiao. Patch-based thermodynamic property models for the subcritical region. In *International Refrigeration and Air-Conditioning Conference at Purdue*, pages 1–10, 2021. Paper 2258.

[10] H. Qiao, V. Aute, and R. Radermacher. Transient modeling of a flash tank vapor injection heat pump system–Part I: Model development. *International Journal of Refrigeration*, 49:169–182, 2015.

[11] Hongtao Qiao, Christopher R Laughman, D Burns, and S Bortoff. Dynamic characteristics of an R410A multi-split variable refrigerant flow air-conditioning system. In *12th IEA Heat Pump Conference*, 2017.

[12] Christopher R. Laughman and Hongtao Qiao. On the influence of state selection on mass conservation in dynamic vapor compression cycle models. *Mathematical and Computer Modeling of Dynamical Systems*, 23(1):262–283, Dec. 2016.

[13] Scott A. Bortoff, Vedang M. Deshpande, Christopher R. Laughman, and Hongtao Qiao. A dynamic analysis of refrigerant mass in vapor compression cycles. In *International Modelica Conference*, 2025.

[14] Yingbo Ma, Shashi Gowda, Ranjan Anantharaman, Chris Laughman, Viral Shah, and Chris Rackauckas. ModelingToolkit: A composable graph transformation system for equation-based modeling, 2021.

[15] Jonathan K Pritchard, Mark T Seielstad, Anna Perez-Lezaun, and Marcus W Feldman. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798, 1999.

[16] Stuart Barber, Jochen Voss, and Mark Webster. The rate of convergence for approximate bayesian computation. 2015.

[17] Simon Tavaré, David J Balding, Robert C Griffiths, and Peter Donnelly. Inferring coalescence times from dna sequence data. *Genetics*, 145(2):505–518, 1997.

[18] Mike West. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):409–422, 1993.

[19] Vaibhav Kumar Dixit and Christopher Rackauckas. Globalsensitivity.jl: Performant and parallel global sensitivity analysis with julia. *Journal of Open Source Software*, 7(76):4561, 2022.