

A Validated Machine Learning Approach to Efficient Thermal Energy Storage Simulation Using Synthetic Data

Falko Schneider¹ , Jan-Niklas Schagen² , Constantin Peters¹, Bodo Kraft² ,
Cristiano José Teixeira Boura¹ , and Ulf Herrmann¹ 

¹Solar-Institut Jülich of FH Aachen University of Applied Sciences, Germany

²Institute for Data-Driven Technologies of FH Aachen University of Applied Sciences, Germany

*Correspondence: Falko Schneider, f.schneider@sj.fh-aachen.de

Abstract. This work presents a validated machine learning approach for developing a surrogate model to simulate thermal energy storage (TES) systems. High-fidelity, physics-based Modelica simulations provide accurate predictions but are computationally expensive and unsuitable for real-time applications like operational assistance systems. To overcome this limitation, we train a Long Short-Term Memory (LSTM)-based surrogate model on synthetic data generated across a wide range of operating conditions, enabling rapid prediction of internal storage temperature profiles. The model is validated using operational data from Synhelion's solar fuel pilot plant in Jülich, Germany. Results demonstrate that the surrogate achieves high accuracy on synthetic test scenarios and promising, though heterogeneous, performance on real plant data. These findings highlight the potential of the surrogate model for real-time applications, including model predictive control and operational support.

Keywords: thermal energy storage, surrogate modelling, CSP optimization

1. Introduction

Dynamic simulations are a powerful tool for investigating Concentrated Solar Power (CSP) systems and for developing optimized control strategies. To this end, a dynamic process model of a solar fuel production plant has been developed in Modelica, where solar heat is used to reform biogas and steam into synthesis gas (syngas), which is subsequently converted into synthetic crude oil [1].

However, high-fidelity physical simulations are computationally intensive and often lack real-time capability, limiting their use in control and optimization applications. To overcome this challenge, we propose a machine learning (ML) based surrogate model for efficient thermal energy storage (TES) simulation. By training the ML model on synthetically generated data from the validated process model, system behavior can be captured without relying on computationally expensive physical equations. This approach enables fast and accurate TES modeling suitable for real-time applications in CSP control. In this work, we generate a diverse set of synthetic scenarios using the Modelica TES model to create a comprehensive training dataset. A surrogate model is then trained and evaluated in terms of predictive accuracy and computational efficiency relative to the original simulation. Finally, we demonstrate the applicability of the trained ML model by applying it to real operational data from the DAWN plant for solar fuel synthesis in Jülich, Germany.

The following section reviews related work on TES modeling, machine learning applications in energy systems, and surrogate modeling approaches, providing context for the present study.

2. Related Work

2.1 Thermal Energy Storage Modeling in CSP

Thermal energy storage (TES) systems are a key component of Concentrated Solar Power (CSP) plants as they allow the decoupling of solar energy collection from electricity or fuel production. Therefore, accurate modeling of TES dynamics is essential in order to optimize plant operation strategies.

2.1.1 Physical Modeling Approaches

High-fidelity physical models typically rely on first-principle equations such as energy balances, heat transfer correlations, and discretized representations of tanks or heat exchangers. Lumped-parameter models offer computational efficiency but may neglect spatial temperature gradients. Distributed-parameter or finite-volume models provide higher accuracy, but require significant computational effort. These models are commonly used in process design and offline optimization, but are often too slow for real-time control applications.

2.1.2 Physical TES Model

The validated process model of previous work [1] is based on the thermal storage built at plant DAWN in Jülich. It contains ceramic storage bricks with ducts for the heat transfer fluid (HTF), i.e. steam, to flow through. The model is spatially discretized into lumped isothermal volumes. For each volume, the energy balances of the HTF and storage material are calculated as well as the convective and radiative heat ex-

change between them. The heat conduction flows within the storage and out through the insulation layers into the environment, i.e. the thermal losses, are determined as well. Depending on the flow direction, the TES is either being charged or discharged. This model setup has been validated with operational data and is used to generate the synthetic training data for a data-driven surrogate model.

2.2 Surrogate Models

Following the definition of Han et al. [2], Surrogate models are computationally inexpensive approximations of high-fidelity models that are costly to evaluate. They are particularly valuable when direct optimization or repeated simulation of the high-fidelity model is impractical due to long computation times, numerical noise, or the lack of readily available gradient information. Surrogate based Optimization replaces direct optimization of the complex model with an iterative process of building, optimizing, and updating a cheaper approximation, the surrogate, thereby drastically reducing computational cost while maintaining acceptable accuracy. This predictor–corrector scheme continues until a termination criterion is met and generally requires far fewer evaluations of the expensive model than direct optimization.

2.2.1 Design of Experiments

The accuracy of a surrogate model depends strongly on the quality and representativeness of the data used for training. Design of Experiments (DoE) techniques are commonly used to strategically select sample points in the input space, referred to as factors, to ensure sufficient coverage while minimizing the number of expensive simulations. Popular approaches include full and fractional factorial designs, Latin Hypercube Sampling (LHS), Sobol sequences, Grid-Search and other space-filling strategies. DoE plays a key role in ensuring that the surrogate adequately represents the relevant design space, which is essential for both global approximation quality and successful optimization. [2]

2.2.2 Types of Surrogate Models

Surrogate models can broadly be categorized into physically based and data-driven approaches. Physically based surrogates rely on simplified physics or coarser discretization of the original system and are sometimes referred to as low-fidelity models. These models can be further refined or corrected using data from the high-fidelity model to improve accuracy. In contrast, data-driven surrogates use statistical or machine learning methods to fit functional approximations directly to sampled data. Common techniques include polynomial regression, radial basis functions, Gaussian process regression (Kriging), and neural networks. Together, these two categories offer a spectrum of approaches, from physics-informed predictors to purely data-driven predictors. [2]

3. Thermal Energy Storage Surrogate Model

In this section, we present our methodology for developing a data-driven surrogate model to simulate the temperature dynamics of a TES, based on the existing high-fidelity process model. In contrast to physics-based surrogates, which rely on manual declarations and simplifications, data-driven machine learning methods operate as black boxes that can directly learn from data without the need for manually declared

rules. Therefore, a machine learning approach was chosen to investigate the potential of black-box modeling techniques.

The approach follows a two-step procedure: first, our design of experiment, the generation of a large, diverse, and physically consistent synthetic dataset; second, the training of a ML based surrogate model to accurately and efficiently approximate the TES temperature evolution under varying operating conditions.

3.1 Synthetic Dataset Creation

Synthetic training data is generated by leveraging the predictive capabilities of the existing Modelica process model. Given the profiles of the HTF mass flow $\dot{m}(t)$ and the HTF inlet temperature $T_{HTF}(t)$, the process model simulates the temperature response $T_{TES}(t)$ measured at 14 sensors on the vertical center line of DAWN's TES over a period of 20 hours. These temperature profiles serve as labels for training a supervised ML model.

$$T_{TES}(t) = \begin{pmatrix} T_{Sensor}^1(t) \\ \vdots \\ T_{Sensor}^{14}(t) \end{pmatrix} \in \mathbb{R}^{14}, \quad t \in [0, 1200] \quad (1)$$

While label generation $T_{TES}(t)$ is fully determined by the underlying physical logic of the process model, $\dot{m}(t)$ and $T_{HTF}(t)$ as well as the initial condition $T_{TES}(0)$ must be declared explicitly as simulation input.

Table 1. Design parameters for the generation of synthetic scenarios. The table lists input functions, their associated parameters, explanations, ranges, and the discretization applied in the grid search for the design of experiments.

Function	Parameter	Explanation	Range	Grid Step Size
$\dot{m}(t)$	m_{peak}	peak HTF mass flow during charging and discharging cycles	0.05–0.25 kg/s	0.02
$\dot{m}(t)$	t_{cycle}	period of charging-discharging cycle	500–480 000 s	59937.5
$T_{HTF}(t)$	$T_{HTF}^{charging}$	HTF inlet temperature during charging	293.15–1473.15 K	73.75
$T_{HTF}(t)$	$T_{HTF}^{discharging}$	HTF inlet temperature during discharging	293.15–1073.15 K	195
$T_{TES}(t)$	$T_{TES}^{initial}$	homogeneous TES temperature at $t = 0$	293.15–973.15 K	136

A synthetic scenario is uniquely defined by a set of five parameters, summarized in Table 1. The HTF mass flow follows a periodic charging–discharging cycle, reaching a peak mass flow m_{peak} in each phase, with the first phase starting as a charging cycle at $t = 0$.

$$\dot{m}(t) = \dot{m}_{peak} \cos\left(\frac{2\pi}{t_{cycle}}t\right) \quad (2)$$

$$T_{HTF}(t) = \begin{cases} T_{HTF}^{charging} & \text{if } \dot{m}(t) > 0 \\ T_{HTF}^{discharging} & \text{if } \dot{m}(t) < 0 \end{cases} \quad (3)$$

To model $\dot{m}(t)$, we employ a cosinusoidal function of period t_{cycle} (see eq. 2). A negative sign of $\dot{m}(t)$ encodes discharging mode, while a positive sign encodes charging mode. The inlet HTF temperature $T_{HTF}(t)$ is defined by two constant values (see eq. 3), for charging and discharging modes respectively. Additionally, the initial TES temperature $T_{TES}(0)$ is defined as a homogeneous value.

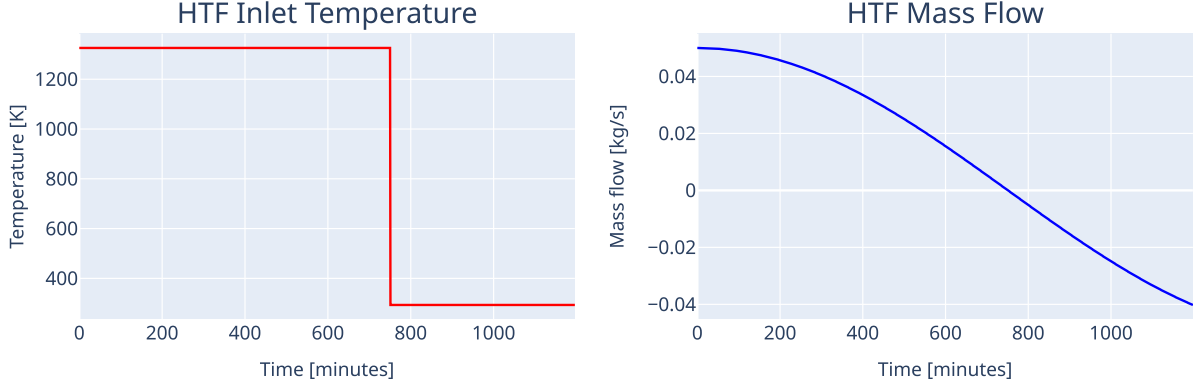


Figure 1. Inputs of an example scenario generated according to our design of experiments (DoE); $m_{peak} = 0.05$, $t_{cycle} = 180312$, $T_{HTF}^{charging} = 1325$, $T_{HTF}^{discharging} = 293$, and $T_{TES}^{initial} = 429$.

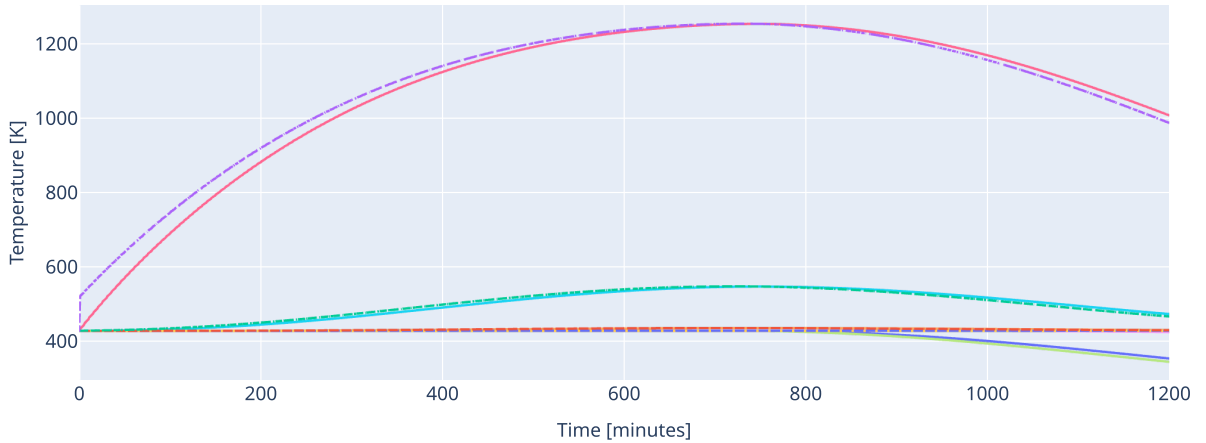


Figure 2. Temperature evolution of the example scenario simulated with the process model.

These inputs based on rules, combined with a grid search of the rules parameters (see Table1), form the basis of the design of experiments used to generate a diverse set of realistic scenarios to train the surrogate model. In total, 50,000 unique scenarios were created, each covering a 20-hour time span. Applying these scenarios to the process model yielded the corresponding synthetic temperature data, which serve as dataset for the development of the data-driven ML surrogate model.

3.2 Surrogate Machine Learning Model

In contrast to the physics-based process model, the ML surrogate model computes predictions by applying learned parameters to the input sequences, enabling rapid inference while approximating the system's complex thermal dynamics based on training data. In the following, we describe the design and architecture of the surrogate model. The final architecture is depicted in fig. 3.

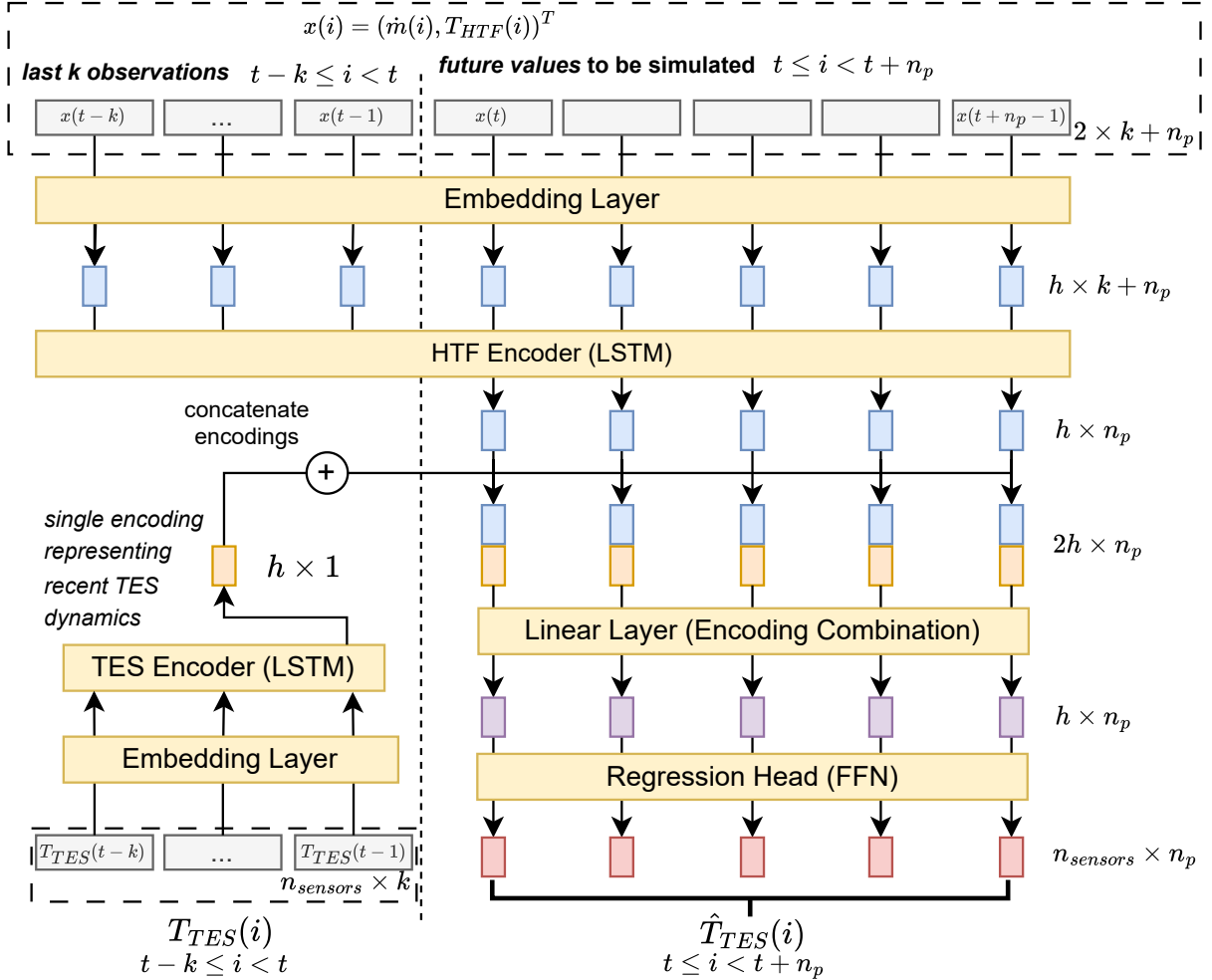


Figure 3. Schematic overview of the surrogate model architecture. The model receives two distinct input sequences: (i) the HTF sequence, including past, current, and planned future mass flow and inlet temperature values, and (ii) the TES sequence, containing historical TES temperatures. These sequences are processed separately and combined within the model to predict the TES temperature profile at the current and future timesteps.

3.2.1 Objective and Input Sequences

Accurate predictions $\hat{T}_{TES}(t)$ of the TES temperature profile $T_{TES}(t)$ require more information than the instantaneous HTF mass flow $\dot{m}(t)$ and inlet temperature $T_{HTF}(t)$ alone. During charging and discharging, the vertical thermocline moves continuously through the storage, while temperatures are measured only at discrete sensor locations. The position of the thermocline at a given timestep can be inferred from the previous TES temperature gradients. Consequently, the historical TES temperatures are essential input features to capture the system's state, while the historical and planned

future HTF inputs provide the boundary conditions of the process. Formally, the surrogate model receives two distinct input sequences at timestep t :

- **HTF sequence:** the mass flow and inlet temperature of the heat transfer fluid, including the past k timesteps, the current timestep, and $n_p - 1$ future timesteps,

$$\mathbf{X}_{HTF}(t) = [(\dot{m}(t - \tau), T_{HTF}(t - \tau))]_{\tau=k}^1 \cup [(\dot{m}(t), T_{HTF}(t))] \cup [(\dot{m}(t + \tau'), T_{HTF}(t + \tau'))]_{\tau'=1}^{n_p-1} \quad (4)$$

where k denotes the number of past timesteps and n_p the number of prediction steps. This sequence provides the model with information about the driving input dynamics that affect the thermal response of the storage system.

- **TES sequence:** the historical TES temperatures from the past k timesteps,

$$\mathbf{X}_{TES}(t) = [T_{TES}(t - \tau)]_{\tau=1}^k. \quad (5)$$

Providing these past $t - k..t - 1$ temperatures allows the model to account for the system's current thermal state, which is critical for accurately predicting future dynamics.

Selecting an appropriate k is crucial, as too short a context can omit important recent dynamics, while unnecessarily large contexts can increase computational overhead and irrelevant information negatively impacting efficiency and accuracy. We intuitively select a historical window of one hour, $k = 60$, without additional hyperparameter optimization or evaluation.

Based on the input sequences, the surrogate model $\phi_{\text{surrogate}}$ then creates temperature predictions for n_p timesteps (current timestep t and following):

$$\phi_{\text{surrogate}}(\mathbf{X}_{HTF}(t), \mathbf{X}_{TES}(t)) = [\hat{T}_{TES}(t')]_{t'=t}^{t+n_p-1} \in \mathbb{R}^{n_p \times n_s}, \quad (6)$$

whereas n_s is the number of sensor locations, each row of the prediction corresponds to a timestep and each column to a sensor location.

3.2.2 Input Embedding and Encoding

Each timestep in the sequences of $\mathbf{X}_{HTF}(t)$ and $\mathbf{X}_{TES}(t)$ is first embedded independently into a latent feature space of size h using a linear layer with ReLU activation, yielding a richer representation that facilitates subsequent sequence encoding. The embedded sequences are then processed by two Long-Short-Term-Memory (LSTM) based encoders. LSTM neural networks have been proven effective in modelling sequential data, as they can capture both long-term and short-term dependencies in input sequences like $\mathbf{X}_{HTF}(t)$ and $\mathbf{X}_{TES}(t)$ [3].

The TES encoder outputs only its final hidden state, providing a compact summary of the past TES temperature trajectory. In contrast, the HTF encoder provides a context vector for every timestep to be predicted. Because the LSTM processes inputs in chronological order, each encoding can incorporate information from earlier timesteps but not from future ones, thus preserving temporal order which aligns with the causal nature of the prediction task.

For each timestep to be predicted, the corresponding HTF context vector is concatenated with the global TES encoding to form a combined representation of size $2h$. This combined vector is then passed through a linear layer with ReLU activation, which not only reduces the dimensionality back to h but is intended to encourage the model

to learn a joint representation that combines information from both input sources. In our experiments, we set $h = 128$ based on heuristic considerations, as this dimensionality offered a reasonable compromise between model complexity and runtime in preliminary tests.

3.2.3 Prediction

In the final step, predictions of future TES temperatures are generated for each timestep by processing the corresponding combined encoding of the HTF and TES representations. Each combined vector is fed independently into a regression head, implemented as a feed-forward neural network, enabling parallel computation across all timesteps to be predicted and thus improving inference efficiency. The regression head maps the latent representation to an output vector whose dimensionality corresponds to the number of TES sensor locations n_s , producing a sequence of predicted temperatures that preserves the temporal alignment with the input sequence $\mathbf{X}_{HTF}(t)$. This architecture ensures that the surrogate model can simultaneously capture the complex nonlinear dependencies between the system inputs and the evolving thermal state while maintaining computational efficiency suitable for optimization applications.

3.2.4 Training and Inference

For training and evaluation, the 50,000 synthetic scenarios were randomly split into training (70%) and test (30%) sets. All input features were subsequently normalized using z-score standardization. During training, the initial $k = 60$ timesteps of T_{HTF} , $\dot{m}(t)$, and $T_{TES}(t)$ of a scenario were provided as context, while the remaining timesteps of each scenario were used as target values for prediction. Therefore during training $n_p = 1200 - 60 = 1140$.

During inference, for any given timestep t in a plant optimization application, the last 60 minutes of measured data are provided as context, and the model predicts future TES temperatures based on planned or simulated HTF inputs for up to 1140 timesteps.

4. Evaluation

In this chapter, the surrogate model introduced in Section 3.2 is evaluated. The evaluation considers both the accuracy of the predicted TES temperatures compared to the results of physics-based simulations and the computational efficiency in terms of inference speed. These analyses provide a quantitative basis for assessing the surrogate model as a fast and reliable alternative for plant optimization tasks.

4.1 Prediction Accuracy

The accuracy of the surrogate model is assessed by comparing its predicted TES temperatures with the results of the physics-based simulations. The root mean square error (RMSE) is employed to quantify the deviation between predicted and reference temperatures for each sensor across all test scenarios. This evaluation provides a clear measure of the model's ability to capture the temporal dynamics of the storage system under the varying operational conditions represented in the dataset. Consistent with the training procedure, a context size of $k = 60$ is used, providing the model with the initial 60 datapoints as input context. Consequently, the prediction horizon for each scenario is $n_p = 1200 - 60 = 1140$ timesteps.

Figure 4 summarizes the RMSE distribution across all sensors and test scenarios. The median RMSE values remain below 3 K for all sensors, indicating consistently accurate predictions across different positions within the TES. Even the maximum errors observed in the test cases do not exceed 11 K, demonstrating that the surrogate model reliably captures the system dynamics, with only minor deviations in extreme cases.

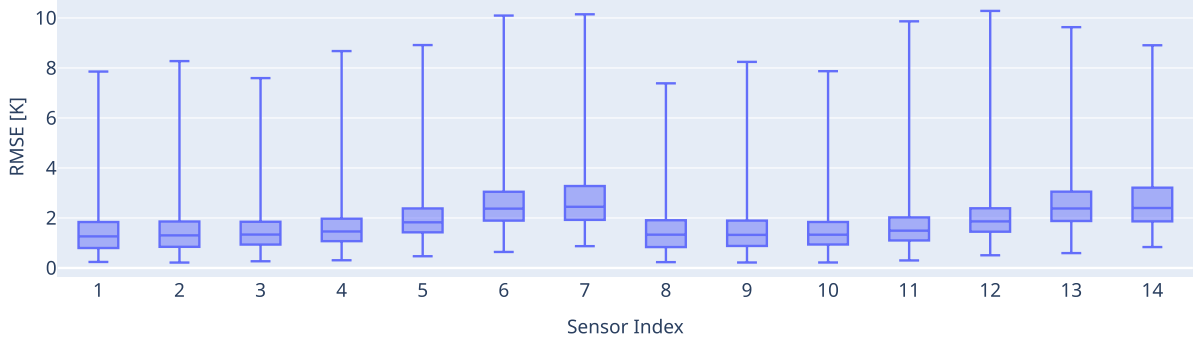


Figure 4. RMSE distribution across all sensors and test scenarios. Median values remain below 3 K, while the maximum error does not exceed 11 K.

Additionally, we conducted an *external validation* of the surrogate model using operational data measured during a Burn-In Test of the DAWN plant (see Fig. 5) to complement the validation performed on synthetic scenarios. This additional evaluation aims to assess the model's generalization capability under real operating conditions.

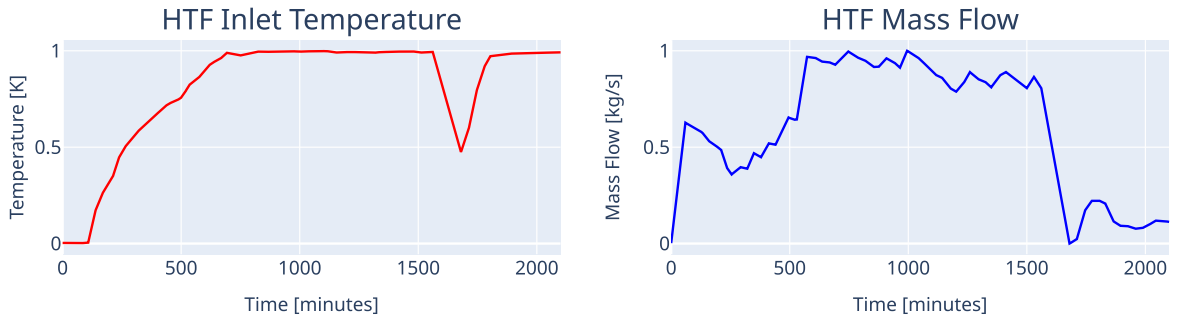


Figure 5. Measured HTF inlet Temperature and HTF Mass Flow Rate during a Burn-In Test. Y-Axis has been non-dimensionalized.

Table 2 compares the RMSE of the surrogate model predictions against those of the physics-based simulation. Although the surrogate model achieves sufficient accuracy on synthetic test scenarios, this additional validation reveals heterogeneous accuracies on operational data. For several sensors (for example indices 1, 3, and 5) the surrogate even outperforms the physics-based model, whereas for sensors 6 and 8–14 the prediction errors are substantially higher. This discrepancy is most likely caused by data drift. A comparison of T_{HTF} and $\dot{m}$ profiles from the synthetic training data (see Fig. 1) and the operational dataset (see Fig. 5) indicates that they originate from different probability distributions. The synthetic scenarios do not cover a build-up of T_{HTF} , whereas the measured $\dot{m}$ signals contain characteristic noise patterns and short-term fluctuations that are absent from the synthetic scenarios.

As a result, during inference the surrogate encounters previously unseen input patterns, which likely degrades prediction accuracy. These results emphasize the importance of including realistic characteristics of real measured operational data in future

dataset generation in order to improve model robustness under real-world conditions as part of future work (see Section 5).

Table 2. Comparison of root mean squared errors (RMSE) between the surrogate model and the physics-based simulation across individual TES sensor locations. While the surrogate achieves lower errors than the physics-based model for some sensors (e.g., indices 1, 3, and 5), it performs considerably worse for others (e.g., indices 6 and 8–14), highlighting heterogeneous generalization performance on operational data.

Sensor Index	Surrogate Model RMSE	Process Model RMSE
1	2.32	8.54
2	9.2	3.33
3	4.1	16.08
4	11.55	6.45
5	7.43	13.49
6	20.17	4.03
7	20.19	18.11
8	35.78	3.71
9	47.93	12.18
10	61.3	27.52
11	50.11	11.83
12	47.93	12.18
13	46.7	24.98
14	49.6	37.74

4.2 Computational Efficiency

In addition to predictive accuracy, the practical applicability of the surrogate model relies on its computational efficiency. Table 3 summarizes the distribution of inference times for both the surrogate and the physics-based simulations across all test scenarios.

Table 3. Comparison of inference times between the surrogate and physics-based process model across all test scenarios. The surrogate achieves consistently low and stable runtimes, whereas the physics-based simulations exhibit significantly higher variability and occasional extreme outliers.

KPI	Surrogate Model	Process Model
Mean	0.01779	5.15763
Standard Deviation	0.00167	10.57368
Minimum	0.01573	0.12891
25% Quantile	0.01719	1.82712
Median (50%)	0.01756	2.03693
75% Quantile	0.01803	2.65189
Maximum	0.06812	145.05

The surrogate model demonstrates consistently low computation times, with a narrow range of values (Mean = 0.01779 s, Maximum = 0.06812 s, Standard Deviation = 0.00167 s), while the physics-based simulations show a wide distribution of runtimes and occasional extreme outliers (Mean = 5.15 s, Maximum = 145.05 s, Standard Deviation = 10.57 s). This indicates that the surrogate not only provides fast evaluations but also ensures predictable and stable efficiency.

Quantitatively, the surrogate achieves an average speed-up of a two order magnitude compared to the physics-based model on the synthetic data. When applying to the measured data of the Burn-In Test (see Fig. 5), the physical model is even more

inefficient due to non-constant input data and therefore reduced discretization of the Solver. In fact, the physical model takes around 67 seconds in contrast to the predictable 0.014 seconds of the surrogate model, resulting in a speed up of three orders.

Such a predictable and stable reduction in computation time as well as the possibility of batched computation enable the simulation of thousands of scenarios within milliseconds to seconds, making the surrogate particularly suitable for optimization loops and real-time predictive control applications.

5. Conclusion & Outlook

In this work, we presented a two-stage methodology for developing a surrogate model capable of accurately and efficiently representing the dynamic behavior of a thermal energy storage system. In the first stage, synthetic training data was generated using a validated physics-based process model. In the second stage, a machine learning architecture was designed and trained on synthetic data to obtain data-driven surrogate model.

The results demonstrate that the surrogate model reproduces the system dynamics with high accuracy on synthetic test data while achieving a computational speed-up of several orders of magnitude compared to the physics-based simulation. This significant gain in efficiency makes the approach well suited for time-critical applications such as model-based optimization or real-time operational support. Additional validation on measured plant data revealed heterogeneous performance, which is likely to be to be attributable to a data drift between synthetic training scenarios and real operating conditions. This highlights the importance of addressing distribution shifts to further improve robustness as part of future work.

Looking forward, a key next step will be to solve the observed data drift. This may involve extending the design of synthetic scenarios to include realistic fluctuations, measurement noise, and temperature ramp-up and ramp-down profiles. Alternatively, the existing synthetic dataset can serve as a basis for pretraining to capture the fundamental thermal dynamics, followed by fine-tuning on measured data to adapt the model to the real process. Furthermore, a continuous monitoring of model performance will be essential to detect future data or concept drifts. Seasonal variations in operating profiles as well as hardware modifications, such as component replacements, may gradually alter the system behavior. Ongoing evaluation and, where necessary, periodic retraining will ensure that the surrogate model remains reliable throughout the plant's lifecycle.

Beyond methodological refinements, the surrogate model opens promising opportunities for deployment. Its integration into model predictive control frameworks and operational assistance systems could enable rapid simulation-based optimization and provide real-time decision support.

Data availability statement

The data that support the findings of this study are not publicly available due to confidentiality agreements and restrictions imposed by the data provider.

Author contributions

- **Falko Schneider:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing.
- **Jan-Niklas Schagen:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing.
- **Constantin Peters:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing.
- **Cristiano José Teixeira Boura:** Funding Acquisition, Supervision, Writing – Review & Editing.
- **Bodo Kraft:** Funding Acquisition, Supervision, Project administration, Writing – Review & Editing.
- **Ulf Hermann:** Funding Acquisition, Supervision, Project administration, Writing – Review & Editing.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors gratefully acknowledge the financial support from Synhelion Germany GmbH and the German Federal Ministry of Research, Technology and Space (BMFTR).

Acknowledgements

This work was carried out with financial support from the German government (FKZ: 13FH064KX2, TwinSF). We also thank our colleagues at Synhelion Germany GmbH for their support and providing the operational data.

References

- [1] F. Schneider, C. Schwager, C. J. Teixeira Boura, N. Oellers, A. Villegas, and U. Herrmann, “Dynamic modeling of a thermochemical system for solar fuel production based on an open-source framework,” *SolarPACES Conference Proceedings*, vol. 2, Aug. 2024. DOI: [10.52825/solarpaces.v2i.908](https://doi.org/10.52825/solarpaces.v2i.908). [Online]. Available: <https://www.tib-op.org/ojs/index.php/solarpaces/article/view/908>.
- [2] Z.-H. Han, K.-S. Zhang, et al., “Surrogate-based optimization,” *Real-world applications of genetic algorithms*, vol. 343, pp. 343–362, 2012.
- [3] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).