Surrogate-driven Variance-based Sensitivity Analysis of Thermal Storage Tanks in Integrated Energy Systems

Seydou Sene[1,2], **Linyu Lin**[3], **Junyung Kim**[3], **Majdi I. Radaideh**[2,*]
[1]*ENSTA Paris, Polytechnic Institute of Paris, 91120 Palaiseau, France*
[2]*Department of Nuclear Engineering and Radiological Sciences, University of Michigan, Ann Arbor, MI 48109, USA*
[3]*Idaho National Laboratory, Idaho Falls, ID 83415, USA*

ABSTRACT

Sensitivity analysis and uncertainty quantification are essential steps for enhancing the accuracy of computational models by identifying and mitigating uncertainties. This study focuses on these steps for the Thermal Energy Delivery System at Idaho National Laboratory, specifically targeting the thermocline tank. Using a Modelica/Dymola simulation model, the study perturbed various design parameters and boundary conditions, including shape factor, porosity, outlet temperature, inlet mass flow rate, and system pressure, to predict and quantify uncertainty in the tank's axial temperature. A dataset of over 1,000 simulations was generated, and surrogate models were developed using the pyMAISE (Michigan Artificial Intelligence Standard Environment) library, which is an Automatic Machine Learning library for nuclear engineering applications. The optimal model, a feedforward neural network with two hidden layers, achieved an R^2 score above 0.99 and a mean absolute error below 1 Kelvin. Sensitivity analyses using Sobol indices and Fourier amplitude sensitivity testing methods on this surrogate model revealed that the inlet mass flow rate at initial timestamps and porosity significantly impacts predicted temperatures across all sensors and time steps.

Keywords: Uncertainty Quantification, Machine Learning, Thermal Energy Storage, Surrogate Modeling, Variance-Based Sensitivity Analysis

1. INTRODUCTION

The Thermal Energy Distribution System (TEDS) is a critical component of the Dynamic Energy Transport and Integration Laboratory at Idaho National Laboratory (INL) [1]. It provides a platform for testing and evaluating heat transfer components, instrumentation, distribution systems, and controls essential for the hybrid generation of electrical and non-electrical products. TEDS is designed for flexible operation, functioning either as a standalone system or part of an integrated network, making it versatile for a variety of research and testing applications. A Modelica-based TEDS model (Figure 1), developed at INL, serves as a dynamic process model for the experimental system. The model, which incorporates key system components, was used to guide experimental design and control strategies, and it was validated and tuned for accuracy using standard validation & verification methodologies and experimental data [1]. The system comprises three main components: a thermocline thermal storage tank, an electrical heater, and a glycol heat exchanger. To verify system parameters, operations, and model accuracy, experimental data is collected from TEDS, covering five operational modes: warm up, charging, standby, discharge, and cool down. For example, during charging, oil heated to 250°C is introduced into the thermocline tank, with the

*M. I. Radaideh (radaideh@umich.edu)

system transitioning to standby mode when the thermal front reaches two-thirds of the tank height. This study will focus on the thermal energy storage (TES) tank.

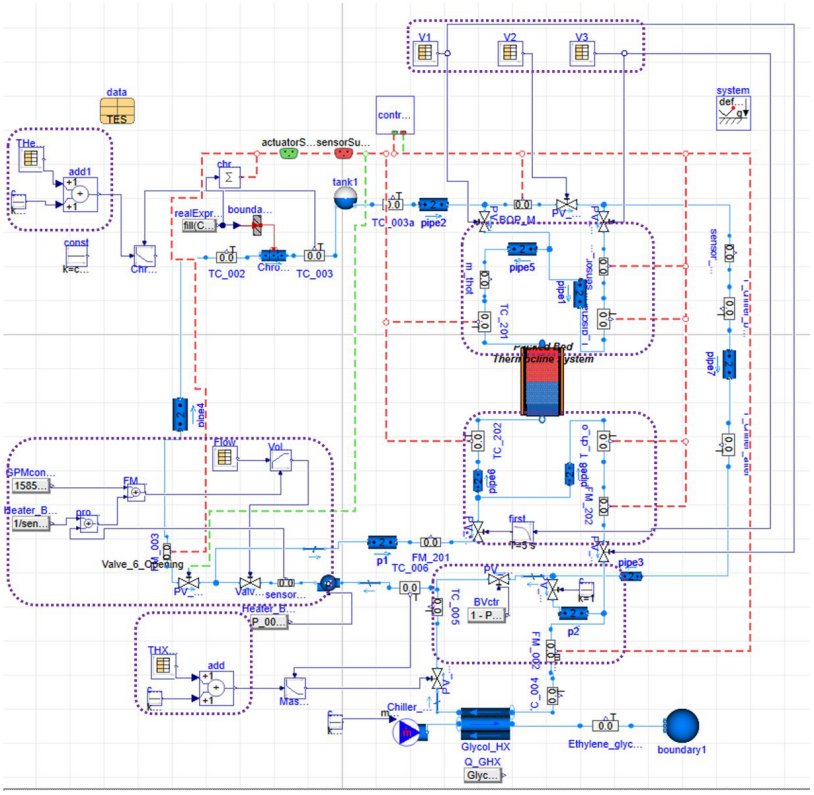


Figure 1: Schematic plot of Modelica-based TEDS model

The research conducted by [2] introduced a Modelica/Dymola model for the INL TEDS facility, demonstrating two simulation scenarios. The first scenario involved a 5 hour test simulating a facility shakedown, while the second applied a typical summer day demand, reflecting mixed commercial and residential electrical needs. More recently, a HYBRID model—a framework for transient process models for various integrated energy systems—was developed for the INL TEDS facility [3], along with a validation of the model using experimental results. Another study focused on developing digital twins to optimize the real-time economy of integrated energy systems through Modelica models [4]. Additional research explored the use of surrogate models and various sensitivity analysis methods, including Sobol indices, in nuclear reactor simulations [5,6] througm methods including Sobol indices and Shapley effect [7,8].

This study conducts a variance-based sensitivity analysis of the TES system in the TEDS facility at INL to identify the model parameters that most significantly influence simulation uncertainty. Given the computational expense of performing such analyses directly using Modelica simulations (Figure 1), a suite of machine learning (ML) models was trained to serve as simulation surrogates. The models with the highest accuracy were selected for further use. The surrogate models, combined with Sobol indices and FAST (Fourier amplitude sensitivity testing), were employed to decompose the total variance in thermocouple measurements, pinpointing the primary simulation parameters driving the effects. This research offers valuable insights into the sensitivity of specific simulation parameters, which can guide future efforts to reduce discrepancies between simulations and experiments and inform experiment design at the TEDS facility.

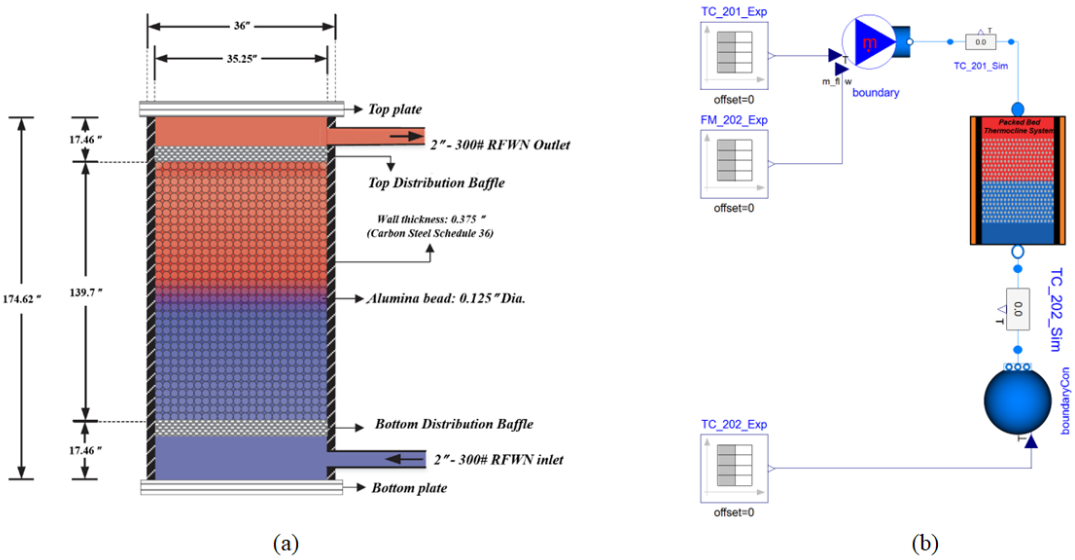


Figure 2: TES system: (a) geometry and (b) Modelica model

2. DATA PREPARATION

2.1. Facility and Simulation Description

The TES system implemented within TEDS consists of a single packed-bed thermocline tank, containing alumina beads with a diameter of 1/8 inch, specifically T-99 PROX-SVERS manufactured by Christy Catalytics, serving as the filler material. The tank dimensions are 174.62 inches in height with an outer diameter of 36 inches. The main body of the vessel is a 36 inch pipe with a standard thickness (STD) pipe schedule to meet the loop's design pressure limits. Inside of the tank is an upper and lower plenum, which will allow for sufficient mixing of the oil before traveling through the packed pebble bed. These plenums are formed by flow distribution plates at the top and bottom of the tank. These plates will also prevent the alumina beads from leaving the tank, with the lower flow distribution plate supporting the weight of the alumina beads. Figure 2 shows the geometry of the TES system and its corresponding Modelica model.

2.2. Data Collection and Preprocessing

The input data is a three dimensional (3D) tensor with shape $(1024, 46, 4)$, where the dimensions represent (samples, time steps, features). There are 1,024 Modelica TES simulations/samples, each defined by four input features: shape factor (sf), porosity (por), inlet mass flow rate (msf), and time index. Shape factor and porosity are static while inlet mass flow rate and the time index evolve with time. The time axis specifies the input value over time for 46 time steps, with each step being 400 seconds. At each time step, mass flow rate follows a certain pattern based on the perturbed time and mass flow rate values, which are shown in Figure 3(a).

Initially, we created a preliminary dataset with 140 samples and three additional input features: inlet temperature ($Tint$), outlet temperature ($Tout$), and outlet pressure ($pout$). These parameters were then removed as our preliminary sensitivity analysis indicated they were not significant in the temperature response of TES. This smaller dataset was used for this prescreening, and then we expanded that to 1,024 samples where only the shape factor, porosity, and mass flow rate were perturbed.

The output data is a 3D tensor with shape $(1024, 13, 6)$, where the dimensions represent (samples, sensors, time slices). There are again 1,024 Modelica TES simulations/samples, and each simulation yields temperature (in Kelvin) measurements from 13 thermocouples distributed around the tank, taken at six specific

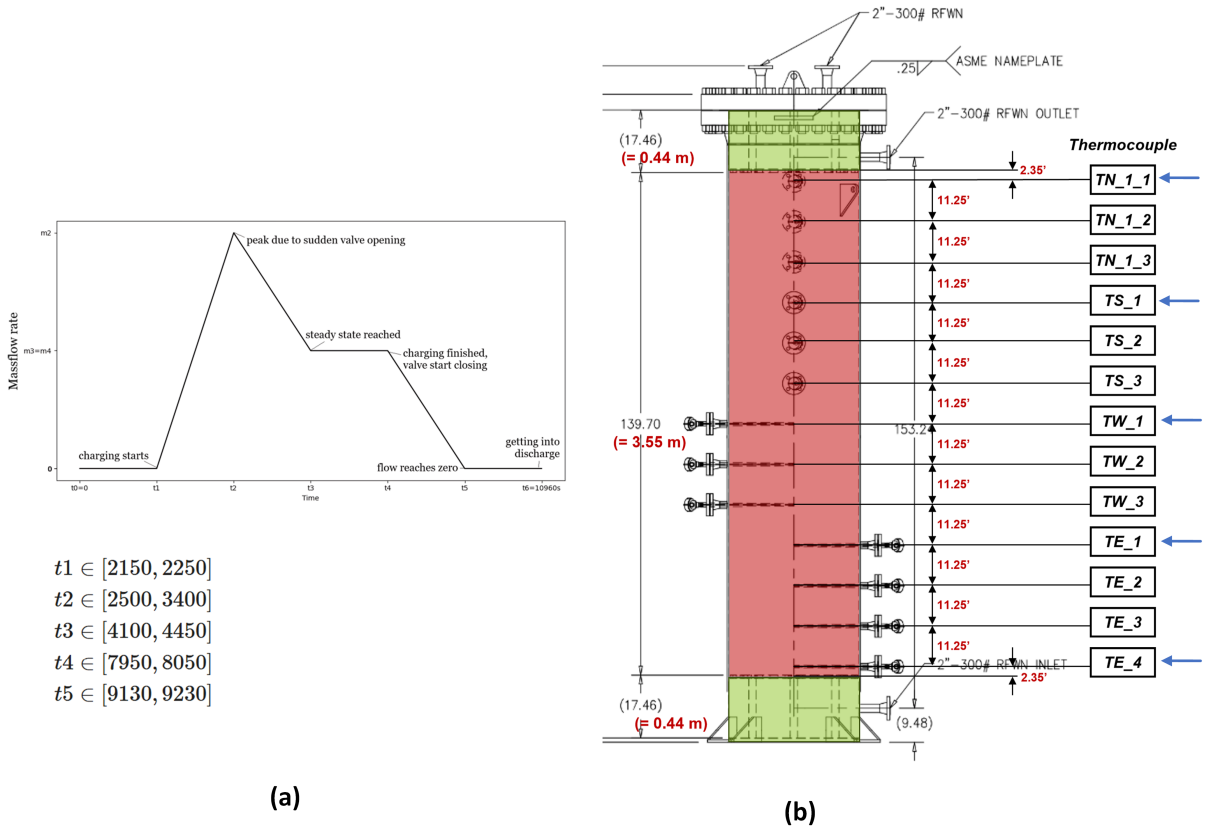


Figure 3: (a) Mass flow rate pattern during the experiment and (b) thermocouple locations with selected ones marked by the blue arrows

time stamps: 4,000, 6,000, 8,000, 10,000, 12,000, and 14,000 seconds. To reduce the volume of the output data, five sensors that cover the axial height of the tank were selected, which are shown in Figure 3(b).

Instead of relying on time-dependent neural networks like recurrent models, we employed a data transformation trick that enables using classical feedforward neural networks and other ML methods. The input data is reshaped into a 2D array with dimensions (samples, features) by removing the "time" feature, retaining only static features, and treating the mass flow rate at each time point as distinct features. This results in 20 input features after excluding time steps where the mass flow rate (msf) remains zero across all samples.

For the output data, measurements from the five selected sensors across six time stamps per sensor are flattened into a 2D array with dimensions (samples, outputs), yielding 30 outputs (five sensors × six time steps). As a result, the final input and output shapes are (1024, 20) and (1024, 30), respectively, making them compatible with most standard ML algorithms.

3. METHODOLOGY

3.1. Automatic Machine Learning with pyMAISE

pyMAISE (Michigan Artificial Intelligence Standard Environment) [9] offers an automated pipeline of evaluating classical and deep neural network models for nuclear engineering applications along with other features, including model selection, cross-validation, hyperparameter tuning, parallel search, explainable ML, and more. In this work, our use of pyMAISE includes three main phases: preprocessing, which handles data splitting, scaling, and encoding; tuning, which optimizes hyperparameters for various models

using user-defined search spaces; and postprocessing, which evaluates models and provides performance diagnostics based on the best hyperparameters. In this work, we have selected six types of ML models as input to the pyMAISE search: *linear regression, lasso regression, decision trees, random forests (RF), k-nearest neighbors, and feedforward neural networks (FNN)*. All these models can be trained using our data described in Section 2.

For all models, 20% of the dataset were used for testing. Both the inputs and outputs are scaled with min-max scaling. Hyperparameter tuning utilized a random search with 200 iterations for classical methods (linear regression, lasso regression, decision trees, RFs, and k-nearest neighbors) and 50 iterations of Bayesian search for the FNN. The selection of random and Bayesian search methods was influenced by the computational cost of training each configuration. Classical methods are relatively inexpensive to train compared to FNNs, allowing for more random search iterations. Each search method included five-fold cross-validation, resulting in 1,000 fits for each classical method and 250 fits for the FNN, each running for 50 epochs. After hyperparameter tuning, the top five configurations of each method based on validation R^2 were retrained on the entire training dataset and evaluated on the testing dataset, with the epochs for FNNs increased to 300.

Model performance is evaluated using traditional regression metrics like R^2 (coefficient of determination), MAE (mean absolute error), MSE (mean squared error), and root MSE (RMSE). pyMAISE then generates a set of diagnostics and validation plots for all models and outputs considered in the search pipeline, which aids in exploring the performance of each model on this dataset. Some of these plots will be provided in the results section.

3.2. Sensitivity Analysis

Following a successful training of an accurate surrogate model with ML, we leverage that model to perform variance-based sensitivity analyses. Here we employ two popular methods: Sobol indices and FAST. These methods help determine how variability in the output (e.g., thermocouple response) can be attributed to different input variables (e.g., sf, por) in the model. The Sobol and FAST indices provide valuable insights into both the individual contributions of input variables and their interactions.

The Sobol method offers two key indices to measure sensitivity. The first-order Sobol index (S_i) quantifies the main effect of an individual input variable X_i on the output Y. It measures the fraction of the output variance explained solely by X_i. The total-order Sobol index (S_{T_i}) accounts for both the main effect of X_i and its interactions with other input variables $X_{\sim i}$. This index captures the total contribution of X_i to the output variance. The corresponding expressions for these indices are [10]:

$$S_i = \frac{\text{Var}(\mathbb{E}[Y \mid X_i])}{\text{Var}(Y)} \tag{1}$$

$$S_{T_i} = 1 - \frac{\text{Var}(\mathbb{E}[Y \mid X_{\sim i}])}{\text{Var}(Y)} \tag{2}$$

where $\text{Var}(Y)$ is the total variance of Y, $\mathbb{E}[Y \mid X_i]$ is the expected value of Y given X_i, and $\text{Var}(\mathbb{E}[Y \mid X_i])$ is variance of the conditional expectation of Y given X_i. Similarly, $X_{\sim i}$ expresses all input variables except X_i, $\mathbb{E}[Y \mid X_{\sim i}]$ is the expected value of Y given $X_{\sim i}$, and $\text{Var}(\mathbb{E}[Y \mid X_{\sim i}])$ is the variance of the conditional expectation of Y given $X_{\sim i}$.

The FAST method aims to achieve the same objectives as Sobol in decomposing the output variance, but it uses a different concept through Fourier transforms to map inputs to output variability. Each input variable X_i is represented as a sinusoidal function with a unique frequency ω_i. This transformation ensures that variations in X_i create periodic changes in the model output Y at a distinct frequency. The formulas for the first-order and total FAST indices are [11]:

$$S_i = \frac{\sum_{k=1}^{\infty} A_{i,k}^2}{\sum_{k=1}^{\infty} A_k^2}$$

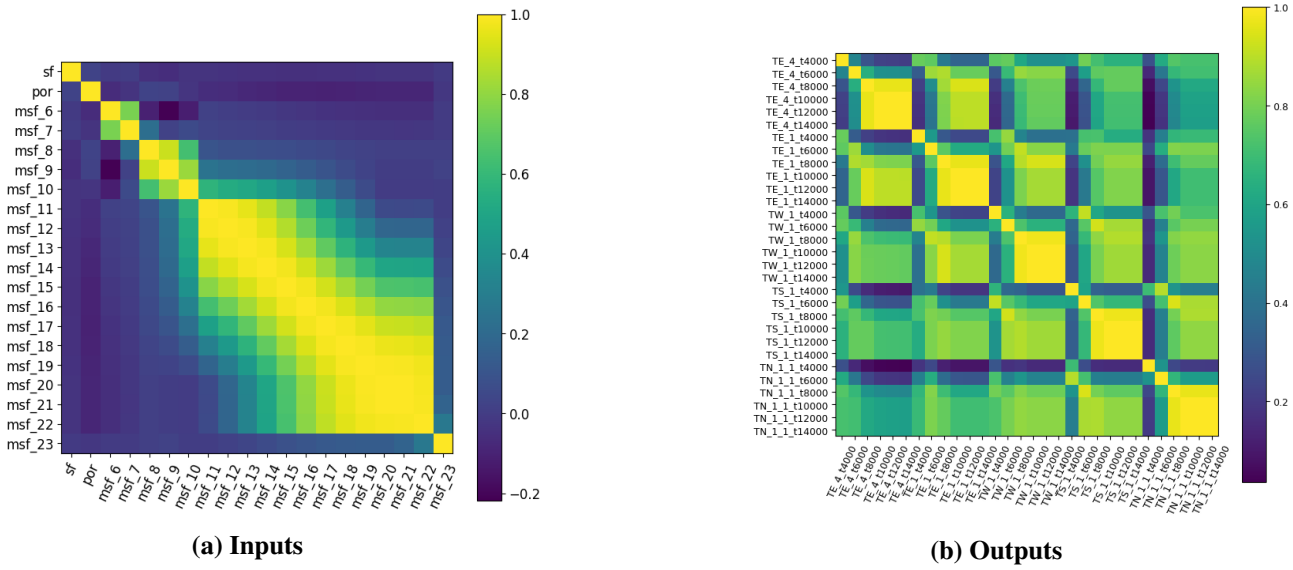
$$S_{T_i} = \frac{\sum_{k \in \mathcal{F}_i} A_k^2}{\sum_{k=1}^{\infty} A_k^2}$$

where $A_{i,k}^2$ is the squared amplitude of the fundamental frequency for X_i, A_k^2 is the squared amplitude of the k-th frequency for all variables (i.e., represents the total spectral energy), and $\mathcal{F}_i$ is the set of harmonics associated with X_i (both direct effect and through interactions).

Overall, both Sobol and FAST sensitivity indices verify parameter sensitivity through two methods. They are essential for understanding the influence of input variables and optimizing model performance by reducing uncertainties associated with key factors.

4. RESULTS

Before we present the results for the ML and sensitivity analyses, Figure 4 presents the correlation matrices for the input and output data, which could also give an impression of the nature of the data we have in this problem. To read the x and y labels in Figure 4(b), the sensor name is given first, then the time stamp is provided after the letter "t". For example, TN_1_1_t14000 refers to the response of the sensor TN_1_1 at time 14,000 seconds, and so on.



(a) Inputs **(b) Outputs**

Figure 4: Correlation matrices for inputs and outputs, where msf_i is the mass flow rate of time step i

Regarding the input data, shape factor and porosity are uncorrelated with other features, which aligns with expectations given that they were sampled independently. Mass flow during time steps 11–22 shows a high correlation, corresponding to the interval where mass flow remains relatively stable and significantly correlated with past time steps. In the output data, for each sensor, the first two time readings are uncorrelated with subsequent ones, while the later readings are highly correlated. This pattern reflects a period of stability in the inlet temperature, mirroring the pattern in the inlet mass flow rate. This also indicates that the correlation between the sensors themselves and within the time span for each sensor is significant and further simplification in the output space may be considered.

Table 1 reveals the top models found by the pyMAISE search, ranked by their highest test R^2 scores, along with their performance across various metrics following the postprocessing phase. The leading FNN demonstrates exceptional results, achieving an average MAE of under 2 K (less than 0.5% of the recorded temperatures) and an average R^2 score exceeding 0.99 in the test set (i.e., no overfitting signs), significantly

surpassing the best results of the classical ML algorithms. RF models come next as second-tier models for this problem.

The best FNN architecture consists of one input layer, two hidden layers, and one output layer. It utilizes the Adam optimizer with a learning rate of 9.6×10^{-4} and a batch size of 8. The input layer, with 20 nodes, aligns with the input shape. The first hidden layer contains 812 neurons with ReLU activation, while the second hidden layer has 596 neurons, also with ReLU activation. The output layer, with 30 neurons, matches the output shape and employs a linear activation function.

Table 1: Performance metrics of the best three models found by pyMAISE for FNN and RF

Model	Train R^2	Train MAE	Test R^2	Test MAE	Test MSE	Test RMSE
FNN 1	0.9966	0.6069	0.9946	0.6727	1.4160	1.1899
FNN 2	0.9941	1.1849	0.9911	1.2523	5.5614	2.3583
FNN 3	0.9833	1.6408	0.9803	1.7602	9.0445	3.0074
RF 1	0.9722	2.2977	0.9076	4.4379	63.2479	7.9529
RF 2	0.9579	2.7451	0.8890	4.7529	71.4078	8.4503
RF 3	0.9524	2.8949	0.8868	4.8594	75.7199	8.7017

Figure 5 shows the diagonal validation plots for the top-performing FNN and RF models. Again, the FNNs performed exceptionally well, as evidenced by the results that display a tight spread close to $y = x$, while RF predictions show a larger spread around the diagonal line. Figure 6 displays the absolute relative error plot for the same models. The diagonal validation and absolute relative error plots are consistent with the performance metrics. The FNN achieved a maximum absolute relative error of 3.5% for one test sample in the sensor TS_1, while the maximum relative error for RF is close to 10%.

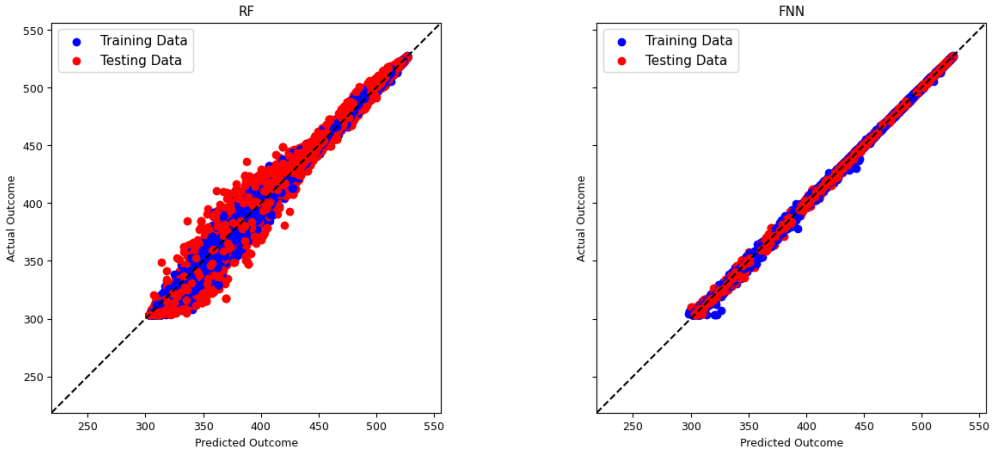


Figure 5: Diagonal validation plots for the top-performing RF and FNN models

4.1. Sensitivity Analysis Results

Moving with the best FNN as a surrogate model, Figure 7 and Figure 8 present the results of the Sobol and FAST indices for thermocouples TE_4 and TW_1, respectively. The analysis reveals that porosity plays a

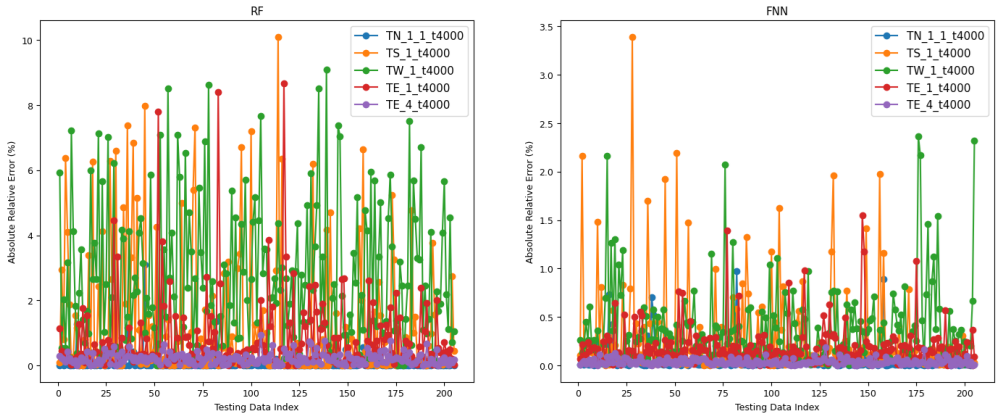


Figure 6: Absolute relative error plots for the top-performing RF and FNN models

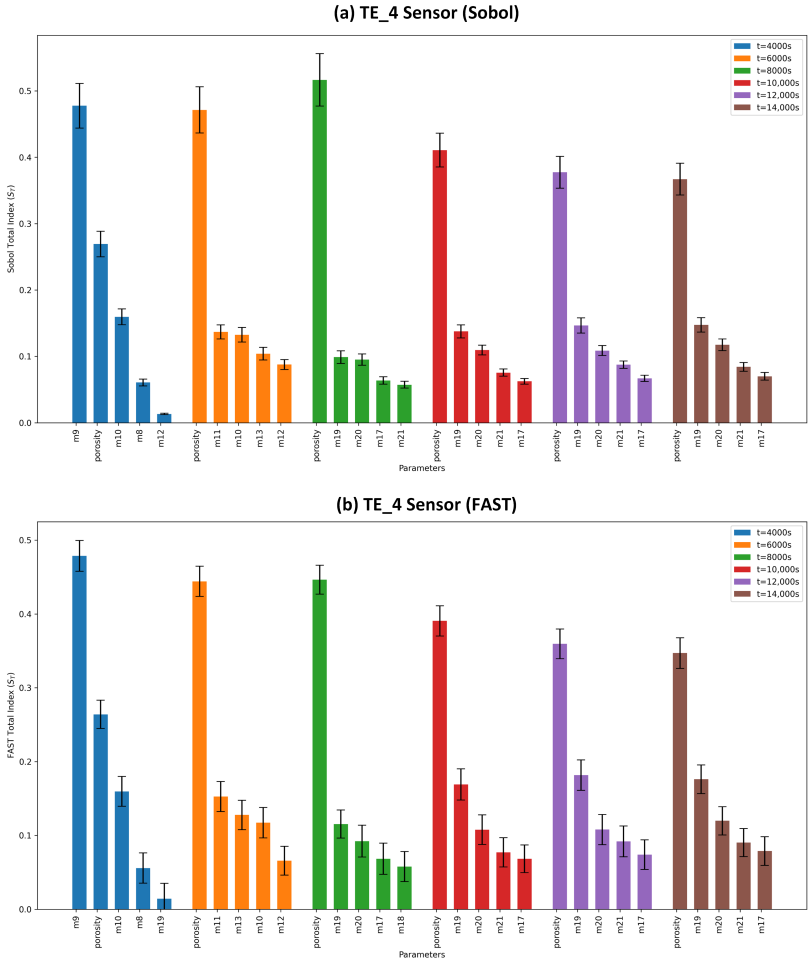


Figure 7: (a) Sobol and (b) FAST sensitivity results for sensor TE_4

significant role in contributing to the thermal response variability and this dominance is visible across all time steps between 4,000 and 14,000 seconds. Also, the peak inlet mass flow rate (m9, m10) shows a strong influence on TE_4, while (m9, m7) shows an influence on the sensor TW_1 response. Peak mass flow rate (m9) was more impactful on TE_4 response at t = 4,000 seconds, while porosity then became dominant for other time steps. Other variables, such as shape factor (sf), show minimal overall impact on all sensors and time steps.

Our analysis reveals that the FAST and Sobol methods produce consistent rankings for the dominant input features, particularly the top three. However, the rankings of lower order (negligible) input features may differ slightly due to their close numerical values, which can be influenced by the statistical uncertainties inherent in the Sobol and FAST estimations. In addition, the results for the three other sensors (TE_1, TS_1, and TN_1_1) are not shown due to space limitation but have a similar trend to the two presented sensors.

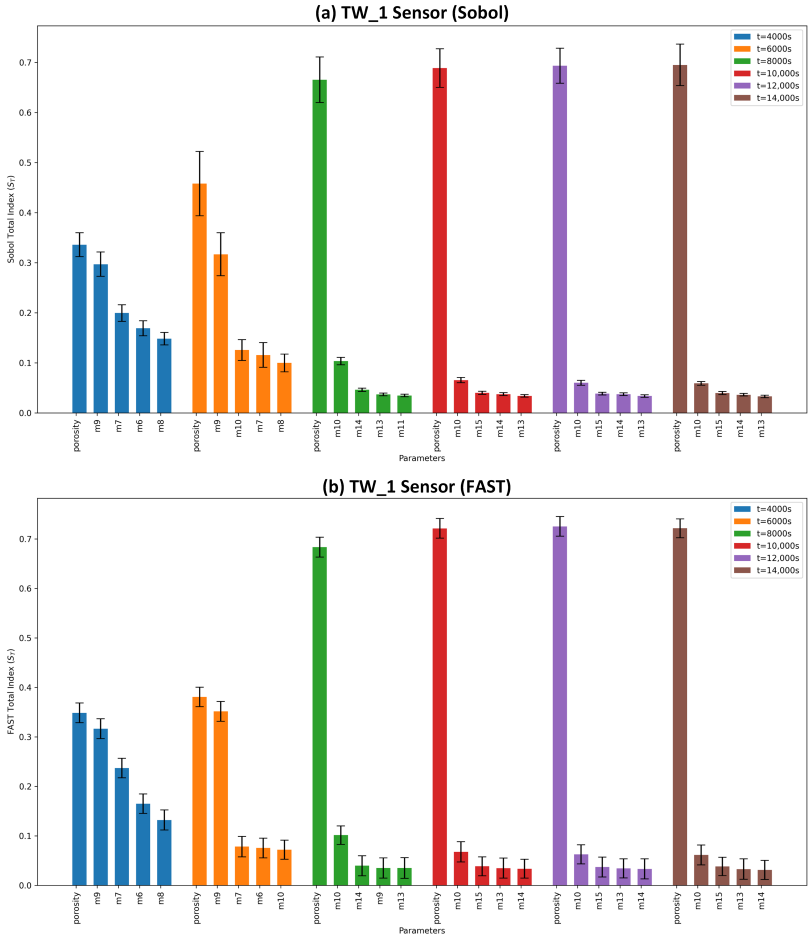


Figure 8: (a) Sobol and (b) FAST sensitivity results for sensor TW_1

5. CONCLUSIONS

This study applies variance-based sensitivity analysis methods to facilitate uncertainty quantification for TES systems, leveraging a Modelica/Dymola-based simulation to enhance model reliability by analyzing the impact of shape factor, porosity, and inlet mass flow rate on the system thermal response. Our preliminary analysis demonstrated that inlet temperature, outlet temperature, and outlet pressure have minimal impact on the TES thermal response. Then, using feedforward neural networks as surrogate models, the

variance-based sensitivity analysis identified porosity and peak inlet mass flow rate as the most influential parameters on thermal storage tank temperature, with peak mass flow rate dominating early time steps and porosity maintaining consistent influence. For surrogate modeling, a feedforward neural network with over 1,300 neurons outperformed other ML models, achieving an R^2 above 0.99 and a mean absolute error below 1 Kelvin. This combination of ML and uncertainty quantification techniques underscores its potential to improve the accuracy and interpretability of complex thermal systems, with sensitivity analysis insights enhancing the Modelica/Dymola TES simulation. Future work aims to expand the analysis to additional components beyond the thermal storage tank, incorporate more intricate boundary conditions, and refine predictions through larger datasets and advanced ML architectures.

ACKNOWLEDGEMENTS

This work was supported through Idaho National Laboratory's Laboratory Directed Research and Development (LDRD) Program Award Number (24A1081-116FP) under Department of Energy Idaho Operations Office contract no. DE-AC07-05ID14517.

REFERENCES

[1] K. L. Frick, S. Bragg-Sitton, and M. Garrouste. "Validation and Verification Methodology for INL Modelica-based TEDS Models Via Experimental Results." Technical report, Idaho National Laboratory, Idaho Falls, ID (2021).

[2] K. Frick, S. Bragg-Sitton, and C. Rabiti. "Modeling the idaho national laboratory thermal-energy distribution system (teds) in the modelica ecosystem." *Energies*, **volume 13**(23), p. 6353 (2020).

[3] V. Novotny, J. Kim, T. J. Morton, S. E. Creasman, and D. M. Mikkelson. "Validation and Verification of TEDS Facility HYBRID Modeling." Technical report, Idaho National Laboratory (INL), Idaho Falls, ID (United States) (2023).

[4] T. Kajihara, D. Garrett, J. Kim, L. Lin, P. W. Talbot, and J. M. Browning. "Digital Twin for Optimizing Real-time Economy of the Integrated Energy Systems." Technical report, Idaho National Laboratory (INL), Idaho Falls, ID (United States) (2023).

[5] M. I. Radaideh and T. Kozlowski. "Surrogate modeling of advanced computer simulations using deep Gaussian processes." *Reliability Engineering & System Safety*, **volume 195**, p. 106731 (2020).

[6] D. Price, M. I. Radaideh, D. O'Grady, and T. Kozlowski. "Advanced BWR criticality safety part II: Cask criticality, burnup credit, sensitivity, and uncertainty analyses." *Progress in Nuclear Energy*, **volume 115**, pp. 126–139 (2019).

[7] M. I. Radaideh and T. Kozlowski. "Analyzing nuclear reactor simulation data and uncertainty with the group method of data handling." *Nuclear Engineering and Technology*, **volume 52**(2), pp. 287–295 (2020).

[8] M. I. Radaideh, S. Surani, D. O'Grady, and T. Kozlowski. "Shapley effect application for variance-based sensitivity analysis of the few-group cross-sections." *Annals of Nuclear Energy*, **volume 129**, pp. 264–279 (2019).

[9] P. A. Myers, N. Panczyk, S. Chidige, C. Craig, J. Cooper, V. Joynt, and M. I. Radaideh. "pyMAISE: A Python platform for automatic machine learning and accelerated development for nuclear power applications." *Progress in Nuclear Energy*, **volume 180**, p. 105568 (2025).

[10] E. Borgonovo and E. Plischke. "Sensitivity analysis: A review of recent advances." *European Journal of Operational Research*, **volume 248**(3), pp. 869–887 (2016).

[11] G. J. McRae, J. W. Tilden, and J. H. Seinfeld. "Global sensitivity analysis—a computational implementation of the Fourier amplitude sensitivity test (FAST)." *Computers & Chemical Engineering*, **volume 6**(1), pp. 15–25 (1982).