Check for updates

Proceedings of the ASME 2025
International Mechanical Engineering Congress and Exposition
IMECE2025
November 16-20, 2025, Memphis, Tennessee

IMECE2025-167746

HYBRID APPROACH BASED ON SIMULATION AND EXPERIMENTAL DATA FOR SURROGATE-BASED MODELING: A CASE STUDY OF A HIGH-TEMPERATURE HEAT PUMP

Loukas Kyriakidis[1,*], Ashish Thapa[1]

[1]German Aerospace Center, Institute of Low-Carbon Industrial Processes, Cottbus, Germany

ABSTRACT

Surrogate models are widely used to approximate complex simulation models, reducing computational costs in real-world applications. However, models based exclusively on simulation data can be inaccurate due to factors like missing real-world effects and discrepancies between simulated and actual performance. Similarly, relying solely on experimental data is costly, time-consuming, and prone to measurement noise and inconsistencies. To overcome these limitations, hybrid modeling has emerged as a promising approach, which integrate both simulation and experimental data, improving the accuracy and reliability of surrogate models. This study investigates methods like transfer learning and adaptive weighting to combine simulation and experimental data effectively, evaluating them in terms of accuracy, robustness, and training time. Specifically, we examine how different hybrid methods perform depending on the amount of available data, dimensionality of the problem, and discrepancy between simulation and experimental data. By testing these techniques on both analytical cases and a real-world high-temperature heat pump, we identify the scenarios in which each method is most effective. Our results demonstrate that hybrid models outperform traditional models based solely on simulation or experimental data, offering improved predictive accuracy and better robustness. These findings highlight the potential of hybrid modeling in real-world industrial applications.

Keywords: Surrogate models, Hybrid modeling, Adaptive weighting, Transfer learning

1. INTRODUCTION

Surrogate models, also known as metamodels or response surface models, play a crucial role in various applications such as design and operational optimization [1, 2]. Their primary advantage lies in their ability to approximate complex and computationally expensive simulation models with significantly reduced computational effort, capturing their input-output behavior and enabling rapid evaluations.

While modeling based on physical equations remains the gold standard for capturing system behavior, it can become extremely complex when applied to systems with many interacting components [3]. In such cases, deriving detailed physical models for each subsystem is not only time-consuming but often impractical due to limited knowledge of internal processes or unmeasurable variables. As a result, data-driven approaches such as surrogate modeling have emerged as attractive alternatives or complements to physics-based models. Surrogate models aim to approximate system behavior using available data, thereby reducing modeling effort while maintaining sufficient predictive accuracy.

Surrogate models can be broadly categorized based on the source of data used to construct them. Simulation-based surrogate models are widely used to approximate complex systems, leveraging simulations to generate data on which predictive frameworks are built, and reducing the dependency on costly and time-consuming experimental measurements. Commonly employed techniques include statistical methods such as Gaussian process regression (GPR, also known as kriging), radial basis functions (RBF), and polynomial regression (PR), as well as machine learning approaches like artificial neural networks (ANNs), support vector regression (SVR), and random forest regression (RFR) [4]. Depending on the application, different types of surrogate models are preferred. For black-box optimization, where the internal structure of the model is not required and the focus is purely on the predictive accuracy, non-parametric models such as GPR, RBF, ANN, and SVR are typically favored. These models are particularly effective for guiding optimization algorithms like Bayesian optimization or evolutionary strategies. In contrast, for applications that demand equation-based representations, PR or other nonlinear transformations, including exponential, logarithmic, ratio, and trigonometric functions [5] are more appropriate. Despite their advantages, simulation-based surrogate models can still face challenges in capturing real-world behavior. Simulations inherently rely on assumptions and idealizations that may

*Corresponding author: loukas.kyriakidis@dlr.de

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

not hold in practice, leading to discrepancies between the surrogate predictions and actual system responses.

On the other hand, surrogate models constructed solely from experimental data can more accurately capture complex physical effects that might be oversimplified or omitted in simulations. However, this approach presents its own set of challenges. Experimental data is often noisy, which can degrade the predictive accuracy of the surrogate model if not properly addressed [6]. Additionally, acquiring high-quality experimental data is typically expensive, time-consuming, and sometimes logistically constrained—especially for systems requiring high-fidelity measurements, extreme operating conditions, or specialized facilities. Furthermore, the number of data points is usually limited, leading to sparse coverage of the input space. These limitations make purely experimental surrogates less practical for many applications. Consequently, hybrid approaches that combine simulation and experimental data are increasingly favored, leveraging the strengths of both while mitigating their individual weaknesses.

To combine the advantages of simulation- and experiment-based surrogate models, hybrid modeling can be applied, leveraging the efficiency and scalability of simulations along with the realism and fidelity of experimental data. This approach overcomes the limitations of using only simulations or experiments by combining them into a single, more robust surrogate model. A notable example of this approach involves the use of neural networks with transfer learning, where knowledge gained from simulation-based pretraining is fine-tuned using experimental data [7]. In this study, we introduce, explore, and compare several forms of hybrid modeling, including transfer learning and adaptive weighting, the latter of which are used to dynamically balance the contribution of experimental and simulated data during model training and prediction. Specifically, we investigate how factors such as the amount of available simulation and experimental data, the dimensionality of the problem, and the level of discrepancy between simulations and experiments influence the effectiveness of each hybrid approach. These methods are applied to both analytical test functions and a real-world case study involving a high-temperature heat pump, providing insight into their robustness and applicability in practical scenarios.

This paper is organized as follows: Section 2 introduces the hybrid methods whose performance is investigated in this work. In Section 3, these methods are evaluated based on two test functions that vary in dimensionality, the number of simulation and experimental data points, and the level of discrepancy between simulations and experiments. Section 4 demonstrates the performance of the hybrid methods on a real-world problem — a high-temperature heat pump. Finally, Section 5 summarizes the main conclusions and provides a perspective for future work.

2. HYBRID MODELS

In this section, we present and describe the hybrid surrogate modeling approaches explored in this work, which aim to combine the strengths of simulation-based and experimental data-driven models.

ANN with Transfer Learning The first hybrid modeling approach investigated in this work is an ANN framework that incorporates transfer learning to effectively integrate simulation and experimental data [7]. In this method, a base neural network is initially trained using a large dataset generated from simulations, enabling it to learn general patterns and trends inherent in the simulated system behavior. This pretraining phase captures the broad structure of the system with relatively low computational cost. Subsequently, the pretrained model is fine-tuned using a smaller, high-fidelity dataset derived from experimental measurements. This fine-tuning step adjusts the network's parameters to better reflect the actual system dynamics, correcting for discrepancies between simulations and reality.

Weighted Surrogate Models The second hybrid modeling approach investigated in this work is based on a weighted combination of surrogate models trained separately on simulation and experimental data. This method allows for leveraging the strengths of both model types while maintaining flexibility in adjusting their contributions depending on data availability and reliability.

The hybrid surrogate model $\hat{y}_{\text{hyb}}$ is constructed as a combination of a simulation-based model $\hat{y}_{\text{sim}}$ and an experimental model $\hat{y}_{\text{exp}}$ as follows:

$$\hat{y}_{\text{hyb}} = w_{\text{sim}}^{\text{norm}} \hat{y}_{\text{sim}} + w_{\text{exp}}^{\text{norm}} \hat{y}_{\text{exp}} \tag{1}$$

where $w_{\text{sim}}^{\text{norm}} + w_{\text{exp}}^{\text{norm}} = 1$. The normalized weights $w_{\text{sim}}^{\text{norm}}, w_{\text{exp}}^{\text{norm}}$ reflect the trust or availability of each data source. In this work, both $\hat{y}_{\text{sim}}$ and $\hat{y}_{\text{exp}}$ are either ANNs or PR models. The degree of the PR depends on the complexity of the problem at hand—simpler systems may require only a low-degree polynomial, while more complex relationships may need higher-degree terms to achieve an accurate approximation.

A critical aspect of this approach is the determination of the weights $w_{\text{sim}}^{\text{norm}}$ and $w_{\text{exp}}^{\text{norm}}$. Simple hyperparameter tuning may not always be sufficient or appropriate—particularly in cases where a model performs well on experimental validation but poorly on simulation data, or vice versa. This can occur due to differences in noise levels, data quantity, or systematic modeling errors.

To address this challenge, this work proposes an adaptive weighting scheme that accounts for both the number of available data points and the variance (as an indicator of uncertainty or noise) within each dataset. The weights are computed as:

$$w_{\text{sim}} = \frac{N_{\text{sim}} \text{Var}_{\text{exp}}}{(N_{\text{sim}} + N_{\text{exp}})(\text{Var}_{\text{sim}} + \text{Var}_{\text{exp}})} \tag{2}$$

$$w_{\text{exp}} = \frac{N_{\text{exp}} \text{Var}_{\text{sim}}}{(N_{\text{sim}} + N_{\text{exp}})(\text{Var}_{\text{sim}} + \text{Var}_{\text{exp}})} \tag{3}$$

Here, N_{sim} and N_{exp} denote the number of simulation and experimental samples, while Var_{sim} and Var_{exp} represent the variance in each respective dataset. The normalized weights are determined as follows:

$$w_{\text{sim}}^{\text{norm}} = \frac{w_{\text{sim}}}{w_{\text{sim}} + w_{\text{exp}}}, \quad w_{\text{exp}}^{\text{norm}} = \frac{w_{\text{exp}}}{w_{\text{sim}} + w_{\text{exp}}} \tag{4}$$

This formulation favors the model trained on the data source that is both more abundant and more reliable (i.e., less noisy), while still ensuring that both sources contribute proportionally to the final prediction. This adaptive scheme ensures a robust and principled combination of models, reducing the risk of overfitting to noisy

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

experimental data or under-representing accurate simulation data, as shown in Sections 3.

Surrogate Model with Weighted Cost Function The third hybrid modeling approach explored in this work avoids the need for two separate surrogate models by fitting a single model that minimizes a weighted cost function combining both simulation and experimental data. The loss function to be minimized is:

$$J = \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \left(\hat{y}_i - y_{\text{sim},i}\right)^2 + \lambda \frac{1}{N_{\text{exp}}} \sum_{j=1}^{N_{\text{exp}}} \left(\hat{y}_j - y_{\text{exp},j}\right)^2 \tag{5}$$

Here, $\hat{y}_i$ and $\hat{y}_j$ denote the model prediction for the i-th and j-th data point in the simulation and experimental dataset, while $y_{\text{sim},i}$ and $y_{\text{exp},j}$ are the corresponding target values from the simulation and experimental datasets. The parameter λ controls the relative importance of the experimental data in the training process.

To ensure balanced learning that reflects data quality and quantity, λ is calculated adaptively as:

$$\lambda = \frac{N_{\text{exp}} \text{Var}_{\text{sim}}}{(N_{\text{sim}} + N_{\text{exp}})(\text{Var}_{\text{sim}} + \text{Var}_{\text{exp}})} \tag{6}$$

This unified modeling approach is conceptually inspired by the physics-informed neural networks [8], where physical laws guide learning. However, instead of relying on governing equations, we incorporate domain knowledge through simulation data and calibrate the model with experimental measurements using data-informed regularization rather than equation-based constraints. The proposed strategy ensures that the surrogate model is not biased toward the dominant dataset, taking into account not only the number of data points but also the variance within each dataset. This allows the model to generalize effectively across both simulation and experimental domains. This unified approach is particularly appealing when a single, compact model is preferred for deployment or interpretability.

3. TEST FUNCTIONS

In this section, we evaluate the performance of the proposed hybrid surrogate modeling strategies using two well-known benchmark functions.

3.1. Function Definition and Setup

The benchmark functions considered in this work are the Rosenbrock and the Goldstein–Price function [9]. These functions are selected due to their distinct levels of nonlinearity [10]. While the Rosenbrock function exhibits moderate nonlinearity, the Goldstein–Price function is significantly more nonlinear and complex. To assess the scalability and robustness of the hybrid models, we vary the input dimensionality as $d = 2, 5, 10, 15, 20$. The goal is to test how well the surrogate modeling methods generalize across varying levels of complexity and input space dimensionality. The Rosenbrock function in d dimensions is defined as:

$$f(\boldsymbol{x}) = \sum_{n=1}^{d-1} \left[100(x_{n+1} - x_n^2)^2 + (1 - x_n)^2\right], \quad \boldsymbol{x} \in \mathbb{R}^d \tag{7}$$

where $\boldsymbol{x}$ is the input vector of the decision variable x. The original Golstein-Price function is limited to two dimensions. To extend it to higher-dimensional input spaces, we define a composite version by applying the standard 2-D Goldstein–Price function to overlapping pairs of consecutive inputs and summing the results. This captures the complex pairwise interactions across the input space and increases the function's nonlinearity with dimension. The generalized form is defined as:

$$\begin{aligned} f(\boldsymbol{x}) = \sum_{n=1}^{d-1} \Big[& 1 + (x_n + x_{n+1} + 1)^2 \Big(19 - 14x_n + 3x_n^2 - 14x_{n+1} \\ & + 6x_n x_{n+1} + 3x_{n+1}^2\Big) \\ & \Big(30 + (2x_n - 3x_{n+1})^2 \Big(18 - 32x_n + 12x_n^2 + 48x_{n+1} \\ & - 36x_n x_{n+1} + 27x_{n+1}^2\Big)\Big)\Big] \end{aligned} \tag{8}$$

with $\boldsymbol{x} \in \mathbb{R}^d$. To mimic real-world scenarios where both simulations and experiments are available, we generate two types of data. The experimental data are considered as the ground truth and are obtained by directly evaluating the benchmark function, i.e., $y_{\text{exp}} = f(\boldsymbol{x})$. The simulation data are generated by varying the true function values with Gaussian noise to reproduce modeling inaccuracies, i.e., $y_{\text{sim}} = f(\boldsymbol{x}) + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Here, σ is the standard deviation of the noise and serves as a parameter for controlling the simulation fidelity. This setup allows us to investigate the influence of the simulation noise on the prediction quality.

Another parameter, which is varied across the investigated scenarios, is the number of simulation and experimental data points. By altering both the data quantity and quality (via noise), we can systematically study the robustness and adaptability of the hybrid surrogate models.

3.2. Results

In the following, we compare the performance of three hybrid surrogate models on the benchmark functions described above. Hybrid model 1 is based on ANN with transfer learning, hybrid model 2 on the weighted surrogate models, and hybrid model 3 on the surrogate model with weighted cost function. For hybrid model 2, both PR and ANN are explored as underlying surrogate models. PR has been selected as a representative white-box model due to its simplicity, analytical tractability, and computational efficiency, making it particularly fast to train and evaluate. ANN, on the other hand, was chosen as the black-box model because it can flexibly learn complex, nonlinear relationships from data, even in high-dimensional spaces. The models are evaluated across a range of input dimensions, number of simulation and experimental data, and noise levels to systematically assess their accuracy (R^2), robustness, and scalability. In each figure, the left subfigure shows the prediction accuracy tested on the experimental data, while the right subfigure presents the results based on simulation data. As in all investigated cases in this section, a third-degree PR is used for the Rosenbrock and Goldstein-Price function. This choice applies both to hybrid model 2 with PR

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

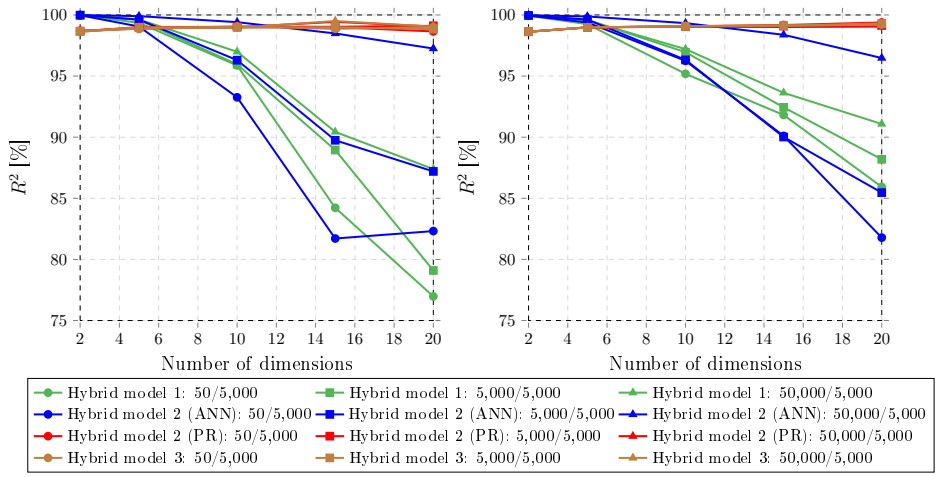


FIGURE 1: R^2 VALUES OF THE HYBRID MODELS EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF DIMENSIONS FOR DIFFERENT NUMBER OF EXPERIMENTAL POINTS, EXEMPLIFIED FOR THE ROSENBROCK FUNCTION WITH AN ERROR OF 0.8. THE LEGEND INDICATES THE NUMBER OF DATA POINTS CONSIDERED: THE FIRST NUMBER REFERS TO EXPERIMENTAL DATA AND THE SECOND TO SIMULATION DATA.

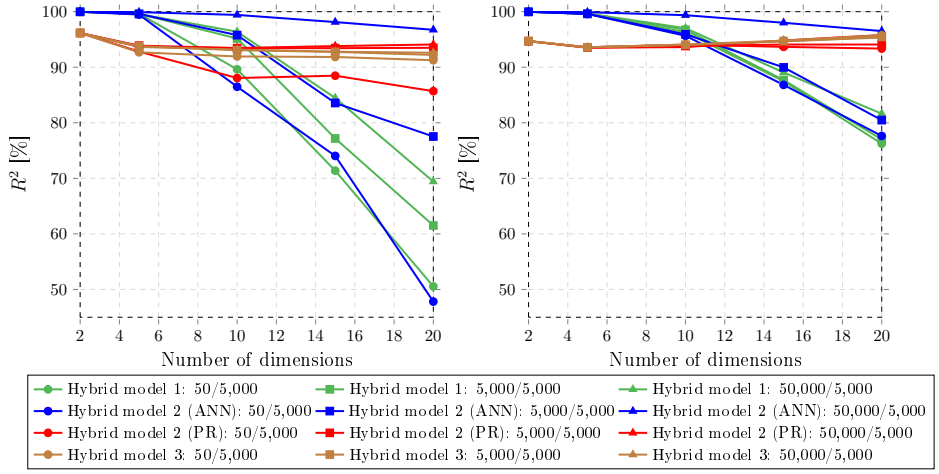


FIGURE 2: R^2 VALUES OF THE HYBRID MODELS EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF DIMENSIONS FOR DIFFERENT NUMBER OF EXPERIMENTAL POINTS, EXEMPLIFIED FOR THE GOLDSTEIN-PRICE FUNCTION WITH AN ERROR OF 5. THE LEGEND INDICATES THE NUMBER OF DATA POINTS CONSIDERED: THE FIRST NUMBER REFERS TO EXPERIMENTAL DATA AND THE SECOND TO SIMULATION DATA.

and hybrid model 3, where the underlying surrogate model in the weighted cost function is a third-degree polynomial for Rosenbrock and for Goldstein-Price function. For hybrid Model 3, the optimization of the unified objective function is performed using the BFGS algorithm from SciPy. Differential Evolution was also tested but it did not show any improvement in the prediction accuracy, while significantly increasing the CPU time; therefore, it is not considered further here. All models are implemented in Python 3.10 and we use the Tensorflow library for ANN and Sklearn for PR. Computations are performed with an Intel(R) Core(TM) i7-8665U CPU. To ensure comparability and stable training behavior, all input and output data are normalized to the range [0, 1] prior to model fitting across all surrogate models.

Figures 1 and 2 show the R^2 values across different input dimensions with 5,000 simulation data and varying number of experimental data for both the Rosenbrock and Goldstein-Price functions. As seen, hybrid model 1 and hybrid model 2 with ANN demonstrate high accuracy for all number of experimental data in low-dimensional spaces. However, as the dimensionality increases, both models require more experimental data to maintain high accuracy; otherwise, the R^2 drops. Hybrid model 2 with PR exhibits almost similar accuracy across the dimensions for all investigated number of experimental data in the Rosenbrock function, but in the case of the Goldstein-Price function, where the nonlinearity and complexity increase, the accuracy on the experimental test set begins to decrease below 90 % when only 50 experimental data points are used, starting from 10 dimensions. With 5,000 and 50,000 experimental data points, the accuracy on this test set remains high. When testing on the simulation data, the prediction accuracy remains consistently high across all

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

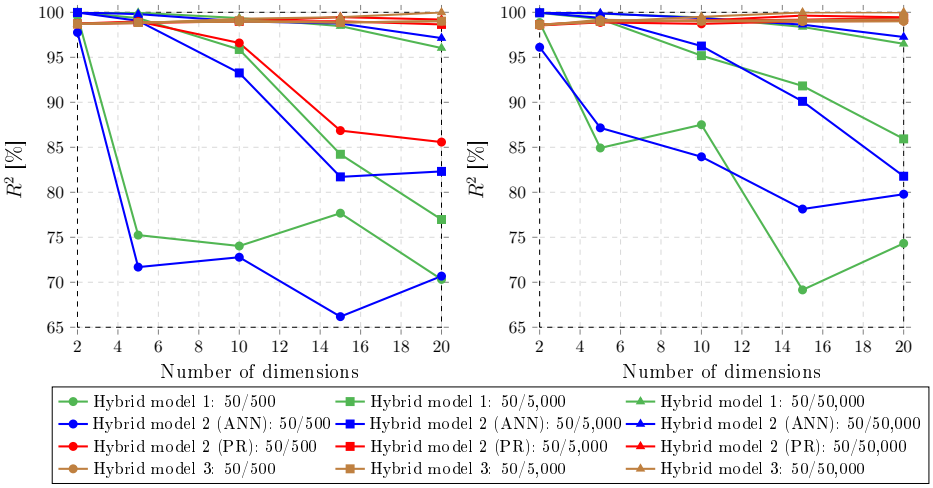


FIGURE 3: R^2 VALUES OF THE HYBRID MODELS EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF DIMENSIONS FOR DIFFERENT NUMBER OF SIMULATION POINTS, EXEMPLIFIED FOR THE ROSENBROCK FUNCTION WITH AN ERROR OF 0.8. THE LEGEND INDICATES THE NUMBER OF DATA POINTS CONSIDERED: THE FIRST NUMBER REFERS TO EXPERIMENTAL DATA AND THE SECOND TO SIMULATION DATA.

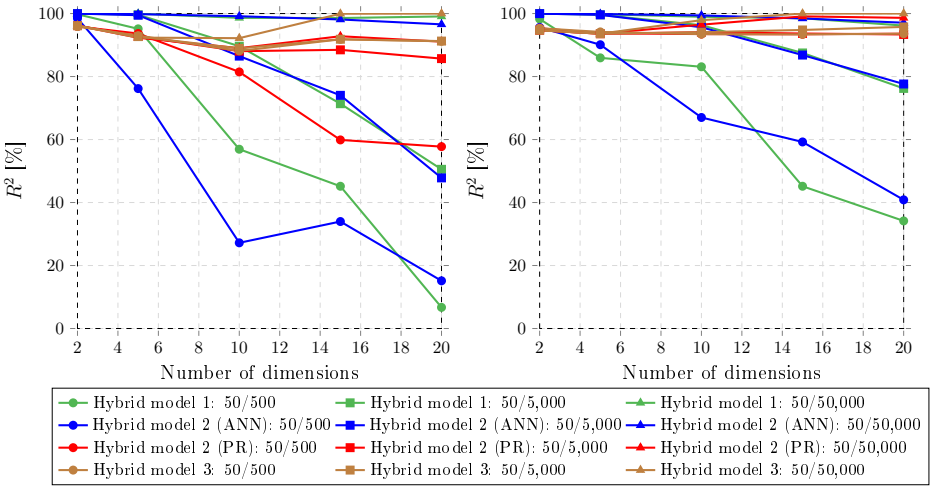


FIGURE 4: R^2 VALUES OF THE HYBRID MODELS EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF DIMENSIONS FOR DIFFERENT NUMBER OF SIMULATION POINTS, EXEMPLIFIED FOR THE GOLDSTEIN-PRICE FUNCTION WITH AN ERROR OF 5. THE LEGEND INDICATES THE NUMBER OF DATA POINTS CONSIDERED: THE FIRST NUMBER REFERS TO EXPERIMENTAL DATA AND THE SECOND TO SIMULATION DATA.

levels of experimental data due to the large and fixed number of simulation points (5,000), which dominates the training and stabilize the model performance. For hybrid model 3, the accuracy follows a similar trend as hybrid model 2 with PR, but for the Goldstein-Price function with experimental data, hybrid model 3 outperforms hybrid model 2, showing higher robustness and scalability, especially when 50 experimental data points are used. This indicates that the surrogate model with a weighted cost function is more effective in handling the increasing complexity of the Goldstein-Price function, particularly when limited experimental data is available, as it is often the case in real-world scenarios. The advantage of hybrid model 3 lies in its integrated training process, where simulation and experimental data are considered through a single weighted loss function. This enables the model to automatically balance the influence of each data source, in contrast to hybrid model 2, which combines two separately trained models. As a result, hybrid model 3 can better extract complementary patterns, leading to improved generalization. It is worth noticing that the prediction accuracy of hybrid model 2 with PR and 3 increases with rising dimensionality. This shows that the increasing dimensions make the function easier to approximate and allow the models to benefit from richer feature spaces.

Figures 3 and 4 present the R^2 values across different input dimensions for both benchmark functions, this time with a fixed number of 50 experimental data points and varying the number of simulation data points. As expected, increasing the number of simulation data helps the hybrid models 1 and 2 with ANN to reach high accuracy, particularly in higher dimensions. This

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

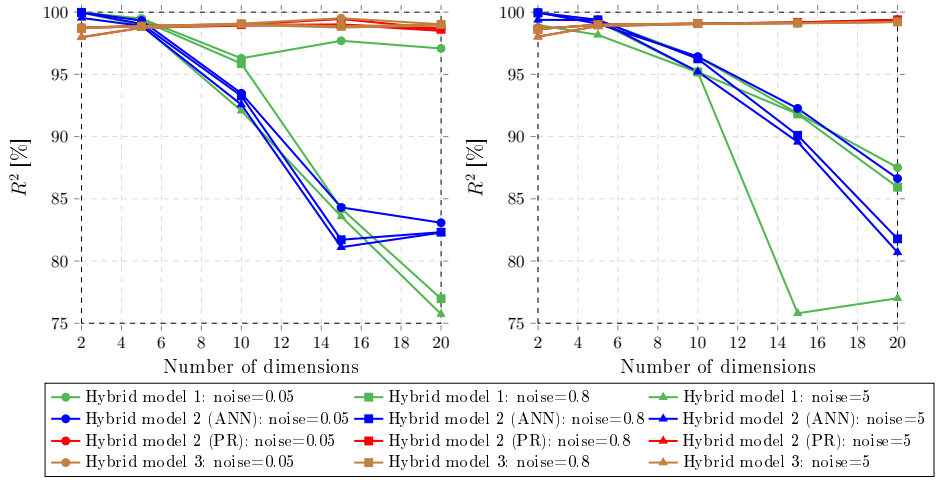


FIGURE 5: R^2 VALUES OF THE HYBRID MODELS EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF DIMENSIONS FOR DIFFERENT NOISE, EXEMPLIFIED FOR THE ROSENBROCK FUNCTION WITH 50 EXPERIMENTAL AND 5,000 SIMULATION POINTS.

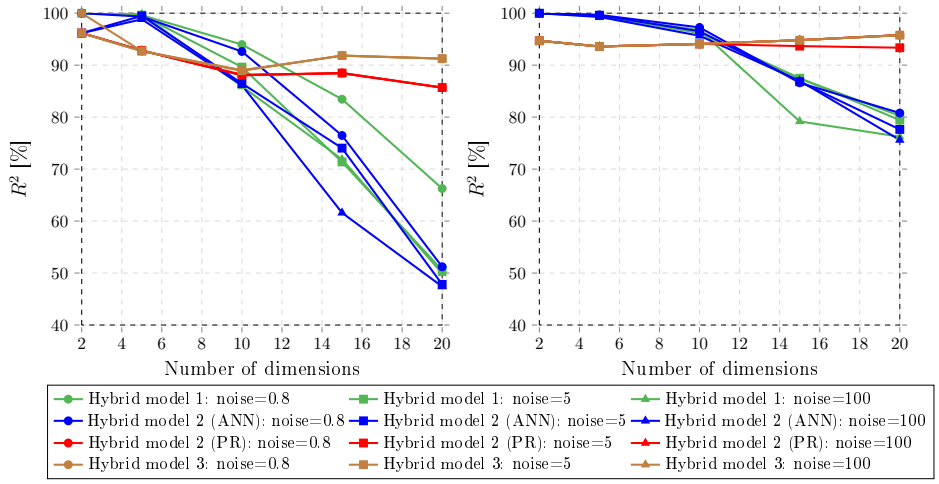


FIGURE 6: R^2 VALUES OF THE HYBRID MODELS EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF DIMENSIONS FOR DIFFERENT NOISE, EXEMPLIFIED FOR THE GOLDSTEIN-PRICE FUNCTION WITH 50 EXPERIMENTAL AND 5,000 SIMULATION POINTS.

demonstrates the models' ability to leverage larger simulation datasets for better generalization. Additionally, hybrid model 3 shows higher accuracy and robustness than hybrid model 2 with PR on the experimental test set, maintaining competitive performance even with limited experimental data. On the simulation test set, hybrid model 3 exhibits similar accuracy to hybrid model 2 with PR, highlighting its stable behavior and scalability across different simulation sizes.

Figures 5 and 6 examine the effect of varying noise levels between simulation and experimental data, using 50 experimental and 5,000 simulation data points. As expected, hybrid model 3 demonstrates strong robustness against noise, with its prediction accuracy remaining almost unaffected across all tested levels. In contrast, higher noise levels significantly reduce the performance of hybrid models 1 and 2 with ANN, particularly in higher dimensions. Notably, different noise levels were investigated for the Rosenbrock and Goldstein-Price functions, with smaller noise magnitudes used for the Rosenbrock function. This is due to the significantly smaller range of y-values in the Rosenbrock function compared to the Goldstein-Price function, ensuring that the noise remains proportionally comparable across both test cases. The performance of hybrid model 2 with PR remains close to that of model 3, further supporting its reliability in noisy environments.

In addition to accuracy and robustness, the computational efficiency of the models was also evaluated using the Rosenbrock function with 50 experimental and 5,000 simulation data points as an example. In low-dimensional settings such as 2-D, hybrid model 2 with PR and hybrid model 3 exhibit very low CPU time due to the simplicity of polynomial regression and the efficiency of the optimization process ($\approx$ 0.04 s for both models). Neural networks, on the other hand, require considerably longer training times even in these small dimensions (107 s for hybrid model 1 and 65 s for hybrid model 2 with ANN without considering the CPU time for hyperparameter tuning). However,

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

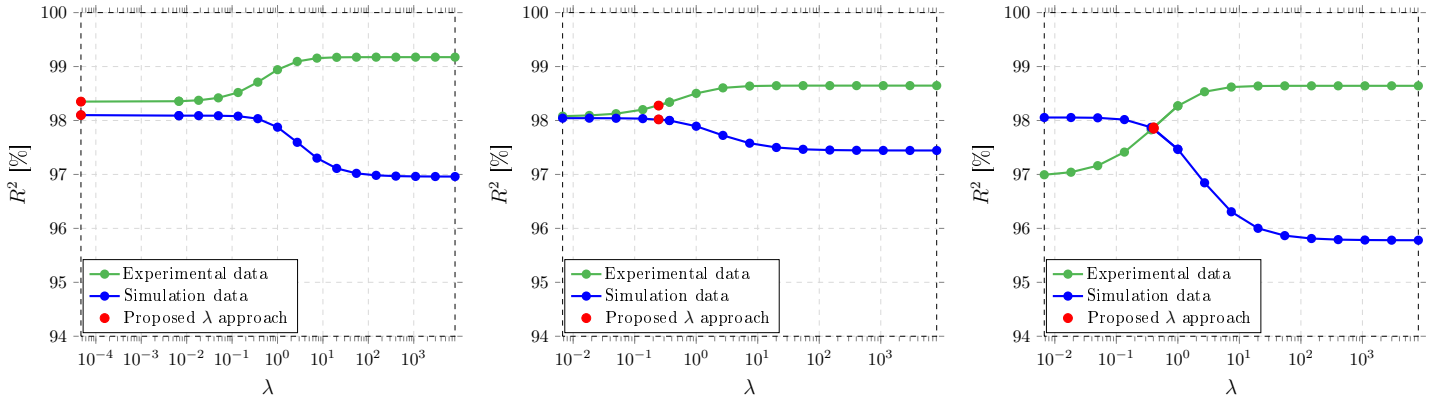


FIGURE 7: R^2 VALUES OF THE HYBRID MODEL 3 EVALUATED ON SIMULATION AND EXPERIMENTAL DATASETS AS A FUNCTION OF THE WEIGHTING PARAMETER λ, EXEMPLIFIED FOR THE 2-D ROSENBROCK FUNCTION WITH AN ERROR OF 5. EACH SUBPLOT CORRESPONDS TO A DIFFERENT DATA DISTRIBUTION SCENARIO: (LEFT) 50 EXPERIMENTAL VS. 500,000 SIMULATION POINTS, (MIDDLE) 5,000 EXPERIMENTAL VS. 5,000 SIMULATION POINTS, AND (RIGHT) 500,000 EXPERIMENTAL VS. 50 SIMULATION POINTS. THE RED POINT INDICATES THE λ VALUE COMPUTED USING THE PROPOSED FORMULA (SEE EQ. (6)).

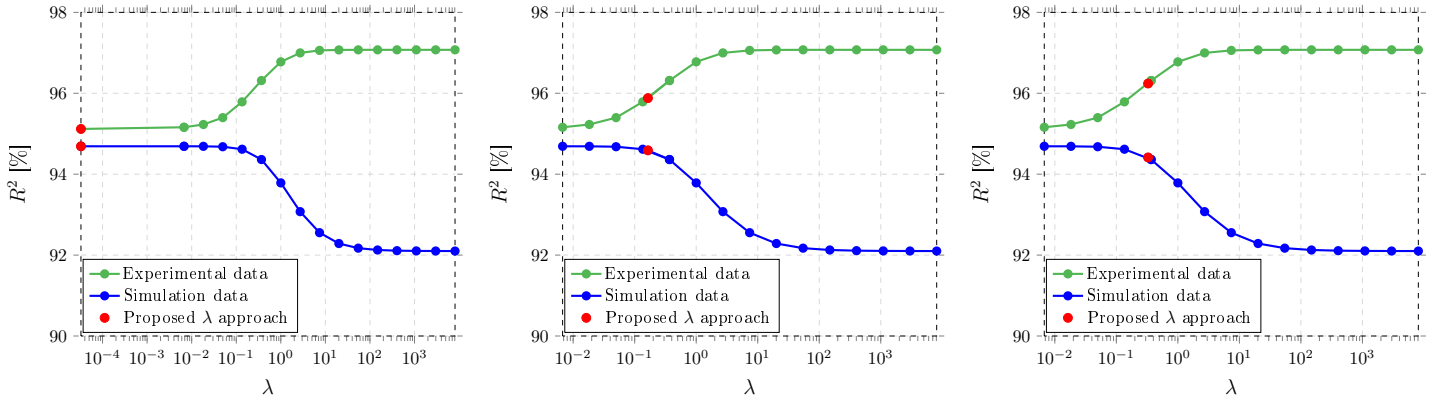


FIGURE 8: R^2 VALUES OF THE HYBRID MODEL 3 ON SIMULATION AND EXPERIMENTAL DATASETS AS A FUNCTION OF THE WEIGHTING PARAMETER λ, EXEMPLIFIED FOR THE 2-D GOLDSTEIN-PRICE FUNCTION WITH AN ERROR OF 5. EACH SUBPLOT CORRESPONDS TO A DIFFERENT DATA DISTRIBUTION SCENARIO: (LEFT) 50 EXPERIMENTAL VS. 500,000 SIMULATION POINTS, (MIDDLE) 5,000 EXPERIMENTAL VS. 5,000 SIMULATION POINTS, AND (RIGHT) 500,000 EXPERIMENTAL VS. 50 SIMULATION POINTS. THE RED POINT INDICATES THE λ VALUE COMPUTED USING THE PROPOSED FORMULA (SEE EQ. (6)).

TABLE 1: COMPARISON OF R^2 FOR PR TRAINED SOLELY ON SIMULATION DATA, PR TRAINED SOLELY ON EXPERIMENTAL DATA, AND THE HYBRID MODEL 3. THE MODELS ARE EVALUATED ON SIMULATION AND EXPERIMENTAL DATASETS FOR THE 20-D GOLDSTEIN-PRICE FUNCTION WITH 50 EXPERIMENTAL AND 5,000 SIMULATION DATA POINTS (NOISE=5).

Models	R^2 (exp)	R^2 (sim)
PR (sim)	85.3 %	95.8 %
PR (exp)	100.0 %	0.0 %
Hybrid Model 3	91.3 %	95.8 %

TABLE 2: THE SAME AS TABLE 1, BUT WITH 500 EXPERIMENTAL AND 5,000 SIMULATION DATA POINTS.

Models	R^2 (exp)	R^2 (sim)
PR (sim)	89.4 %	95.8 %
PR (exp)	100.0 %	43.3 %
Hybrid Model 3	93.7 %	95.8 %

as the dimensionality increases, particularly beyond 10-D (e.g., 20-D), the CPU time of hybrid model 3 increases significantly and can exceed that of the ANN-based models (without considering the CPU time for hyperparameter tuning in the ANN-based models), leading to a CPU time of 598 s, while hybrid model 1 and hybrid model 2 with ANN and PR require 149 s, 115 s, and 15 s, respectively. This is primarily due to the added complexity of optimizing the weighted cost function using the BFGS algorithm. While this model offers strong accuracy and robustness, its computational scalability may be improved in future work by exploring more efficient optimizers, implementing parallelization, or leveraging more powerful computing hardware. For the ANN-based models, the CPU time does not increase as significantly with dimensionality as it does for Hybrid Model 3. This is

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

mainly due to the use of early stopping, which limits the number of training iterations once convergence is reached.

Another important aspect is how the hybrid models with self-adaptive weights perform compared to cases where the weights are manually fine-tuned as hyperparameters. This comparison is illustrated in Figs. 7 and 8 for the Rosenbrock and Goldstein–Price function, respectively. Each plot shows the performance of hybrid model 3 on both the experimental and simulation test sets across a range of manually selected weight values, alongside the performance using the self-adaptive weight computed via Eq. (6). The three subplots represent different data availability scenarios: 50 experimental and 500,000 simulation data points (left), 5,000 experimental and 5,000 simulation data points (middle), and 500,000 experimental and 50 simulation data points (right). In all cases, the self-adaptive weighting approach leads to balanced and consistent performance between the two test sets. It avoids regions in which the performance diverges significantly—particularly where over-reliance on either the simulation or experimental data would degrade generalization. The adaptive method dynamically adjusts its trust in each data source based on both variance and sample size. A similar pattern of behavior is also observed in hybrid model 2 when using adaptive weighting. However, these results are not shown here for brevity.

Tables 1 and 2 further illustrate the advantages of hybrid modeling by comparing the performance of hybrid model 3 against traditional PR models trained solely on either simulation or experimental data. The results, shown for the 20-D Goldstein-Price function at a noise level of 5, clearly highlight the limitations of using only one data source: while PR trained on simulation data performs well on the simulation test set, it generalizes poorly to experimental data; conversely, PR trained solely on experimental data overfits, achieving perfect accuracy on the experimental set but failing on the simulation set. In contrast, hybrid model 3 achieves consistently high R^2 scores on both test sets, even when the number of experimental samples is small (Table 1). With more experimental data (Table 2), its accuracy further improves, underscoring its ability to effectively combine information from both sources.

4. HIGH-TEMPERATURE HEAT PUMP

In this section, we briefly introduce the high-temperature heat pump considered in this study and present the results of the hybrid modeling techniques applied to this specific use case.

4.1. Description

This work investigates a high-temperature heat pump (HTHP) operating on a closed-loop Brayton cycle using air as the working fluid [11]. The system, located at a test facility in Cottbus, Germany, is designed to reach supply temperatures up to 250 °C and consists of a combination of compression, heat exchange, expansion, and recuperation stages. The core components include a two-stage radial compressor, a high-temperature heat exchanger (HTHX), a recuperator for internal heat recovery, a turbine, and a low-temperature heat exchanger (LTHX). Flow control is managed by a three-way valve, which directs the fluid through the recuperator as needed, and a turbine bypass valve, which adjusts the resistance within the loop, allowing control over the operating points of the compressors.

The process begins as air is compressed, raising its temperature. Heat is then extracted in the HTHX and then partially used for internal heat recovery in the recuperator before the fluid expands in the turbine, reducing both pressure and temperature. The cooled air subsequently absorbs heat in the LTHX before being preheated again in the recuperator and returning to the compressor. This configuration allows for efficient thermal energy transfer and makes the setup suitable for industrial applications requiring high supply temperatures.

4.2. Results

In this section, we compare the performance of the hybrid surrogate modeling approaches for the HTHP system introduced above. For each hybrid model, two separate surrogate models are developed: one predicts the outlet temperature on the sink side, which represents the temperature available for industrial applications, based on five input variables: source-side inlet mass flow, source-side inlet temperature, sink-side inlet mass flow, sink-side inlet temperature, and motor speed. The other model predicts the electric power consumption based on these five input variables. These two quantities—the outlet temperature and electric power consumption—are also the output variables considered in the surrogate-based optimization presented in [2]. A total number of 26 experimental data points are available from the test facility, while the simulation data are generated using a detailed model in Modelica 4.0.0. To generate the simulation dataset, a one-at-a-time sensitivity approach is used: for each input variable, its value is systematically varied across a defined range while keeping the remaining four inputs fixed. This procedure is repeated for each input, enabling the characterization of the system's response to individual input variations and the capture of nonlinear dependencies in the output variables. The impact of data availability on the model performance is assessed by varying the number of simulation data points: 50, 500, 5,000, and 25,000. As with the test function study, all input and output data are normalized to the range [0, 1] for consistency and to ensure comparability across different model types. For hybrid model 2 with PR and hybrid model 3, a third-degree PR model is trained for the outlet temperature and the electric power consumption.

Figures 9 and 10 show the R^2 values for the outlet temperature and electric power prediction, respectively, evaluated for different numbers of simulation data points. Each model is tested both on simulation data and on the available experimental data. As seen in both figures, hybrid model 3 consistently achieves high prediction accuracy on the simulation test set, even with 50 simulation samples. This highlights the benefit of its integrated training strategy, where both data sources are used simultaneously through a unified weighted cost function. In contrast, the other models require more simulation data to reach similar levels of accuracy. This also happens with hybrid model 2, which relies on two separately trained models, and if one of them (e.g., the experimental-based surrogate) performs poorly (see Table 3, the overall hybrid prediction also suffers.

When tested on the experimental dataset, hybrid model 3 again achieves high accuracy across all simulation data sizes in

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

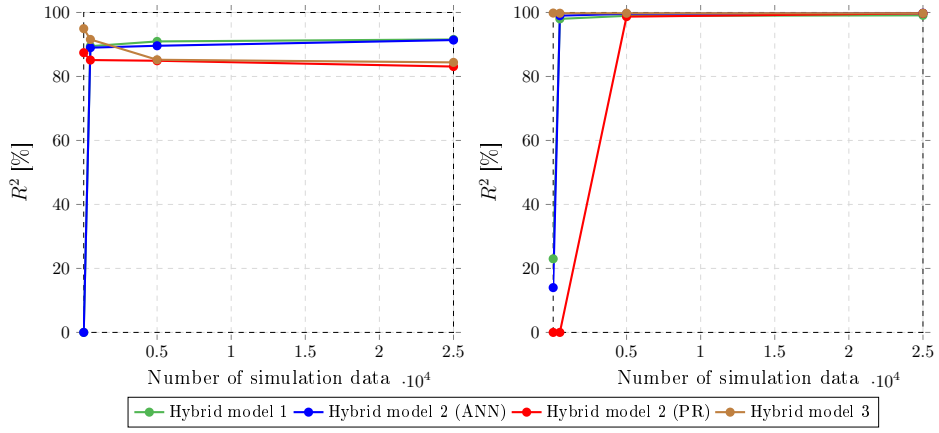


FIGURE 9: R^2 VALUES OF THE HYBRID MODELS FOR PREDICTING THE OUTLET TEMPERATURE OF THE HTHP, EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF SIMULATION POINTS.

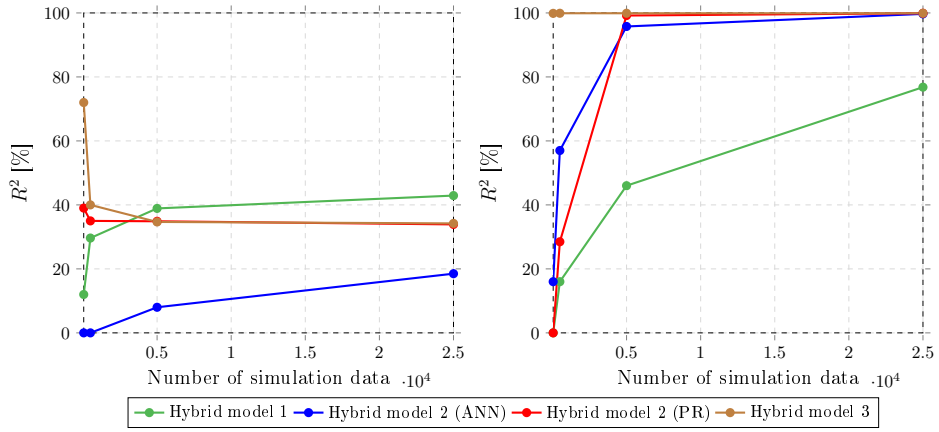


FIGURE 10: R^2 VALUES OF THE HYBRID MODELS FOR PREDICTING THE ELECTRIC POWER CONSUMPTION OF THE HTHP, EVALUATED ON EXPERIMENTAL (LEFT) AND SIMULATION (RIGHT) DATASETS AS A FUNCTION OF THE NUMBER OF SIMULATION POINTS.

TABLE 3: COMPARISON OF R^2 FOR PR TRAINED SOLELY ON SIMULATION DATA, PR TRAINED SOLELY ON EXPERIMENTAL DATA, AND THE HYBRID MODEL 3. THE MODELS ARE EVALUATED ON SIMULATION AND EXPERIMENTAL DATASETS FOR THE OUTLET TEMPERATURE PREDICTION OF THE HTHP WITH 26 EXPERIMENTAL AND 50 SIMULATION DATA POINTS.

Models	R^2 (exp)	R^2 (sim)
PR (sim)	83.2 %	99.8 %
PR (exp)	100.0 %	0.0 %
Hybrid Model 3	95.0 %	99.8 %

Fig. 9 on the left, underlining its robustness and ability to generalize from the available data. The ANN-based models (hybrid model 1 and hybrid model 2 with ANN) show improved accuracy as the number of simulation samples increases, reflecting their higher data requirements and flexibility in learning from large datasets. Hybrid model 2 with with PR demonstrates a performance close to that of hybrid model 3 on the experimental test set. While its accuracy on the simulation test set is low for small data sizes, its accuracy on the experimental test set is higher because both underlying surrogate models (for simulation and experiment) are well trained in this setting. This contrast illustrates that the separate model training in hybrid model 2 can be particularly vulnerable when either surrogate underperforms. A similar trend can be observed for the power prediction in Fig. 10, but with lower accuracy. This happens because the simulation model was calibrated in [11] based on the outlet temperature but not on the electric power consumption, resulting in a better match between simulated and experimental temperature data. When evaluating hybrid model 2 with PR and hybrid model 3 on the experimental dataset, the highest prediction accuracy is observed at a simulation dataset size of 50 points. Beyond this, the accuracy, especially for electric power prediction in hybrid model 3, declines. This drop is due to the self-adaptive weighting in the unified loss function, which increases the influence of simulation data as more of it becomes available. Since discrepancies exist between the experimental and simulation datasets, the model becomes less accurate on the experimental data as it increasingly prioritizes the simulation data.

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026

Regarding computational cost, hybrid model 3 starts with a CPU time of 0.38 s with 50 simulation data points and increases to 10 s with 25,000 simulation data points. In comparison, hybrid model 1, hybrid model 2 with ANN, and hybrid model 3 with PR require CPU training times between 43 s and 110 s, 25 s and 75 s (without considering the CPU time for hyperparameter tuning in the ANN-based models), and 0.02 s and 0.30 s, respectively. As with the test functions, hybrid model 2 with PR consistently achieves the lowest CPU time across all conditions. Although the CPU times in the heat pump results fall within the expected range, they can be reduced, as discussed in Section 3.2, and this reduction will be explored in future work.

5. CONCLUSION

This work presents a hybrid modeling framework that effectively combines simulation and experimental data to improve predictive performance in complex engineering systems. Within this framework, we introduce and compare three hybrid modeling approaches: (1) a transfer-learning-based ANN model, (2) a two-step weighted surrogate model, and (3) a novel surrogate model trained via a unified weighted cost function. For the second approach, both black-box (ANN) and white-box (PR) surrogates are considered, allowing an investigation of interpretability and training effort.

The performance of the proposed approaches has been tested using the Rosenbrock and Goldstein-Price test functions under varying conditions of dimensionality, data quantity, and noise, as well as on the HTHP system under varying amounts of simulation data. Across all settings, hybrid model 3 consistently achieves the best balance between accuracy, robustness, and generalization, particularly in higher dimensions or when simulation and experimental data differ. Its integrated training and adaptive data weighting enable dynamic adjustment of trust in each data source. Even with a small number of simulation data, hybrid model 3 demonstrates high accuracy, outperforming other models. While hybrid model 2 with PR performs well in many cases, hybrid model 3's unified loss function offers superior performance by considering both data types during training. In contrast, ANN-based models require larger datasets to reach high performance, especially in higher dimensions, making them less practical in data-scarce settings. Regarding computational cost, hybrid model 3 shows increased CPU time in higher-dimensional spaces (>=15-D) but remains time-efficient in lower dimensions and in cases with low number of data points.

In future work, other black-box models such as Gaussian Processes could be explored as alternatives to ANN. Moreover, the optimization of the weighted cost function in hybrid model 3, currently performed using the BFGS algorithm, could be replaced by alternative optimization methods to investigate their impact on the model accuracy, robustness, and CPU time. Last, the presented methodology can be applied to other real-world systems to further validate its generalizability and practical relevance.

ACKNOWLEDGMENTS

The authors declare no competing interests. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] Lédée, François, Crawford, Curran and Evins, Ralph. "Improved surrogate modeling for multi-energy system design: Model architecture, sampling and scaling choices." *Applied Energy* Vol. 390 (2025): pp. 1–20. DOI https://doi.org/10.1016/j.apenergy.2025.125812.

[2] Kyriakidis, Loukas, Mendez, Miguel Alfonso and Bähr, Martin. "A hybrid algorithm based on Bayesian optimization and Interior Point OPTimizer for optimal operation of energy conversion systems." *Energy* Vol. 312 (2024): pp. 1–10. DOI https://doi.org/10.1016/j.energy.2024.133416.

[3] Wu, Yuandi, Sicard, Brett and Gadsden, Stephen Andrew. "Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring." *Expert Systems with Applications* Vol. 255 (2024): pp. 1–40. DOI https://doi.org/10.1016/j.eswa.2024.124678.

[4] McBride, Kevin and Sundmacher, Kai. "Overview of Surrogate Modeling in Chemical Process Engineering." *Chemie Ingenieur Technik* Vol. 91 (2019): pp. 228–239. DOI https://doi.org/10.1002/cite.201800091.

[5] Cozad, Alison, Sahinidis, Nikolaos V. and Miller, David C. "Learning surrogate models for simulation-based optimization." *AIChE Journal* Vol. 60 (2014): pp. 2211–2227. DOI https://doi.org/10.1002/aic.14418.

[6] Letham, Benjamin, Karrer, Brian, Ottoni, Guilherme and Bakshy, Eytan. "Constrained Bayesian Optimization with Noisy Experiments." *Bayesian Analysis* Vol. 14 No. 2 (2019): pp. 495–519. DOI 10.1214/18-BA1110.

[7] Humbird, K. D., Peterson, J. L., Salmonson, J. and Spears, B. K. "Cognitive simulation models for inertial confinement fusion: Combining simulation and experimental data." *Physics of Plasmas* Vol. 28 (2021): pp. 1–8. DOI 10.1063/5.0041907.

[8] Raissi, M., Perdikaris, P. and Karniadakis, G. E. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational Physics* Vol. 378 (2019): pp. 686–707. DOI https://doi.org/10.1016/j.jcp.2018.10.045.

[9] Wikipedia contributors. "Test Functions for Optimization." (2025). URL https://en.wikipedia.org/wiki/Test_functions_for_optimization. Accessed: 13/05/2025.

[10] Hao, Jia, Ye, Wenbin, Jia, Liangyue, Wang, Guoxin and Allen, Janet. "Building surrogate models for engineering problems by integrating limited simulation data and monotonic engineering knowledge." *Advanced Engineering Informatics* Vol. 49 (2021): pp. 1–15. DOI https://doi.org/10.1016/j.aei.2021.101342.

[11] Tran, A. Phong and Stathopoulos, Panagiotis. "Dynamic simulation and experimental validation of a high-temperature Brayton heat pump." *Applied Thermal Engineering* (2025): pp. 1–42DOI https://doi.org/10.1016/j.applthermaleng.2025.126536.

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Downloaded from http://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2025/89374/V006T09A075/7595077/v006t09a075-imece2025-167746.pdf by Deutsches Zentrum für Luft- und Raumfahrt (DLR) user on 07 August 2026