Neurocomputing 491 (2022) 464–488

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Structure optimization of prior-knowledge-guided neural networks

Mohamed Atwya *, George Panoutsos

Department of Automatic Control and Systems Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

ARTICLE INFO

Article history:
Received 17 March 2021
Revised 15 November 2021
Accepted 4 March 2022
Available online 8 March 2022

Keywords:
Prior-guided neural network
Machine learning
Constrained optimization
Structure optimization

ABSTRACT

Prior-knowledge use in neural networks, for example, knowledge of a physical system, allows network training to be tailored to specific problems. Literature shows that prior-knowledge in neural network training enhances predictive performance. Research to date focuses on parametric optimization rather than structure optimization. We present a new framework to optimize the structure of a neural network using prior-knowledge. This is achieved through optimizing the number of hidden units via a line search and cross-validation using the empirical error to eliminate data-set/model-structure application dependency for prior-knowledge guided neural networks. In addition to using the prior-knowledge in the model training step, we propose utilizing the prior errors as part of the cross-validation performance index to improve generalization. Results demonstrate that the proposed training framework enhances the model's prediction accuracy and prior-knowledge consistency for convex data sets with a unique minimum and non-convex multi-modal data sets. The presented results yield a new understanding of physics-guided neural networks in terms of their structural and parametric optimization.

1. Introduction

Prior-knowledge is typically not exploited in machine learning frameworks when modeling physical systems. Prior information, such as first principle physical laws and empirically supported relationships, could guide a learning algorithm towards a more theory-consistent solution. Model consistency to any underlying physical laws can improve the model's generalization ability, especially when only a few training examples are available, which is a common challenge when dealing with problems where data generation is very resource-intensive.

Using prior information in model training is not a new challenge in the literature. In 2017, Karpatne et al. [1] classified theory-informed data-driven models into the following five categories: (1) choosing a suitable model design, (2) refining data model outputs, (3) augmenting theory-based models [2], (4) hybrid models [3,4], and (5) Navigating the search space of candidate models. Navigating the search space of candidate models can be influenced by prior-knowledge using four different approaches: (1) model parameter initialization, (2) probabilistic models, (3) theory-based regularization, and (4) constrained optimization [1].

Theory-guided parameter initialization methods such as the matrix completion algorithm and pre-training using computational simulation data can help guide machine learning algorithms towards generalizable and theory-consistent models [5,6]. Probabilistic models in the context of theory-guided data-driven modeling involve including theory-guided data distributions as Bayesian priors [7]. Theory-guided probabilistic models have shown success in applications such as predicting subsurface aquifers connectivity and electrical activity within the heart wall [8,9]. Theory-guided data-driven modeling regularization methods include using and developing new Lasso variants such as the sparse group Lasso to impose theory-specific structures on the parameters of DS models [10,11].

Constrained optimization methods include (1) developing constrained optimization methods for PDEs/non-linear transformations and (2) integrating theory-based constraints with existing optimization techniques. In this work, for the first time in the literature, we focus on integrating theory-based constraints within optimization methods to navigate the search space of candidate feedforward neural network models.

Note that in the literature of theory-guided machine learning, several terminologies are used to refer to the application-dependent domain knowledge utilized (theory, prior, physics, etc.). In theory-based constrained optimization, the methods developed can be applied to any domain knowledge that can be expressed in the form of a constraint. In this work, we choose the term prior-knowledge-guided neural networks.

In [12], the authors construct physics-guided neural networks with a hybrid-physics-data model that uses a physics-based model's simulated output as an additional feature when training the

* Corresponding author.
E-mail address: Matwya1@sheffield.ac.uk (M. Atwya).

https://doi.org/10.1016/j.neucom.2022.03.008
0925-2312/

neural network (NN). The authors introduce a method to add two forms of physics relationships, equality and inequality constraints, to the loss function. They use the mean square error (MSE) for the empirical loss, a rectified linear unit (ReLU) function for the physics-consistency loss, and elastic net regularization for model complexity control. The model structure was set to 3 hidden layers with 12 hidden units each. Using labeled and unlabeled data, they demonstrate the prediction error and physics-consistency error improvements of the physics-guided NN on a spatio-temporal problem.

Similar to [12], the authors in [13] also add equality and inequity physics-based constraints to the NN loss function and use $\mathscr{L}_2$ regularization. The network architecture consists of two hidden layers, with the first layer having 64 hidden units and the second layer having 128 units. They set the $\mathscr{L}_2$ regularization weight value as $\lambda_R = 1$ and empirically optimize the physics-based constraint loss weight λ_p. The authors use a $60-20-20$ split to train, validate, and test the models. The results show empirical error percentage improvements on noise-free and noisy synthetic (Bohachevsky function) and real data sets.

In [6], Jia et al. use recurrent NNs (RNN) and advance the work of [12] to estimate the spatio-temporal lake temperature. Their first contribution is the use of theory-based simulation data to pre-train the RNN. They also propose running a theory-based energy balance model in parallel with the RNN. If the temporal regression of the temperature violates the energy balance constraint, a penalty is added to the loss function. They perform sensitivity studies on the impact of the physics-based constraint weights by varying one weight and keeping the second weight constant.

Wu et al. [14] demonstrated how a novel statistical covariance-based constraint in the optimization loss function of deep convolutional generative adversarial networks (DCGANs) can improve training stability and convergence properties. They demonstrated the performance improvement of their GANs-based physical system emulator on the Rayleigh–Benard convection.

Liu and Wang [15] developed a multi-fidelity physics-constrained NN (MF-PCNN) and proposed an adaptive weighting scheme for regularization. They apply their method to PDEs from two-dimensional heat transfer, phase transition, and dendritic growth problems. The authors add a physics-equation-based constraint and initial condition (IC) and boundary condition (BC) constraints to the cost function. They successfully train several MF-PCNN variants with four layers but with different hidden unit numbers identified by conducting sensitivity studies. The authors express that their method requires searching and sampling procedures to find the optimal architecture and set it as future work. Even though their proposed method does not use model complexity control (e.g., $\mathscr{L}_1$ and $\mathscr{L}_2$ regularization), they claim that the sensitivity studies avoid over-fitting. However, in the results, an increase in the MSE of one of the MF-PCNN variants is attributed to over-fitting. Their proposed adaptive weighting scheme results in higher prediction accuracy and shorter training time than standard NNs.

In [16], the authors compute data-driven solutions to PDEs using physics-guided neural networks (PGNN). Their method involves training two NNs in two steps. Step one is to train a 9-layer NN with 20 units per layer to model a data set extracted from the initial and boundary conditions. Step two is to compute the partial derivatives of the first NN, substitute the partial derivatives into the PDE to define the second 9-layer NN with 20 units per layer, and train it on collocation points. They demonstrate this method on Burger's equation with Dirichlet boundary conditions in one space dimension.

The more recent work in [17] proposes a new PGNN using the system's variational energy as the loss function combined with the transfer learning method. The authors argue that the proposed loss function is easier to minimize, and hence, the proposed PGNN performs better relative to PGNNs with residual loss functions such as in [16,18,19]. They demonstrate the effectiveness of the proposed method on six fracture mechanics problems. The neural network had three hidden layers in the six fracture mechanics problems, each consisting of 50 neurons.

In [20], the authors propose incorporating prior-knowledge in neural network models using genetic algorithms for a metallurgy application. The authors train and test a vanilla NN on a continuous cooling transformation data set with seven inputs and six outputs. They find that the resulting NN does not conform to metallurgical engineering prior-knowledge. Accordingly, they propose a genetic-algorithm-based multi-objective cost function constrained via metallurgy prior-knowledge. The vanilla NN and the proposed NN models have one hidden layer with twenty neurons. The proposed model improves the mean-square error and almost completely removes the metallurgy prior-knowledge inconsistencies studied.

Jagtapa et al. [21] introduce an optimize-able adaptive hyperparameter that changes the slope of the activation function to tune the PGNN architecture. They show that the proposed adaptive activation function results in a faster loss function decay and a smaller $\mathscr{L}_2$ regularization error. Even though the proposed method is tuneable for any number of hidden layers, the authors choose different numbers of hidden layers and units for each case study with no reasoning. A more comprehensive study would include a sensitivity study on the effects of the number of hidden layers, number of hidden units, and training data set size on the empirical performance of the proposed method.

The work in [22] proposes a Probabilistic Physics-guided Neural Network (PPgNN) with a novel model architecture to extend the standard neural network approach for fatigue data analysis. Among other methods, the proposed PPgNN utilizes a custom loss function, physics-constrained loss optimization, and a custom partially connected neural network architecture. Their proposed PPgNN variant has two hidden layers. Hidden layer one consists of 20 hidden units using the tanh activation function. Hidden layer two consists of two units; unit one uses the linear activation function, and unit two uses the exponential linear unit activation function. The output layer has one neuron, which uses the linear activation function. The model's final outputs are the output layer neuron and the second hidden unit of the second hidden layer. The proposed PPgNN is tested on fatigue data analysis data and is more consistent with the domain knowledge relative to a neural network without physics guidance. With further validation, the authors claim that the proposed framework is not limited to fatigue data analysis and can be applied to other survival data analyses via adjusting the constraints and the network architecture according to physics knowledge.

In [23], the authors propose novel methods to integrate semantic knowledge into the learning step to improve the accuracy of convolutional neural networks (CNNs) on image classification tasks. The work presents a novel use of knowledge distillation, collective classification, training phases, and hyperparameter cross-validation for prior guided CNNs. They use cross-validation as a heuristic hyperparameter tuning method to tune the penalty of each predicate. The authors test different modeling case studies, data set sizes, and CNN architectures. The data set size ablation study shows that the smaller the training data set size, the more significant the reduction of the empirical error of the proposed methods. The work in [23] has no investigation on the random

model weight initialization effects on the empirical error mean and variance.

The number of hidden units dictates the NN/PGNN model's complexity and is commonly between 5 and 100 [24]. The number of units should be sufficiently large to capture the data non-linearity but not too large to avoid over-fitting and maintain the model's smoothness between the in-sample training data points. Typically, the number of hidden units is chosen via application-specific expert knowledge or empirical experimentation. The work in [12,6,16,13,25], among several other works, either sets the number of hidden units based on expert knowledge or reports utilizing sensitivity studies without providing a systematic framework or a statistical analysis on the empirical and prior-consistency performance. The authors in [16] note that the interplay between the physics-guided NN architecture/training procedure and the complexity of the modeling data is poorly understood and propose a Bayesian approach to monitor the variance of the predictive posterior distribution as possible future work.

The authors in [15] note that their method requires systematic searching and sampling procedures to find the optimal physics-guided NN architecture and set it as future work. The work in [16] states that designing the correct physics-guided NN architecture is an open research question and confirms observing that a specific physics-guided NN architecture that yields accurate results for one data set can fail for another. Therefore, there is a need for a systematic framework along with statistical analysis.

Another gap in the PGNN literature is a lack of statistical analysis:

- There is a lack of statistical analysis on the mean and variance of the empirical and prior-consistency errors across different random weight initialization of PGNN models.
- There is a lack of investigation with statistical tests on the effect of the data set size on the mean of the empirical and prior-consistency errors.
- There is a lack of investigation with statistical tests on the correlation between the empirical and prior-consistency errors.
- There is a lack of investigation with statistical tests on the effect of the loss weight hyperparameters on the mean of the empirical and prior-consistency errors.

For more detailed literature and surveys on PGNN gaps and methods of incorporating prior-knowledge into machine learning, we direct the reader to [26–28]. In this work, we summarize the literature review by highlighting the need for (1) a systematic framework to set the number of hidden units for PGNNs, (2) statistical analysis on the performance of PGNNs under varying settings, and (3) methods to verify if and when a PGNN model can be trusted. In this work, we aim to address the gaps discussed via the following contributions towards prior-knowledge guided feedforward neural networks (PGNNs):

1. Model structure optimization via the Hidden Unit Number: A framework to optimize the number of hidden units via a line search and cross-validation using the empirical error to eliminate data-set/model-structure application dependency for prior-knowledge guided feedforward neural networks (model $PGNN_1$).
2. Model structure optimization via Prior-knowledge-guided cross-validation: In addition to using the prior-knowledge in the model training step, we propose utilizing the prior errors as part of the cross-validation performance index (Eq. 3). Incorporating the prior-consistency error along with the empirical error in the validation process improves generalizability (model $PGNN_2$).
3. Statistical study on the empirical and prior-consistency error performance of PGNNs under the effects of random model weight initialization, data set size, hyperparameters, and the correlation between the empirical prior-consistency errors.

The remainder of the paper is organized as follows. The Methodology section provides the problem formulation and methods contributions 1 and 2. The Evaluation section covers the methods used to test and benchmark the proposed methods against reference models. The Results and Analysis Sections introduce the benchmark testing data, demonstrate the empirical and physics-consistency errors of the NN/PGNN models, and discuss the results. The conclusion covers the proposed framework's main benefits, limitations, and future work.

2. Methodology

This work, utilizes prior-knowledge constraint losses in the cost function of a multi-layer perceptron feedforward neural network model. The first PGNN model, $PGNN_1$, differs from the NN model by incorporating the prior-based losses in the weight optimization function. This work's contributions are: (1) $PGNN_1$ differs from the literature on PGNN models by optimizing the number of hidden units, and (2) $PGNN_2$ incorporates the same methods in $PGNN_1$ but additionally incorporates the prior error in the cross-validation process.

In the following Subsections, the problem formulation and preliminaries are provided in SubSection 2.1. SubSections 2.3 and 2.2 will discuss the methodologies of contributions 1 and 2, respectively. Finally, SubSection 3.1 provides the model performance testing methods.

2.1. Preliminaries

A typical NN optimization criterion typically has two loss functions: an empirical loss function $\mathscr{L}_e$ and a regularization loss function $\mathscr{L}_r$ (Eq. 1). To incorporate prior-knowledge, we add the prior-knowledge-based loss function $\mathscr{L}_p$ to the optimization criterion as in literature (Eq. 2) [15,13,21]. The regularization and prior-based loss functions have the weights $\rho_r \geqslant 0$ and $\rho_p \geqslant 0$, respectively (Eq. 1 and 2). Accordingly, the framework proposed leads to three hyper-parameters: regularization loss weight (ρ_r), prior-based loss weights (ρ_p), and the hidden units number (j).

$$\mathscr{L}_{NN} = \mathscr{L}_e + \rho_r \mathscr{L}_r \tag{1}$$

where $\mathscr{L}_e$ is the empirical loss, $\mathscr{L}_r$ is the regularization loss, and ρ_r is the regularization loss weight.

$$\mathscr{L}_{PGNN} = \mathscr{L}_e + \rho_r \mathscr{L}_r + \rho_p \mathscr{L}_p, \tag{2}$$

where $\mathscr{L}_p$ is the prior loss and ρ_p is the prior-based loss weight.

The cost function is non-linear and multi-modal (Eq. 2). Therefore, the network's initial random weights determine the cost function solution (a local minimum), and a training algorithm does not guarantee to find the global minimum. Furthermore, if the training data set is randomly sampled from an available data set, then the training data set might adversely be biased (i.e., not a representative of the physics phenomenon of study). Therefore, data models must be trained and validated with reliable empirical evidence. It is also necessary to ensure that the performance differences between any proposed methods and reference models are due to the proposed methods and not effects such as biased data sets and random weight initialization. Accordingly, to validate the models, repeated k-fold cross-validation (CV) is applied for the following reasons: (1) make efficient use of the data to do training

and validation, (2) reduce the chance of bias and variance in the validation performance index, and (3) address the local-minimum problem. The literature's CV performance index for vanilla NNs and PGNN models has been the empirical error. The number of CV folds was chosen to be $k = 5$ to reduce the variance in the validation performance index. Finally, for every set of m values (number of hidden units) and random initial weights, the 5-fold CV is repeated 3 times using re-divided data subsets to minimize the occurrence of misleading results due to bias-dominated data subsets.

2.2. Model structure optimization via the hidden unit number

In contribution (1), we train the PGNN model across a linearly spaced vector of 8 values within $[30, 100]$ to estimate the best number of hidden unit values m^* value based on the empirical error. A second search is then performed with a finer vector of 10 linearly spaced points around the best-estimated value $m^* \pm 10$. The model ($PGNN_1$) with the best cross-validation empirical error is chosen. Optimizing the number of hidden units via a line search eliminates the data-set/model-structure application dependency for prior-knowledge-guided feedforward neural networks.

2.3. Model structure optimization via prior-knowledge-guided cross-validation

In contribution (2), we propose utilizing the prior error as part of the cross-validation performance index to optimize the number of hidden units ($PGNN_2$, Eq. 3). Note that the prior-based errors $\mathscr{E}_p$ are application dependent and would therefore have different means and standard deviations per application. If the empirical error ($\mathscr{E}_e$) and the prior-based errors differ by orders of magnitude, then the cross-validation error function can be dominated by either errors. Accordingly, the empirical and prior errors in Eq. 3 have been re-scaled via min–max normalization.

$$\underline{\mathscr{E}} = \frac{\underline{\mathscr{E}}_e - min(\underline{\mathscr{E}}_e)}{max(\underline{\mathscr{E}}_e) - min(\underline{\mathscr{E}}_e)} + \frac{\underline{\mathscr{E}}_p - min(\underline{\mathscr{E}}_p)}{max(\underline{\mathscr{E}}_p) - min(\underline{\mathscr{E}}_p)}. \tag{3}$$

3. Evaluation

The half sum squared error (HSSE) metric was chosen for the empirical loss function in Eq. 1 and 2. The regularization loss function was selected as $\mathscr{L}_2$ regularization to penalize large weights in the model and reduce data over-fitting.

$$\mathscr{L}_e = \frac{1}{2}\sum_{i=1}^{n}\left(\underline{\hat{y}}^{(i)} - \underline{y}^{(i)}\right)^2, \tag{4}$$

where n is the number of data points and $\hat{y}_i$ and y_i are the prediction and target for the k^{th} input vector, respectively.

$$\mathscr{L}_r = \frac{1}{2}\underline{w}^T\underline{w}, \tag{5}$$

where $\underline{w}$ is a vector of the model weights.

We chose to use the boundary condition constraint (scenario 1) in Eq. 6 as the prior loss function to demonstrate the proposed methods in the main result in SubSection 4.1. We also investigate using the initial condition constraint (Scenario 2) and a horizontal symmetry constraint (Scenario 3) as prior loss functions (Eq. 7 and 8) to assess the impact of using different priors on the proposed methods (SubSection 4.2). The prior losses are applied separately to the reference and proposed models, where scenarios 1 to 3 are the boundary, initial, and symmetry prior, respectively. To quantify the prior-knowledge violations, we utilize the HSSE metric (Eq. 6, 7 8).

$$\mathscr{L}_p = \frac{1}{2}\sum_{i=1}^{p}\left(\underline{\hat{y}}^{(i)}\left(\underline{X}_b^{(i)}\right) - \underline{y}^{(i)}\left(\underline{X}_b^{(i)}\right)\right)^2, \tag{6}$$

where p is the number of samples in the boundary dataset.

$$\mathscr{L}_p = \frac{1}{2}\sum_{i=1}^{1}\left(\hat{y}^{(i)}(0,0) - y^{(i)}(0,0)\right)^2. \tag{7}$$

$$\mathscr{L}_p = \frac{1}{2}\sum_{i=1}^{n}\left(\underline{\hat{y}}^{(i)}(-\underline{x}_1, \underline{x}_2) - \underline{y}^{(i)}(\underline{x}_1, \underline{x}_2)\right)^2. \tag{8}$$

The $\mathscr{L}_2$ regularization loss weight (ρ_r) and the prior loss weight (ρ_p) dictate the importance of model smoothness (complexity) and prior-consistency, respectively. Since ρ_r and ρ_p are scalars, it is possible to train NN/PGNNs with different combinations of ρ values and test the NN/PGNNs on a validation data set to identify the best model (empirical error wise). A typical NN example of complexity control is a one-grid search to find the optimal regularization weight (ρ_r), which results in the best PI. Similarly, a two-grid search is used to find the optimal function width and regularization weight of a Gaussian radial basis function model. State-of-the-art methods involve online adaptive methods to update the weights (ρ_r and ρ_p) during training. Optimizing the weights is outside the scope of this work; therefore, we chose to set them as constants for the main result and perform a hyperparameter sensitivity study in SubSections 4.1 and 4.6, respectively.

The scaled conjugate gradient (SCG) back-propagation method was used to minimize the loss function (Eq. 1 and 2) and find the optimal weights. Note that each input feature from each case study was transformed separately by z-score standardization (i.e., zero mean and unit standard deviation). The optimization algorithm has two stopping criteria; the precision of the objective function and the weights at the solution. The two criteria were set to 10^{-4}. The precision in the line search parameter space was set to Netlab's default value (10^{-3}) [29]. The optimization algorithm was allowed a maximum of 1000 iterations to find the solution.

The root mean square error (RMSE) metric was chosen for the empirical and prior-based CV error functions (Eq. (9)–(12)). Note that the summation stop values in Eq. 9, 10, and 12 are divided by k to take into account the CV folds.

$$\mathscr{E}_e = \sqrt{\frac{\sum_{i=1}^{n/k}\left(\underline{\hat{y}}^{(i)} - \underline{y}^{(i)}\right)^2}{n/k}}. \tag{9}$$

$$\mathscr{E}_p = \sqrt{\frac{\sum_{i=1}^{p/k}\left(\underline{\hat{y}}^{(i)}(\underline{X}_b^{(i)}) - \underline{y}^{(i)}(\underline{X}_b^{(i)})\right)^2}{p/k}}. \tag{10}$$

$$\mathscr{E}_p = \sqrt{\left(\hat{y}(0,0) - y(0,0)\right)^2}. \tag{11}$$

$$\mathscr{E}_p = \sqrt{\frac{\sum_{i=1}^{n/k}\left(\underline{\hat{y}}^{(i)}\left(-\underline{x}_1^{(i)}, \underline{x}_2^{(i)}\right) - \underline{y}^{(i)}\left(\underline{x}_1^{(i)}, \underline{x}_2^{(i)}\right)\right)^2}{n/k}}. \tag{12}$$

The reference and proposed models consist of two input neurons ($d = 2$), a hidden layer with a non-linear ReLu activation function, and an output layer with one neuron ($c = 1$) and a linear activation function. Below is a list of the differences between the reference and proposed models.

Reference Models:

- NN−100: A vanilla NN with 100 hidden units (no prior usage in the loss or validation performance index).

- PGNN−30: A PGNN with 30 hidden units (no prior usage in the validation performance index).
- PGNN−65: A PGNN with 65 hidden units (no prior usage in the validation performance index).
- PGNN−100: A PGNN with 100 hidden units (no prior usage in the validation performance index).

Proposed Models:

- $PGNN_1$: A PGNN network where the number of hidden units is optimized via a line search as detailed in SubSection 2.2 (no prior usage in the validation function).
- $PGNN_2$: A PGNN network where the number of hidden units is optimized via a line search (SubSection 2.2) and the prior error is utilized in the validation performance index (SubSection 2.3).

3.1. Model testing

Following training and validation, an out-of-sample (OOS) empirical data set with 1000 data points was used to test the models RMSE empirical performance (Eq. 9 with $k = 1$). The OOS empirical data set was also utilized to test the symmetry prior-consistency RMSE (Eq. 12 with $k = 1$). Similarly, an out-of-sample (OOS) boundary data set with 4000 data points was used to test the models RMSE empirical performance ((Eq. 10 with $k = 1$). The initial prior-consistency RMSE performance was tested via Eq. 11.

The OOS data sets were reserved and only used to test the models after they were trained and validated. Note that each PGNN/NN model was trained and validated 24 times using randomly initialized weights, and the averaged results are provided in SubSection 4.

4. Results and analysis

4.1. Benchmark function testing

To test the proposed methods' empirical and prior-consistency error performance, we test 14 benchmark case studies varying from convex functions with a unique minimum to non-convex multi-modal functions (Table A). Further, we analyze why the proposed methods improve the empirical and prior-consistency error performance for some functions but not others and what dictates the level of improvement. For each modeling case study, a boundary data set $[X_b, y(X_b)]$ of $p = 4000$ uniformly distributed data points was used for the prior-knowledge as discussed in SubSection 2.1. Tables 1 and 2 show the empirical and prior-consistency RMSE mean and standard deviation using a data set of 1000 samples.The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. The median and $[0.25, 0.75]$ quantiles of the optimal number of hidden units selected by the proposed methods are shown in Table 7.

The percentage change in the empirical and prior-consistency RMSEs between the proposed methods and the reference models are provided in Tables 3 and 4. The cells highlighted in Tables 3 and 4 indicate a statistically significant percentage change ($P< 0.05$) based on an independent two-sample multivariate Hotelling's test (Appendix B Tables B.11 and B.12). Similarly, the percentage change in the variance of the empirical and prior-consistency RMSEs between the proposed methods and the reference models is provided in Tables 5 and 6. The cells highlighted in Tables 5 and 6 indicate a statistically significant percentage change ($P< 0.05$) based on a two-sample F-test for equal variance (Appendix B Tables B.13, B.14, B.15, and B.16).

The first result to note is the lack of any statistically significant difference in the empirical and prior-consistency error means and variances between the vanilla NN and PGNN−100 for the 14 benchmark functions (Tables 3, 5, and 6). The vanilla NN and PGNN−100 performance show that given an identical network architecture, a PGNN model may not out-perform a vanilla NN if the loss function weight hyperparameters are not tuned. There is also no statistically significant difference between $PGNN_1$ and PGNN2 empirical and prior-consistency mean errors except for the Ackley function (Table 4). However, the effect of the different number of hidden units between $PGNN_1$ and $PGNN_2$ is seen in the performance improvements of $PGNN_1$ and $PGNN_2$ relative to the reference models.

By optimizing the network architecture, $PGNN_1$ outperformed (empirical and prior-consistency error-mean-wise) the reference NN and PGNN-100 models for 3 and 6 functions, respectively (Tables 3 and 4). $PGNN_1$ reduced the empirical error variance relative to the reference NN and PGNN-100 for 6 and 3 functions, respectively (Table 5). $PGNN_1$ also reduced the prior-consistency error variance relative to the reference NN and PGNN-100 for 4 and 3 functions, respectively (Table 5).

For the McCormick function, $PGNN_1$ degraded the percentage change in the empirical and prior-consistency errors variance relative to NN−100 and PGNN−100. The performance degradation in $PGNN_1$ with the McCormick function can be associated with the hidden unit results in Table 7 since $PGNN_1$ had the largest quantile range $[0.25, 0.75]$ of 39 hidden units. A larger hidden unit quantile range is likely to correlate with the variance in the empirical and prior-consistency errors positively. The reason for $PGNN_1$'s performance with the McCormick function is not clear, but it is likely to

Table 1
Empirical RMSE mean and standard deviation range across 14 benchmark functions using a training data set of 1000 samples, a regularization weight of $1e-3$ for the NN model, and a regularization and a prior weights of $1e-5$ and $1e-6$ for PGNN models, respectively.

Benchmark Function	NN−100	PGNN−100	$PGNN_1$	$PGNN_2$
McCormick	0.11 ± 0.01	0.11 ± 0.02	0.11 ± 0.04	0.10 ± 0.01
Griewank	0.14 ± 0.03	0.15 ± 0.03	0.13 ± 0.01	0.13 ± 0.01
Ackley	0.59 ± 0.00	0.59 ± 0.01	0.58 ± 0.01	0.58 ± 0.01
Branin	0.88 ± 0.13	0.88 ± 0.16	0.84 ± 0.14	0.88 ± 0.14
Styblinski-Tang	0.90 ± 0.24	0.92 ± 0.22	0.93 ± 0.22	0.93 ± 0.24
Booth	1.36 ± 0.12	1.41 ± 0.41	1.30 ± 0.11	1.31 ± 0.12
Three-hump Camel	2.13 ± 0.63	2.01 ± 0.43	1.82 ± 0.30	1.87 ± 0.31
Bukin N.6	2.17 ± 0.12	2.21 ± 0.17	2.16 ± 0.14	2.16 ± 0.14
Himmelblau's	3.57 ± 1.74	3.72 ± 1.71	2.95 ± 0.38	3.18 ± 0.71
Zakharov	15.11 ± 5.06	14.76 ± 3.07	13.11 ± 2.48	13.90 ± 3.00
Bohachevsky	21.11 ± 3.67	21.74 ± 4.08	19.06 ± 2.96	19.21 ± 2.85
Rosenbrock	620.93 ± 279.05	644.92 ± 274.50	487.82 ± 136.64	538.26 ± 215.55
Beale	1252.28 ± 313.79	1250.56 ± 359.89	1146.71 ± 286.44	1086.34 ± 252.09
Goldstein-Price	19110.69 ± 6111.54	18938.97 ± 5574.39	15712.85 ± 3940.58	15916.60 ± 3574.84

Table 2
Boundary prior-consistency RMSE mean and standard deviation across 14 benchmark functions using a training data set of 1000 samples, a regularization weight of $1e-3$ for the NN model, and a regularization and a prior weights of $1e-5$ and $1e-6$ for the PGNN models, respectively.

Benchmark Function	NN−100	PGNN−100	$PGNN_1$	$PGNN_2$
McCormick	0.21 ± 0.03	0.23 ± 0.04	0.22 ± 0.09	0.20 ± 0.02
Griewank	0.31 ± 0.05	0.32 ± 0.05	0.28 ± 0.03	0.27 ± 0.04
Ackley	1.01 ± 0.02	1.00 ± 0.02	0.99 ± 0.02	0.97 ± 0.02
Branin	2.50 ± 0.47	2.50 ± 0.51	2.33 ± 0.54	2.31 ± 0.48
Styblinski-Tang	2.91 ± 0.75	2.85 ± 0.81	2.99 ± 0.83	2.78 ± 0.74
Booth	2.24 ± 0.31	2.35 ± 0.25	2.19 ± 0.21	2.18 ± 0.19
Three-hump Camel	12.16 ± 3.48	11.08 ± 2.72	10.34 ± 1.78	10.19 ± 1.80
Bukin N.6	2.94 ± 0.21	2.99 ± 0.27	2.90 ± 0.19	2.90 ± 0.23
Himmelblau's	9.40 ± 2.59	9.68 ± 2.38	8.17 ± 1.05	8.41 ± 1.32
Zakharov	42.21 ± 15.65	42.04 ± 13.16	35.87 ± 6.48	36.36 ± 7.83
Bohachevsky	40.42 ± 8.17	40.57 ± 7.29	34.73 ± 7.50	33.73 ± 7.91
Rosenbrock	1992.64 ± 548.91	2056.07 ± 600.60	1869.55 ± 403.32	1921.96 ± 451.53
Beale	4262.21 ± 831.87	4339.32 ± 1025.90	4020.72 ± 851.43	3843.76 ± 761.80
Goldstein-Price	66484.35 ± 16126.71	67140.18 ± 13743.65	59491.66 ± 11529.53	60214.83 ± 10469.91

Table 3
Percentage change in the empirical and prior-consistency RMSEs for the 14 benchmark functions comparing the proposed methods to the NN reference model using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$ for the NN model. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$, respectively. Shading indicates a statistically significant difference between the values (critical P value 0.05).

Benchmark	[NN,PGNN−100]		[NN,$PGNN_1$]		[NN,$PGNN_2$]	
Function	Empirical	Prior	Empirical	Prior	Empirical	Prior
McCormick	1.54	4.78	2.74	2.11	−5.63	−7.89
Griewank	1.73	2.55	−12.09	−11.41	−11.84	−12.17
Ackley	−0.07	−0.45	−0.44	−1.86	−0.37	−3.93
Branin	0.09	0.02	−4.56	−6.96	−0.31	−7.88
Styblinski–Tang	1.46	−1.84	2.41	2.96	3.10	−4.51
Booth	3.48	5.00	−4.37	−2.39	−3.77	−2.64
Three-hump Camel	−5.80	−8.85	−14.62	−14.91	−12.11	−16.16
Bukin N.6	1.75	1.74	−0.44	−1.32	−0.75	−1.09
Himmelblau's	4.41	3.02	−17.40	−13.10	−10.7	−10.51
Zakharov	−2.36	−0.41	−13.26	−15.03	−8.04	−14.10
Bohachevsky	2.96	0.38	−9.73	−14.08	−9.04	−16.54
Rosenbrock	3.86	3.18	−21.44	−6.18	−13.31	−3.55
Beale	−0.14	1.81	−8.43	−5.67	−13.25	−9.82
Goldstein–Price	−0.90	0.99	−17.78	−10.52	−16.71	−9.43

do with the data set size relative to the complexity of the function and the untuned hyperparameters.

$PGNN_2$ outperformed (empirical and prior-consistency error-wise) the reference NN and PGNN-100 models for 6 and 8 functions, respectively (Tables 3 and 4). $PGNN_2$ reduced the empirical error variance relative to the reference NN and PGNN-100 for 5 and 3 functions, respectively (Table 5). $PGNN_2$ also reduced the prior-consistency error variance relative to the reference NN and PGNN-100 for 5 and 3 functions, respectively (Table 5).

The discrepancy in the error improvements between $PGNN_1$ and $PGNN_2$ is due to $PGNN_2$ validating the prior-consistency of the model in addition to its empirical generalizability. In Table 7, $PGNN_2$ has resulted in a smaller, larger, and equal number of hidden units median for 7, 5, and 2 functions. We also note that $PGNN_2$ has a larger hidden unit quantile range $[0.25, 0.75]$ than $PGNN_1$ for 12 out of the 14 benchmark functions. By incorporating the prior-consistency error in the validation performance index, $PGNN_2$ re-enforces prior-consistency. $PGNN_2$ outperforms (empirical/prior error mean and variance) $PGNN_1$ relative to the reference models by selecting a suitable number of hidden units to avoid under-fitting or over-fitting depending on the training/prior data landscape versus the data set size.

The discrepancy in the error improvements of the proposed methods between the benchmark functions is likely to be related to (1) discrepancy in the value of the prior-knowledge for each benchmark function, (2) training data set size, (3) prior data set size. The training data set size effects are examined in SubSection 4.2. The prior data set size effects are not examined and are left for future work.

A possible explanation for the discrepancy in the error improvements is that the prior-knowledge has a different value for each benchmark function depending on the training/prior data landscape. For example, the boundary prior is likely to be less valuable if, for a given 3-dimensional function, the target values across a line on the 2-dimensional landscape are highly non-linear and non-monotonic. We propose a simple measure of boundary/training mean output ratio to give insight into the value of the boundary prior. The prior output mean, training output mean, and boundary/training output mean ratio for each benchmark function are shown in Table 8). Using the proposed measure, we measure

Table 4
Percentage change in the empirical and prior-consistency RMSEs for the 14 benchmark functions comparing the proposed methods to thePGNN reference models using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$, respectively. Shading indicates a statistically significant difference between the values (critical P value 0.05).

Benchmark	[PGNN−100,PGNN$_1$]		[PGNN−100,PGNN$_2$]		[PGNN$_1$,PGNN$_2$]	
Function	Empirical	Prior	Empirical	Prior	Empirical	Prior
McCormick	1.17	−2.55	−7.06	−12.09	−8.14	−9.79
Griewank	−13.58	−13.61	−13.34	−14.36	0.28	−0.86
Ackley	−0.37	−1.42	−0.30	−3.50	0.07	−2.11
Branin	−4.65	−6.98	−0.40	−7.90	4.45	−0.99
Styblinski–Tang	0.93	4.89	1.61	−2.72	0.67	−7.25
Booth	−7.59	−7.04	−7.01	−7.28	0.63	−0.26
Three-hump Camel	−9.36	−6.65	−6.70	−8.02	2.94	−1.47
Bukin N.6	−2.15	−3.02	−2.46	−2.78	−0.31	0.24
Himmelblau's	−20.89	−15.64	−14.54	−13.14	8.03	2.97
Zakharov	−11.17	−14.68	−5.82	−13.75	6.02	1.09
Bohachevsky	−12.33	−14.40	−11.65	−16.86	0.77	−2.86
Rosenbrock	−24.36	−9.07	−16.54	−6.52	10.34	2.80
Beale	−8.30	−7.34	−13.13	−11.42	−5.26	−4.40
Goldstein–Price	−17.03	−11.39	−15.96	−10.31	1.30	1.22

Table 5
Percentage change in the variance of the empirical RMSE across the model variants for the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$ for the PGNN models, respectively. Shading indicates a statistically significant difference between the values (critical P value 0.05).

Benchmark	Percentage change (%)					
Function	[NN−100, PGNN−100]	[NN−100, PGNN$_1$]	[NN−100, PGNN$_2$]	[PGNN−100, PGNN$_1$]	[PGNN−100, PGNN$_2$]	[PGNN$_1$, PGNN$_2$]
McCormick	46.31	951.52	−20.86	618.71	−45.91	−92.47
Griewank	−6.32	−73.56	−70.02	−71.77	−68.00	13.36
Ackley	83.12	26.99	10.48	−30.65	−39.67	−13.00
Branin	40.35	6.99	5.81	−23.77	−24.61	−1.10
Styblinski–Tang	14.35	−16.23	−0.31	−2.19	16.40	19.00
Booth	31.33	−19.60	−2.87	−38.78	−26.04	20.81
Three-hump Camel	−54.14	−76.99	−75.72	−49.83	−47.07	5.50
Bukin N.6	106.06	29.75	31.38	−37.03	−36.24	1.25
Himmelblau's	−3.36	−95.20	−83.63	−95.03	−83.06	241.16
Zakharov	−35.13	−75.87	−64.91	−62.81	−45.91	45.45
Bohachevsky	23.80	−34.73	−39.50	−47.27	−51.13	−7.31
Rosenbrock	−3.23	−76.02	−40.33	−75.22	−38.34	148.86
Beale	31.54	−16.67	−35.46	−36.65	−50.93	−22.55
Goldstein–Price	−16.81	−58.43	−65.79	−50.03	−58.87	−17.70

Spearman's correlation and the statistical significance between the empirical and prior-consistency RMSE percentage change and the boundary/training mean output ratio (Table 9). From Table 9, the boundary/training mean output ratio increases, the percentage change between the NN and the proposed methods (PGNN$_1$ and PGNN$_2$) decreases with a strong negative correlation ($\rho < -0.60$). Similarly, as the boundary/training mean output ratio increases, the percentage change between the PGNN−100 and the proposed methods decreases with a moderate negative correlation ($\rho < -0.40$). The negative correlation indicates that the prior is more valuable (empirical and prior-consistency error-wise) when the prior target and training data-target means differ more. This result does not rule out the influence of other factors in the impact of PGNN$_1$ and PGNN$_2$ on the empirical and prior-consistency error performance relative to the reference models.

This study has shown that optimizing the network architecture (PGNN$_1$) statistically significantly improves the empirical and prior-consistency RMSE means relative to the reference models but can result in a larger empirical and prior-consistency variance across randomly initialized model weights. Using the prior-knowledge and the empirical error in the validation process, PGNN$_2$, results in a more informed performance index for model candidate selection and statistically significantly improves the empirical and prior-consistency RMSE mean and variance relative to the reference models. PGNN$_2$ statistically significantly improves the empirical and prior-consistency RMSE mean and variance for more benchmark functions than PGNN$_1$ relative to the reference models. Finally, The results invite further research on (1) the relationship between the empirical and prior data in the context of physics guided loss and validation PI functions, (2) training data

Table 6
Percentage change in the variance of the boundary prior-consistency RMSE across the model variants for the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$, respectively. Shading indicates a statistically significant difference between the values (critical P value 0.05).

Benchmark Function	Percentage change (%)					
	[NN−100, PGNN−100]	[NN−100, $PGNN_1$]	[NN−100, $PGNN_2$]	[PGNN−100, $PGNN_1$]	[PGNN−100, $PGNN_2$]	[$PGNN_1$, $PGNN_2$]
McCormick	23.88	660.26	−49.92	513.69	−59.58	−93.41
Griewank	−10.40	−59.27	−46.13	−54.54	−39.88	32.25
Ackley	1.67	11.52	−30.41	9.68	−31.55	−37.59
Branin	21.93	36.66	8.42	12.08	−11.08	−20.67
Styblinski–Tang	15.57	19.48	−3.31	3.39	−16.34	−19.08
Booth	−35.58	−52.19	−60.70	−25.79	−38.99	−17.79
Three-hump Camel	−39.24	−74.01	−73.29	−57.22	−56.04	2.75
Bukin N.6	63.33	−21.47	14.05	−51.92	−30.17	45.23
Himmelblau's	−15.62	−83.69	−73.86	−80.67	−69.03	60.21
Zakharov	−29.31	−82.85	−74.98	−75.74	−64.60	45.93
Bohachevsky	−20.26	−15.66	−6.08	5.77	17.79	11.36
Rosenbrock	19.72	−46.01	−32.34	−54.90	−43.48	25.33
Beale	52.09	4.76	−16.14	−31.12	−44.86	−19.95
Goldstein–Price	−27.37	−48.89	−57.85	−29.62	−41.97	−17.54

Table 7
The median and $[0.25, 0.75]$ qauntiles of the optimal number of hidden units selected by the proposed methods across the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$, respectively.

Benchmark Function	$PGNN_1$ Median (quantile)		$PGNN_2$ Median (quantile)	
McCormick	80.00	(59.00 – 98.00)	83.00	(60.00 – 100.00)
Griewank	56.50	(42.00 – 70.00)	51.50	(38.50 – 66.50)
Ackley	77.00	(62.00 – 91.50)	64.00	(46.00 – 80.00)
Branin	96.00	(86.00 – 104.00)	84.50	(73.00 – 98.50)
Styblinski-Tang	95.50	(86.50 – 106.00)	97.00	(89.00 – 107.00)
Booth	102.00	(99.00 – 107.00)	105.00	(101.00 – 110.00)
Three-hump Camel	98.50	(86.00 – 103.00)	90.00	(81.50 – 100.00)
Bukin N.6	100.00	(90.50 – 105.00)	100.50	(93.00 – 103.00)
Himmelblau's	103.00	(94.50 – 110.00)	101.00	(92.00 – 107.00)
Zakharov	103.00	(94.00 – 108.50)	95.00	(83.00 – 101.50)
Bohachevsky	97.00	(90.00 – 100.50)	97.00	(88.00 – 105.00)
Rosenbrock	92.00	(90.00 – 96.00)	90.50	(87.00 – 98.00)
Beale	90.00	(78.50 – 100.00)	90.00	(77.00 – 100.00)
Goldstein-Price	92.50	(69.00 – 99.00)	94.00	(70.00 – 100.50)

Table 8
The mean of the boundary data and the training data of the 14 benchmark functions.

Benchmark Function	Boundary output mean	Training output mean	Boundary/training output mean ratio
McCormick	13.79	7.63	1.81
Griewank	0.94	1.00	0.94
Ackley	12.05	9.68	1.24
Branin	88.15	54.93	1.60
Styblinski-Tang	108.12	−7.59	−14.24
Booth	747.39	389.24	1.92
Three-hump Camel	1139.11	255.57	4.46
Bukin N.6	149.14	121.38	1.23
Himmelblau's	357.00	137.34	2.60
Zakharov	8221.47	3695.84	2.22
Bohachevsky	199945.39	9940.28	2.01
Rosenbrock	316950.15	125540.62	2.52
Beale	41572.18	8614.62	4.83
Goldstein-Price	197803.51	50959.70	3.88

landscape versus prior-knowledge value, and (3) adaptive-weighting of the prior-knowledge in PGNNs.

4.2. Data set size multivariate sensitivity study

This study aims to investigate if the data set size affects the proposed methods' multivariate empirical and prior-consistency RMSE performance. To the best of the authors' knowledge, this is the first study to quantitatively examine the effect of the data set size on the multivariate empirical and prior-consistency error variance for PGNNs. Five training data sets were created for the Bohachevsky and the Ackley functions; sample sizes: $200, 600, 1000, 2000$, and 4000. The samples in each data set are uniformly distributed and bounded by the input ranges in Table A.10. We set a regularization weight of $1e-3$ for the NN model and regularization and a prior weight of $1e-5$ and $1e-6$ for the PGNN models. Three prior case studies were tested per benchmark function: (1) boundary prior, (2) initial prior, and (3) symmetry prior. The empirical and prior-consistency RMSEs are presented in Appendix C Tables C.17 and C.18.

We perform a one-way Multivariate Analysis of Variance (MANOVA) for comparing the multivariate means of the empirical and prior-consistency errors, grouped by data set size. The dimension result estimates the dimension of the space containing the group

Table 9
Spearman's rho correlation and P value between the percentage change in RMSEs and the boundary/training mean output ratio across the 14 benchmark functions using a training data set of 1000 samples.The model hyperparameters for this test include a regularization weight of $1e-3$ for the NN model, and a regularization and a prior weights of $1e-5$ and $1e-6$ for the PGNN models, respectively. P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

	$\rho(P)$					
	[NN−100, PGNN−100]	[NN−100, $PGNN_1$]	[NN−100, $PGNN_2$]	[PGNN−100, $PGNN_1$]	[PGNN−100, $PGNN_2$]	[$PGNN_1$, $PGNN_2$]
Empirical	−0.24 (0.40)	−0.67 (0.01)	−0.76 (0.00)	−0.54 (0.05*)	−0.57 (0.03)	0.35 (0.23)
Prior	0.09 (0.77)	−0.53 (0.06)	−0.42 (0.14)	−0.47 (0.09)	−0.34 (0.24)	0.28 (0.33)

means. The P values test whether the means lie in the space of dimensions 0, 1, or 2 (critical P value 0.05). The null hypothesis is that the multivariate means of all data set sizes are equal, and any difference observed in the multivariate means is random.

From Table C.19, $PGNN_1$ and $PGNN_2$ MANOVA results for the Bohachevsky case study show a 1-dimensional statistically significant difference ($P < 0.05$) in the multivariate means for the three prior scenarios. However, the multivariate means may lie on the same line. From Table C.20, $PGNN_1$ and $PGNN_2$ MANOVA results for the Ackley case study show a 2-dimensional statistically significant difference ($P < 0.05$) in the multivariate means for the three prior scenarios. The multivariate means may lie on the same plane in 2-dimensional space but not on the same line.

Qualitatively, the canonical analysis of the Bohachevsky case for $PGNN_1$ in Fig. C.1a, C.2a, and C.3a shows that: (a) data set size 200 is separate from and does not overlap the other data set sizes, (b) data set size 600, 1000, and 2000 are overlapping, but with distinct centers, (c) data set sizes 2000 and 4000 are not separable, (d) the first canonical variable approximately separates data set sizes 200, 600, 1000, and 2000, and (e) the second canonical variable, does not show a clear separation between the data set sizes. The canonical analysis of the Bohachevsky case for $PGNN_2$ in Fig. C.4a, C.5a, and C.6a approximately has the same results as $PGNN_1$.

Quantitatively, the Mahalanobis distance matrix of the Bohachevsky case for $PGNN_1$ and $PGNN_2$ in Fig. C.1b, C.2b, C.3b, C.4b, C.5b, and C.6b confirms the canonical analysis findings and demonstrates that the distance between the 2000 and 4000 is negligible relative to the distances between other data set sizes. To assess if the univariate distance between data set sizes is statistically significant, we perform a one-way analysis of variance (ANOVA) and a multiple comparison test in SubSection D.

Qualitatively, the canonical analysis of the Ackley case for $PGNN_1$ in Fig. C.7a, C.8a, and C.9a shows that: (a) all data set sizes overlap with one or more other data set sizes, but each has a distinct center, (d) the first canonical variable approximately separates the data set sizes into three groups (1) 200, (2) 600 and 1000, and (3) 2000 and 4000, and (e) the second canonical variable does not show a clear separation between the data set sizes. The canonical analysis of the Ackley case for $PGNN_2$ in Fig. C.10a, C.11a, and C.12 approximately has the same results as the Canonical analysis for $PGNN_1$.

Quantitatively, the Mahalanobis distance matrix of the Ackley case for $PGNN_1$ and $PGNN_2$ in Fig. C.7(b), C.8(b), C.9(b), C.10(b), C.11(b), and C.12(b) confirms the canonical analysis findings and demonstrates that the data set sizes can be separated into three groups (1) 200, (2) 600 and 1000, and (3) 2000 and 4000. Note that the distance between the 2000 and 4000 data set sizes is smaller than the distance between the 600 and 1000 data set sizes. To assess if the univariate distance between data set sizes is statistically significant, we perform a one-way ANOVA and a multiple comparison test in SubSection D.

The MANOVA study shows a 1-dimensional statistically significant difference ($P < 0.05$) in the multivariate empirical and prior-consistency means for the three prior scenarios in the Bohachevsky case study. However, the Ackley case study has a 2-dimensional statistically significant difference ($P < 0.05$) in the multivariate means. The change in the dimensional difference of the multivariate means between the Bohachevsky and Ackley functions is likely due to (1) the functions' complexity and (2) the value of the prior per function.

A prior is likely to significantly impact the empirical and prior-consistency errors as the function complexity increases with a fixed data set size. The value of a prior in a given case study depends on (1) the training data set size and (2) the correlation between the empirical and the prior-consistency errors. If the data set size is small relative to the function complexity, then a prior is likely to be more useful in training and validating PGNNs, and vice versa. Similarly, if there is a strong correlation between the empirical and the prior-consistency errors, then a prior is likely to be more useful in training and validating PGNNs. In SubSection 4.4, we report and analyze the correlation between the empirical and the prior-consistency errors for the Bohachevsky and the Ackley functions.

For the Bohachevsky and Ackley functions, the canonical analysis and Mahalanobis distance matrix for the proposed methods with all three priors separated the 200 data set size results from the other data sets for $PGNN_1$ and $PGNN_2$. However, the 2000 and 4000 data set sizes were not separable for both functions. The separation of the 200 data set size and lack of separation between the 2000 and 4000 data set sizes are in line with previous studies and show that as the data set size increases, the impact on the empirical and prior-consistency errors reduces.

The canonical analysis and Mahalanobis distance matrix for the proposed methods with all three priors approximately separated the multivariate errors of the 600 and 1000 data set sizes for the Bohachevsky, but not the Ackley function. A possible explanation for the difference in multivariate error variance between the Bohachevsky and Ackley functions for the 600 and 1000 sample sizes is that the Ackley function is more complex than the Bohachevsky function. Since the Ackley function is more complex, smaller changes in the data set size are likely to impact the multivariate error performance less.

We conclude that the data set size can result in a 1-dimensional or 2-dimensional statistically significant difference ($P < 0.05$) in the multivariate empirical and prior-consistency means for convex, and non-convex modeling problems up to a given data set size. The functions' complexity, the prior value, and the correlation between the empirical and the prior-consistency errors dictate the magnitude of the multivariate difference between data set sizes.

4.3. Data set size univariate sensitivity study

This study examines if and how the data set size and the modeling problem complexity affect the proposed methods' effect on the univariate empirical and prior-consistency RMSE performance. We perform an ANOVA test (omitted but available upon request) followed by a multiple comparison test using Tukey's honestly sig-

nificant difference procedure to determine which data set sizes statistically differ ($P > 0.05$, Table D.21). The ANOVA and multiple comparison tests are performed on the Bohachevsky, and the Ackley function results in SubSection 4.2.

For the Bohachevsky case study, the proposed methods ($PGNN_1$ and $PGNN_2$) have a statistically significant impact on the empirical RMSE mean for the three prior scenarios up to a data set size of 600 samples. The prior-consistency error means for the initial and symmetry priors are statistically significantly different up to a data set size of 600 samples. The boundary prior-consistency error mean is statistically significantly different up to a data set size of 1000 samples.

For the Ackley case study, the proposed methods ($PGNN_1$ and $PGNN_2$) have a statistically significant impact on the empirical RMSE means for all the prior scenarios up to a data set size of 4000 samples. The proposed methods ($PGNN_1$ and $PGNN_2$) also statistically significantly impact the prior-consistency RMSE means for the initial prior scenario up to a data set size of 4000 samples. For the boundary and symmetry prior scenarios, the proposed methods have a statistically significant impact on the prior-consistency RMSE means for data set sizes up to 2000 samples (Table D.21). Now that we have confirmed that the proposed methods' have a statistically significant impact on the empirical and prior-consistency RMSEs for some data set sizes, we will assess if there is a correlation between the data set size and the empirical/prior-consistency RMSEs.

The ANOVA study and multiple comparison test results align with previous studies; as the data set size increases, PGNN models are less likely to impact the empirical and prior-consistency errors. The proposed methods ($PGNN_1$ and $PGNN_2$) have a statistically significant impact on the empirical RMSE mean for the three prior scenarios up to a data set size of 600 samples for the Bohachevsky function and up to 4000 samples for the Ackley function. Similarly, the prior-consistency error means for the Bohachevsky function do not statistically differ for data set sizes larger than 600 for the initial and symmetry priors and 1000 samples for the boundary prior. However, for the Ackley function, the prior-consistency error means do not statistically differ for data set sizes larger than 4000 for the initial prior and 2000 for the boundary and symmetry priors. The Bohachevsky function is convex with a unique minimum and therefore requires a smaller training data set size relative to the Ackley function, which is non-convex and multi-modal. The difference in complexity between the two functions explains why the empirical and prior-consistency error variance saturates at a smaller data set size for the Bohachevsky function.

For data set sizes 1000 and larger, $PGNN_1$ for the Bohachevsky function achieves the same empirical error mean across all the three prior scenarios, indicating that with the proposed framework, the number of hidden units set by $PGNN_1$ dictates the performance, instead of the prior type/value. $PGNN_1$ empirical performance with the Bohachevsky function for data set sizes 1000 and larger is also likely because of the use of a constant prior weight instead of a tuned/adaptive weight in the loss function (Eq. 2).

Note that the initial prior is less beneficial than the boundary and symmetry priors for the Bohachevsky function since the Bohachevsky function is convex. Hence, there is no statistically significant impact on the empirical and prior-consistency RMSE means for data sets larger than 600 samples.

We conclude this section by noting that the proposed methods ($PGNN_1$ and $PGNN_2$) affect the empirical and prior-consistency RMSEs depending on the modeling problem complexity and data set size. As the modeling problem complexity decreases and the data set size increases, the proposed methods are less likely to have a statistically significant impact ($P > 0.05$) on the empirical and prior-consistency RMSE.

4.4. Data set size and performance correlation study

The objective of this study is to determine the strength and direction of correlation between the data set size and the empirical and prior-consistency errors. We use Superman's rank correlation to measure the strength and direction of monotonic association between the data set size and the empirical RMSE in the results in SubSection 4.2. In Table E.22, we see a strong ($\rho > 0.60$) to very strong ($\rho > 0.80$) statistically significant ($P < 0.05$) negative correlation between the empirical RMSE and the data set size for the three prior scenarios of the Bohachevsky and Ackley case studies.

Similarly, we use Superman's rank correlation to measure the strength and direction of monotonic association between the data set size and the prior-consistency RMSE. From Table E.23, we see a very strong ($\rho > 0.80$) statistically significant ($P < 0.05$) negative correlation between the boundary and symmetry prior RMSE, and the data set size for the three prior scenarios of the Bohachevsky and Ackley case studies. For the initial prior, there is a weak ($\rho > 0.30$) statistically significant ($P < 0.05$) negative correlation for the Bohachevsky case study and a strong ($\rho > 0.60$) statistically significant ($P < 0.05$) negative correlation for the Ackley case study.

The statistically significant ($P < 0.05$) negative correlation between the data set size and the empirical and prior-consistency RMSE is in line with previous studies. The weaker correlation between the initial prior-consistency error and data set size for the Bohachevsky function further supports the results obtained in SubSection 4.3, which show that the initial prior is less beneficial than the boundary and symmetry priors for the Bohachevsky function since the function is convex.

4.5. Empirical and prior-consistency error correlation study

This study aims to determine the strength and direction of correlation between the empirical and prior-consistency errors across the data set sizes examined in this work. A prior may not have value or correlate to the empirical error where a sufficiently representative training data set is available. Where the prior is valuable, the relationship between the empirical and prior-consistency errors can be a positive linear or monotonic correlation. The empirical and prior-consistency errors may not be strongly correlated if the prior is not significantly informative relative to the training data (e.g., the initial prior scenario with the Bohachevsky function). We use Superman's rank correlation to measure the strength and direction of monotonic association between the empirical RMSE and the prior-consistency RMSEs from the data set size sensitivity study results in SubSection 4.2. Table F.24 shows the correlation and P value for the three prior case studies in the Bohachevsky and the Ackley functions.

In Table F.24, for the Bohachevsky function, the reference and proposed models have a very strong ($\rho > 0.8$) statistically significant ($P < 0.05$) positive correlation between the empirical RMSE and the prior-consistency RMSE for the boundary and symmetry prior scenarios. For the initial prior scenario with the Bohachevsky function, the proposed methods have a statistically significant ($P < 0.05$) weak ($[0.20, 0.39]$) to moderate ($[0.40, 0.59]$) positive correlation between the empirical RMSE and the prior-consistency RMSE. For the initial prior scenario with the Bohachevsky function, the reference models NN−100, PGNN−65, and PGNN−100 have a statistically significant ($P < 0.05$) weak ($[0.20, 0.39]$) positive correlation between the empirical RMSE and the initial prior-consistency RMSE. The reference model PGNN−30 has no statistically significant ($P < 0.05$) correlation between the empirical RMSE and the initial prior-consistency RMSE.

In Table F.24, for the Ackley function, the reference and proposed models have a very strong ($\rho > 0.8$) statistically significant

($P< 0.05$) positive correlation between the empirical RMSE and the prior-consistency RMSE for the boundary and symmetry prior scenarios. For the initial prior scenario with the Ackley function, the reference and proposed models have a statistically significant ($P< 0.05$) strong positive correlation ($\rho > 0.6$) between the empirical RMSE and the initial prior-consistency RMSE.

The correlation between the empirical RMSE and the prior-consistency RMSE for the Bohachevsky and Ackley functions mostly differ in the initial prior case study. The Bohachevsky function has a weak to moderate positive correlation, while the Ackley function has a strong positive correlation. The correlation difference for the initial prior between the two functions is consistent with the findings regarding the smaller benefit of the initial prior, especially for the Bohachevsky function discussed in SubSections 4.4 and 4.3.

The data set size and correlation between the empirical and the prior-consistency errors are also linked. For example, a boundary prior is likely more beneficial for small data sets if a strong symmetrical two-dimensional monotonic relation between the function's output and input data exists over the input data range up to the boundary condition (e.g., Bohachevsky, Ackley, or Griewank). On the other hand, the boundary prior can be less valuable where the data set size is small and the correlation between the empirical and prior-consistency errors is weak (e.g., Zakharov or Three-hump Camel). The association between the data set size and the value of the boundary prior accords with our earlier observations in SubSection 4.1, which showed better empirical/prior-consistency RMSE mean error improvements for the Bohachevsky, Ackley, and Griewank relative to the Zakharov and Three-hump Camel functions.

In summary, these results show a statistically significant ($P< 0.05$) positive correlation between the empirical RMSE and the prior-consistency RMSE for the three prior scenarios of the Bohachevsky and Ackley case studies. The positive correlation between the empirical RMSE and the prior-consistency RMSEs demonstrates the importance of developing prior-consistent NNs. The positive correlation between the empirical RMSE and the prior-consistency RMSEs further supports using the prior-consistency error in the validation performance index of PGNNs ($PGNN_2$).

4.6. Hyperparameter sensitivity investigation

The objective of this investigation is to determine if the regularization loss weight and the prior loss weight hyperparameters impact the RMSE performance of the proposed methods and the multivariate empirical and prior-consistency RMSE performance difference between the reference and the proposed models. To perform this investigation, we train the reference and proposed models with different combinations of empirical and prior-consistency weights and compare the multivariate empirical and prior-consistency RMSEs via a two-sample Hotelling's T^2 test for independent samples.

Throughout the results in this Subsection, the NN−100 reference model was trained with a regularization weight $\rho_r = 1e-3$. The PGNN−30, PGNN−65, PGNN−100, $PGNN_1$, and $PGNN_2$ models were trained with the following $[\rho_r, \rho_p]$ loss weight combinations: $[1e-3, 1e-4], [1e-3, 1e-6], [1e-3, 1e-6], [1e-5, 1e-4]$, and $[1e-7, 1e-4]$ (Table G.25 and G.27). All the models were trained 24 times with randomly initialized weights for the Bohachevsky function using a training data set of 1000 samples.

The two-sample Hotelling's T_2 test is sensitive to violations of the assumption of equal variance and covariance. Therefore, we must assess if the variance–covariance matrices statisitcally significantly differ to decide whether to perform a homoscedastic or a heteroscedastic test. In this work, we use the multivariate statistical test Box's M (assuming multivariate normality) to check the equality of multiple covariance matrices. The Box M test is sensitive to violations of multivariate non-normality [30]. The tested data does not have a univariate or multivariate normal distribution (results omitted). However, instead of taking on a Chi_2 approximation, Box's M test can take on an F-test approximation which is largely robust to normality violations [30,31]. Although the Chi_2 approximation is more accurate for group sample sizes larger than or equal to 20, we use the F-test approximation due to its robustness to normality violations.

The null hypothesis for the Box M test is that the observed covariance matrices for the dependent variables are equal across groups. In other words, a non-significant test result ($P> 0.05$) indicates that the covariance matrices are equal. If the covariance matrices are not significantly different (homoscedastic) and the groups' sample size is at least 50, Hotelling's T_2 test takes a Chi_2 approximation; otherwise, it takes an F approximation. If the covariance matrices are significantly different (heteroscedastic), Hotelling's T_2 test takes a Chi_2 approximation.

We perform the Box's M and two-sample Hotelling's T_2 tests to assess if changing the prior weight or the regularization weight affects the multivariate empirical and prior-consistency RMSEs of the proposed models (Table G.26 and G.28).

The results from the prior weight $1e-4$ and the three regularization weights (72 samples per model variant, 24 random weight initializations by 3 regularization weights) were combined in a group for each model variant, and the Box's M and two-sample Hotelling's T_2 tests were applied to compare the models (Table G.29). Finally, we combine the results from the three prior weights and the three regularization weights for each model variant (144 samples per model variant, 24 random weight initializations by 3 regularization weights, and 24 random weight initializations by 3 prior weights) and apply the Box's M and two-sample Hotelling's T_2 tests (Table G.30).

We report the Box's M statistic, Box's M statistic P significance level, T_2 statistic (F statistic or chi-square statistic), and T_2 statistic P significance level. The hyperparameter investigation study results in Table G.26 and G.28 show that $PGNN_1$ and $PGNN_2$ performance for the Bohachevsky function is not statically significantly affected by the prior loss weight or the regularization loss weight. Table G.29 and G.30 show that the proposed methods $PGNN_1$ and $PGNN_2$ have a statistically significant impact on the multivariate empirical and prior-consistency error means relative to the reference models for different prior and regularization loss weights.

In conclusion, (1) the proposed methods are not affected by the loss weight hyperparameters and (2) for a given set of hyperparameters, the proposed methods consistently improve the empirical and prior-consistency RMSE means relative to the reference models (Table G.29 and G.30). The exception to the empirical and prior-consistency RMSE mean improvements is for the initial prior case study, where there is no statically significant difference in the multivariate empirical and prior-consistency RMSE performance between $PGNN_2$ and PGNN−100. The lack of statically significant difference in the multivariate errors is consistent with the findings regarding the smaller benefit of the initial prior especially for the Bohachevsky function discussed in SubSections 4.4, 4.3, and 4.5.

5. Conclusion

The literature on physics-guided NNs (PGNN) shows that when PGNNs are expertly tailored to specific problems, they can perform better than traditional NNs for small and noisy synthetic/experimental data sets. Expertly tailored PGNNs have also been used as computationally-efficient and accurate counterparts to physics-based numerical simulations [16,32]. However, the need for expert

knowledge to tailor the PGNN structure (number of hidden layers and units) limits the applications and objectivity of PGNNs [15,16].

In this investigation, the aim was to develop and statistically test a systematic framework to optimize the structure of PGNNs while ensuring empirical generalizability and prior-consistency. We propose a framework to optimize the number of hidden units via a line search and cross-validation using the empirical error to eliminate data-set/model-structure application dependency ($PGNN_1$). In addition to using the prior-knowledge in the model training step, we propose utilizing the prior errors as part of the cross-validation performance index to reinforce prior-consistency ($PGNN_2$). The third contribution of this work was to perform statistical analysis on the empirical and prior-consistency error performance of PGNNs under varying settings to determine when and why the proposed methods are effective.

Optimizing the network architecture ($PGNN_1$) with repeated cross-validation statistically significantly improves the empirical and prior-consistency RMSE means relative to the reference models by selecting the number of hidden units with the smallest empirical error. However, a limitation of $PGNN_1$ is the possibility of a larger hidden unit quantile range across random model weight initialization. The $PGNN_1$ limitation is hypothesized to be due to the data set size versus the function's complexity and the untuned hyperparameters.

Using the prior-knowledge and the empirical error in the validation process, $PGNN_2$ statistically improves the empirical and prior-consistency RMSE mean and variance relative to the reference models. By incorporating the prior-consistency error in the validation performance index, $PGNN_2$ re-enforces prior-consistency. $PGNN_2$ outperforms (empirical/prior error mean and variance) $PGNN_1$ relative to the reference models by selecting a suitable number of hidden units to avoid under-fitting or over-fitting depending on the training/prior data landscape versus the data set size. Further, the positive correlation between the empirical RMSE and the prior-consistency RMSEs supports utilizing the prior-consistency error in the validation performance index of PGNNs ($PGNN_2$).

The data set size investigation has shown that PGNNs result in a 1-dimensional or 2-dimensional statistically significant difference ($P < 0.05$) in the multivariate empirical and prior-consistency means for convex and non-convex modeling problems up to a 2000 sample data set size. The univariate analysis has shown a statistically significant difference ($P < 0.05$) in the empirical and prior-consistency errors for up to sample data set size 4000 and 2000, respectively. The functions' complexity, the prior value, and the correlation between the empirical and the prior-consistency errors dictate the magnitude of the multivariate difference between data set sizes. In the hyperparameter sensitivity investigation, we show that the proposed methods are not affected by the loss weight hyperparameters and consistently improve the empirical and prior-consistency RMSE means relative to the reference models.

Future work includes investigating practical applications, multi-prior loss functions, prior adaptive weighing, and the boundary data set size effect on the performance of PGNNs. Additional studies should focus on the relationship between the prior type/value and the modeling problem landscape complexity to develop a complete picture of the importance of priors in the context of PGNNs.

CRediT authorship contribution statement

Mohamed Atwya: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **George Panoutsos:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by The University of Sheffield and the UK EPSRC Future Manufacturing Hub – Manufacture using Advanced Powder Processes (MAPP) through Grant EP/P006566/1.

Appendix A. Benchmark functions

The 14 benchmark functions studied in this work are presented in Table A.10.

Table A.10
The 14 benchmark functions and their corresponding input ranges used for test the empirical and prior-consistency RMSEs of the proposed methods.

Benchmark Function	Function $y(x_1, x_2)$	x_1 range	x_2 range
McCormick	$\sin x_1 + x_2 + (x_1 - x_2)^2 - 1.50x_1 + 2.50x_2 + 1.00$	$[-1.50, 4.00]$	$[-3.00, 4.00]$
Griewank	$x_1^2/4000.00 + x_2^2/4000.00 - \cos x_1 \cos\left(x_2/\sqrt{2.00}\right) + 1.00$	$[-5.00, 5.00]$	$[-5.00, 5.00]$
Ackley	$-20\exp\left[-0.2\sqrt{0.5(x_1^2 + x_1^2)}\right]$ $-\exp\left[0.5(\cos(2\pi x_1) + \cos(2\pi x_2)\right] + \exp(1) + 20$	$[-5.00, 5.00]$	$[-5.00, 5.00]$
Branin	$(x_2 - (5.10/(4.00\pi^2))x_1^2 + (5.00/\pi)x_1 - 6.00)$ + $10.00(1 - (1/(8.00\pi)))\cos(x_1) + 10$	$[-5.00, 10.00]$	$[0.00, 15.00]$
Styblinski-Tang	$({x_1}^4 - 16.00x_1^2 + 5.00x_1 + x_2^4 - 16.00x_2^2 + 5.00x_2)/2$	$[-5.00, 5.00]$	$[-5.00, 5.00]$
Booth	$(x_1 + 2.00x_2 - 7.00)^2 + (2.00x_1 + x_2 - 5.00)^2$	$[-10.00, 10.00]$	$[-10.00, 10.00]$
Three-hump Camel	$2.00x_1^2 - 1.05x_1^4 + x^6/6.00 + x_1x_2 + x_2^2$	$[-5.00, 5.00]$	$[-5.00, 5.00]$
Bukin N.6	$100.00\sqrt{\lvert x_2 - 0.01x_1^2\rvert} + 0.01\lvert x_1 + 10.00\rvert$	$[-15.00, -5.00]$	$[-3.00, 3.00]$
Himmelblau's	$\left({x_1}^2 + x_2 - 11.00\right)^2 + \left(x_1 + x_2^2 - 7.00\right)^2$	$[-5.00, 5.00]$	$[-5.00, 5.00]$
Zakharov	$x_1^2 + x_2^2 + (0.50x_1 + x_2)^2 + (0.50x_1 + x_2)^4$	$[-5.00, 10.00]$	$[-5.00, 10.00]$
Bohachevsky	$x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7$	$[-100.00, 100.00]$	$[-100.00, 100.00]$
Rosenbrock	$100.00(x_2 - x_1^2)^2 + (1.00 - x_1)^2$	$[-5.00, 10.00]$	$[-5.00, 10.00]$
Beale	$(1.50 - x_1 + x_1x_2)^2 + \left(2.25 - x_1 + x_1x_2^2\right)^2 + \left(2.625 - x_1 + x_1x_2^3\right)^2$	$[-4.50, 4.50]$	$[-4.50, 4.50]$
Goldstein-Price	$\left[1.00 + (x_1 + x_2 + 1.00)^2(19.00 - 14.00x_1 + 3.00x_1^2 - 14.00x_2\right.$ $\left.+6.00x_1x_2 + 3.00x_2^2)\right] \times \left[30.00 + (2.00x_1 - 3.00x_2)^2 + (18.00\right.$ $\left.-32.00x_1 + 12.00x_1^2 + 48.00x_2 - 36.00x_1x_2 + 27.00x_2^2)\right]$	$[-2.00, 2.00]$	$[-2.00, 2.00]$

Appendix B. Benchmark functions statistical tests

Tables B.11 and B.12 present the independent two-sample multivariate Hotelling's test on the percentage change in the empirical and prior-consistency RMSEs between the proposed methods and the reference models (Tables 3 and 4). Similarly, Tables B.13, B.14, B.15, and B.16 present the two-sample F-test for equal variance results for the percentage change in the variance of the empirical and prior-consistency RMSEs between the proposed methods and the reference models (Tables 5 and 6).

Table B.11
Independent two-sample multivariate Hotelling's test of the empirical and prior-consistency RMSEs for the 14 benchmark functions comparing the proposed methods to the reference methods using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. P values are shaded to indicate a statistically significant difference between the values (critical P value 0.05).

Benchmark	[NN,PGNN−100]				[NN,PGNN$_1$]				[NN,PGNN$_2$]			
Function	F	P	T^2	P	F	P	T^2	P	F	P	T^2	P
McCormick	0.46	0.71	4.97	0.16	9.32	0.00	0.20	0.90	1.76	0.15	4.66	0.11
Griewank	0.90	0.44	0.73	0.70	6.10	0.00	7.90	0.02	4.51	0.00	7.91	0.02
Ackley	1.05	0.37	0.46	0.80	0.43	0.73	8.23	0.02	0.27	0.85	41.95	0.00
Branin	0.43	0.73	0.00	1.00	0.36	0.78	1.43	0.50	0.68	0.56	6.91	0.04
Styblinski–Tang	0.30	0.82	0.64	0.73	0.61	0.61	0.14	0.93	0.66	0.58	2.51	0.30
Booth	0.96	0.41	2.29	0.34	1.09	0.35	3.33	0.21	2.05	0.10	2.23	0.35
Three-hump Camel	2.20	0.09	1.77	0.43	3.89	0.01	5.23	0.07	3.58	0.01	7.29	0.03
Bukin N.6	1.35	0.26	0.96	0.63	2.34	0.07	1.78	0.43	0.97	0.41	0.26	0.88
Himmelblau's	0.41	0.75	0.16	0.92	13.81	0.00	4.71	0.09	5.55	0.00	3.20	0.20
Zakharov	0.34	0.79	0.36	0.84	5.45	0.00	3.39	0.18	3.30	0.02	3.90	0.14
Bohachevsky	0.67	0.57	0.80	0.68	0.54	0.65	6.37	0.05	0.73	0.53	8.67	0.02
Rosenbrock	0.11	0.95	0.15	0.93	3.51	0.01	4.65	0.10	0.60	0.62	1.48	0.49
Beale	1.18	0.32	3.55	0.19	2.07	0.10	2.43	0.31	2.45	0.06	4.59	0.12
Goldstein–Price	0.91	0.44	1.60	0.46	3.27	0.02	11.38	0.00	3.36	0.02	11.62	0.00

Table B.12
Independent two-sample multivariate Hotelling's test of the empirical and prior-consistency RMSEs for the 14 benchmark functions comparing the proposed methods to thereference methods using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. P values are shaded to indicate a statistically significant difference between the values (critical P value 0.05).

Benchmark	[PGNN−100,PGNN$_1$]				[PGNN−100,PGNN$_2$]				[PGNN$_1$,PGNN$_2$]			
Function	F	P	T^2	P	F	P	T^2	P	F	P	T^2	P
McCormick	7.73	0.00	3.89	0.14	2.65	0.05*	14.40	0.00	11.12	0.00	1.63	0.44
Griewank	4.79	0.00	12.34	0.00	4.83	0.00	12.67	0.00	0.89	0.45	0.85	0.66
Ackley	0.35	0.79	4.36	0.13	0.93	0.42	35.87	0.00	0.70	0.55	14.93	0.00
Branin	1.13	0.36	1.36	0.52	1.70	0.16	7.56	0.03	0.77	0.51	4.51	0.12
Styblinski–Tang	0.16	0.92	0.96	0.63	0.90	0.45	0.78	0.69	1.75	0.16	3.12	0.23
Booth	0.49	0.69	9.86	0.01	0.50	0.68	8.86	0.02	0.27	0.85	0.14	0.93
Three-hump Camel	1.87	0.13	3.22	0.22	2.04	0.11	1.88	0.41	0.12	0.99	1.81	0.42
Bukin N.6	1.62	0.18	2.41	0.32	0.74	0.53	1.43	0.50	1.93	0.12	0.61	0.74
Himmelblau's	12.80	0.00	8.52	0.01	5.30	0.00	6.11	0.05*	3.23	0.02	2.11	0.35
Zakharov	3.73	0.01	4.32	0.12	1.99	0.11	5.61	0.08	0.90	0.44	1.90	0.40
Bohachevsky	1.29	0.28	8.09	0.03	2.04	0.11	9.70	0.01	0.14	0.94	0.83	0.67
Rosenbrock	3.52	0.01	6.52	0.04	0.71	0.55	2.31	0.33	1.55	0.20	0.95	0.63
Beale	0.70	0.55	1.42	0.50	1.31	0.27	3.63	0.18	0.13	0.95	0.60	0.75
Goldstein–Price	3.42	0.02	6.11	0.05*	3.31	0.02	6.04	0.05*	0.10	0.96	0.08	0.96

Table B.13
Two-sample F-test for equal variances of the empirical RMSE across the model variants for the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. P values are shaded to indicate a statistically significant difference between the values (critical P value 0.05).

Benchmark	[NN−100,PGNN−100]		[NN−100,$PGNN_1$]		[NN−100,$PGNN_2$]	
Function	F	P	F	P	F	P
McCormick	0.68	0.37	0.10	0.00	1.26	0.58
Griewank	1.07	0.88	3.78	0.00	3.34	0.01
Ackley	0.55	0.15	0.79	0.57	0.91	0.81
Branin	0.71	0.42	0.93	0.87	0.95	0.89
Styblinski–Tang	1.17	0.71	1.19	0.67	1.00	0.99
Booth	0.76	0.52	1.24	0.61	1.03	0.95
Three-hump Camel	2.18	0.07	4.35	0.00	4.12	0.00
Bukin N.6	0.49	0.09	0.77	0.54	0.76	0.52
Himmelblau's	1.03	0.94	20.84	0.00	6.11	0.00
Zakharov	1.54	0.31	4.14	0.00	2.85	0.02
Bohachevsky	0.81	0.61	1.53	0.31	1.65	0.24
Rosenbrock	1.03	0.94	4.17	0.00	1.68	0.22
Beale	0.76	0.52	1.20	0.67	1.55	0.30
Goldstein–Price	1.20	0.66	2.41	0.04	2.92	0.01

Table B.14
Two-sample F-test for equal variances of the empirical RMSE across the model variants for the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. P values are shaded to indicate a statistically significant difference between the values (critical P value 0.05).

Benchmark	[PGNN−100,$PGNN_1$]		[PGNN−100,$PGNN_2$]		[$PGNN_1$,$PGNN_2$]	
Function	F	P	F	P	F	P
McCormick	0.16	0.00	2.47	0.03	15.18	0.00
Griewank	2.20	0.06	1.66	0.23	0.76	0.51
Ackley	0.91	0.83	1.46	0.37	1.60	0.27
Branin	0.89	0.79	1.12	0.78	1.26	0.58
Styblinski–Tang	0.97	0.94	1.20	0.67	1.24	0.62
Booth	1.35	0.48	1.64	0.24	1.22	0.64
Three-hump Camel	2.34	0.05*	2.27	0.05	0.97	0.95
Bukin N.6	2.08	0.09	1.43	0.40	0.69	0.38
Himmelblau's	5.17	0.00	3.23	0.01	0.62	0.27
Zakharov	4.12	0.00	2.83	0.02	0.69	0.37
Bohachevsky	0.95	0.89	0.85	0.70	0.90	0.80
Rosenbrock	2.22	0.06	1.77	0.18	0.80	0.59
Beale	1.45	0.38	1.81	0.16	1.25	0.60
Goldstein–Price	1.42	0.41	1.72	0.20	1.21	0.65

Table B.15
Two-sample F-test for equal variances of the boundary prior-consistency RMSE across the model variants for the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. P values are shaded to indicate a statistically significant difference between the values (critical P value 0.05).

Benchmark	[NN−100,PGNN−100]		[NN−100,$PGNN_1$]		[NN−100,$PGNN_2$]	
Function	F	P	F	P	F	P
McCormick	0.81	0.61	0.13	0.00	2.00	0.10
Griewank	1.12	0.79	2.46	0.04	1.86	0.15
Ackley	0.98	0.97	0.90	0.80	1.44	0.39
Branin	0.82	0.64	0.73	0.46	0.92	0.85
Styblinski–Tang	0.87	0.73	0.84	0.67	1.03	0.94
Booth	1.55	0.30	2.09	0.08	2.54	0.03
Three-hump Camel	1.65	0.24	3.85	0.00	3.74	0.00
Bukin N.6	0.61	0.25	1.27	0.57	0.88	0.76
Himmelblau's	1.19	0.69	6.13	0.00	3.83	0.00
Zakharov	1.41	0.41	5.83	0.00	4.00	0.00
Bohachevsky	1.25	0.59	1.19	0.69	1.26	0.88
Rosenbrock	0.84	0.67	1.85	0.15	1.48	0.36
Beale	0.66	0.32	0.95	0.91	1.19	0.68
Goldstein–Price	1.38	0.45	1.96	0.11	2.37	0.04

Table B.16
Two-sample F-test for equal variances of the boundary prior-consistency RMSE across the model variants for the 14 benchmark functions using 24 random weight initializations per model variant and a training data set of 1000 samples. The regularization weight for the NN model was set to $1e-3$. The regularization and prior weights for the PGNN models were set to $1e-5$ and $1e-6$. P values are shaded to indicate a statistically significant difference between the values (critical P value 0.05).

Benchmark	[PGNN−100,$PGNN_1$]		[PGNN−100,$PGNN_2$]		[$PGNN_1$,$PGNN_2$]	
Function	F	P	F	P	F	P
McCormick	0.14	0.00	1.85	0.15	13.29	0.00
Griewank	3.54	0.00	3.12	0.01	0.88	0.77
Ackley	1.44	0.39	1.66	0.23	1.15	0.74
Branin	1.31	0.52	1.33	0.50	1.01	0.98
Styblinski–Tang	1.02	0.96	0.86	0.72	0.84	0.68
Booth	1.63	0.25	1.35	0.48	0.83	0.65
Three-hump Camel	1.99	0.11	1.89	0.13	0.95	0.90
Bukin N.6	1.59	0.27	1.57	0.29	0.99	0.98
Himmelblau's	20.14	0.00	5.90	0.00	0.29	0.00
Zakharov	2.69	0.02	1.85	0.15	0.69	0.38
Bohachevsky	1.90	0.13	2.05	0.09	1.08	0.86
Rosenbrock	4.04	0.00	1.62	0.25	0.40	0.03
Beale	1.58	0.28	2.04	0.09	1.29	0.55
Goldstein–Price	2.00	0.10	2.43	0.04	1.22	0.64

Appendix C. Data set size sensitivity MANOVA study

Tables C.17 and C.18 present the empirical and prior-consistency RMSEs for the Bohachevsky and the Ackley functions with different data set sizes. Tables C.19 and C.20 present the one-way Multivariate Analysis of Variance (MANOVA) results for the Bohachevsky and Ackley case studies, respectively. Figs. C.1a, C.2a, C.3a, C.4a, C.5a, and C.6a demonstrate the canonical analysis

Table C.17
Empirical and prior-consistency RMSE data set size study for the Bohachevsky case study using a regularization weight of $1e-5$ and a prior weight of $1e-6$.

Model Type	Prior	Scenario 1		Scenario 2		Scenario 3	
	Weight	Empirical Error	Prior_1	Empirical	Prior_2	Empirical	Prior_3
200	NN−100	66.02 ± 14.75	174.86 ± 39.55	66.02 ± 14.75	66.35 ± 41.32	66.02 ± 14.75	81.64 ± 21.72
	PGNN−30	118.07 ± 24.08	264.62 ± 74.78	118.08 ± 25.20	76.82 ± 70.93	118.58 ± 25.19	141.93 ± 31.46
	PGNN−65	76.58 ± 13.72	191.64 ± 37.77	76.11 ± 13.05	63.31 ± 40.69	76.61 ± 13.52	96.33 ± 20.01
	PGNN−100	65.24 ± 13.26	173.57 ± 33.37	65.24 ± 13.26	70.79 ± 48.33	65.24 ± 13.26	80.64 ± 20.27
	PGNN_1	62.97 ± 11.62	167.72 ± 29.03	60.13 ± 8.43	68.43 ± 40.33	62.02 ± 10.26	76.08 ± 14.98
	PGNN_2	60.66 ± 9.76	159.63 ± 29.69	64.82 ± 12.68	57.92 ± 44.83	59.71 ± 8.37	72.63 ± 12.00
600	NN−100	25.95 ± 4.36	55.03 ± 10.03	25.95 ± 4.36	26.54 ± 24.69	25.95 ± 4.36	32.68 ± 5.39
	PGNN−30	67.25 ± 14.93	124.81 ± 21.03	67.24 ± 14.94	62.38 ± 43.80	67.27 ± 14.93	67.75 ± 18.84
	PGNN−65	35.23 ± 6.79	72.22 ± 18.84	35.32 ± 7.11	40.20 ± 24.10	35.37 ± 7.15	41.24 ± 7.71
	PGNN−100	26.60 ± 4.86	55.44 ± 11.73	26.60 ± 4.86	26.02 ± 20.74	33.30 ± 6.46	33.30 ± 6.46
	PGNN_1	22.36 ± 2.78	46.42 ± 7.80	22.27 ± 3.21	31.36 ± 22.68	22.59 ± 2.92	28.03 ± 3.88
	PGNN_2	22.81 ± 3.08	45.91 ± 6.76	24.44 ± 4.24	19.03 ± 13.58	22.50 ± 3.20	27.60 ± 4.37
1000	NN−100	21.11 ± 3.67	40.42 ± 8.17	21.11 ± 3.67	23.78 ± 16.67	21.11 ± 3.67	27.86 ± 5.55
	PGNN−30	61.78 ± 15.65	111.01 ± 29.81	61.79 ± 15.67	63.19 ± 44.10	61.79 ± 15.67	63.72 ± 24.58
	PGNN−65	27.83 ± 4.40	50.57 ± 10.19	27.83 ± 4.40	28.12 ± 19.00	27.83 ± 4.40	34.48 ± 6.80
	PGNN−100	21.74 ± 4.08	40.57 ± 7.29	21.74 ± 4.08	23.56 ± 17.01	21.74 ± 4.08	28.35 ± 6.19
	PGNN_1	19.06 ± 2.96	34.73 ± 7.50	19.06 ± 2.96	26.88 ± 17.88	19.06 ± 2.96	24.52 ± 4.42
	PGNN_2	19.21 ± 2.85	33.73 ± 7.91	20.95 ± 3.89	27.32 ± 15.56	19.44 ± 3.00	24.58 ± 4.43
2000	NN−100	19.33 ± 3.53	33.42 ± 7.01	19.33 ± 3.53	28.73 ± 17.92	19.33 ± 3.53	24.82 ± 4.71
	PGNN−30	59.33 ± 18.24	100.96 ± 31.15	59.33 ± 18.24	83.02 ± 42.94	59.33 ± 18.24	58.11 ± 25.85
	PGNN−65	25.30 ± 4.84	43.19 ± 7.77	25.30 ± 4.84	29.70 ± 15.25	25.30 ± 4.84	30.30 ± 7.85
	PGNN−100	19.39 ± 3.74	33.88 ± 8.74	19.39 ± 3.74	28.98 ± 16.24	19.39 ± 3.74	25.24 ± 5.16
	PGNN_1	16.16 ± 2.27	26.66 ± 5.36	16.16 ± 2.27	23.20 ± 14.65	16.16 ± 2.27	19.95 ± 3.49
	PGNN_2	16.27 ± 2.34	26.10 ± 5.31	17.97 ± 2.71	17.81 ± 8.10	15.94 ± 1.68	19.63 ± 2.76
4000	NN−100	19.18 ± 3.32	33.35 ± 6.52	19.18 ± 3.32	23.69 ± 11.50	19.18 ± 3.32	24.36 ± 4.41
	PGNN−30	61.54 ± 14.82	104.46 ± 27.65	61.54 ± 14.82	78.11 ± 48.07	61.54 ± 14.82	58.15 ± 21.03
	PGNN−65	24.54 ± 4.83	43.61 ± 8.11	24.54 ± 4.83	27.28 ± 14.90	24.54 ± 4.83	28.41 ± 8.29
	PGNN−100	18.70 ± 3.37	33.31 ± 6.85	18.70 ± 3.37	27.32 ± 16.09	18.70 ± 3.37	23.75 ± 4.19
	PGNN_1	16.29 ± 2.31	27.38 ± 4.21	16.29 ± 2.31	22.87 ± 10.57	16.29 ± 2.31	20.00 ± 3.34
	PGNN_2	16.14 ± 2.16	26.81 ± 4.07	18.23 ± 2.57	17.44 ± 9.63	16.55 ± 2.54	19.99 ± 3.50

Table C.18
Empirical and prior-consistency RMSE data set size study for the Ackley case study using a regularization weight of $1e-5$ and a prior weight of $1e-6$.

Model Type	Prior	Scenario 1		Scenario 2		Scenario 3	
	Weight	Empirical Error	Prior_1	Empirical	Prior_2	Empirical	Prior_3
200	NN−100	0.65 ± 0.01	1.22 ± 0.04	0.65 ± 0.01	2.93 ± 0.31	0.65 ± 0.01	0.27 ± 0.03
	PGNN−30	0.64 ± 0.01	1.17 ± 0.04	0.64 ± 0.01	3.33 ± 0.22	0.64 ± 0.01	0.26 ± 0.03
	PGNN−65	0.64 ± 0.01	1.18 ± 0.04	0.64 ± 0.01	3.11 ± 0.25	0.64 ± 0.01	0.28 ± 0.03
	PGNN−100	0.65 ± 0.01	1.21 ± 0.04	0.65 ± 0.01	2.91 ± 0.33	0.65 ± 0.01	0.27 ± 0.03
	PGNN_1	0.64 ± 0.01	1.18 ± 0.04	0.64 ± 0.01	3.12 ± 0.31	0.64 ± 0.01	0.27 ± 0.03
	PGNN_2	0.63 ± 0.01	1.13 ± 0.04	0.64 ± 0.01	2.75 ± 0.27	0.64 ± 0.01	0.26 ± 0.02
600	NN−100	0.59 ± 0.01	1.07 ± 0.03	0.59 ± 0.01	2.06 ± 0.29	0.59 ± 0.01	0.21 ± 0.02
	PGNN−30	0.60 ± 0.01	1.04 ± 0.03	0.60 ± 0.01	2.66 ± 0.24	0.60 ± 0.01	0.20 ± 0.02
	PGNN−65	0.59 ± 0.01	1.05 ± 0.04	0.59 ± 0.01	2.25 ± 0.32	0.59 ± 0.01	0.21 ± 0.02
	PGNN−100	0.59 ± 0.01	1.07 ± 0.03	0.59 ± 0.01	2.01 ± 0.29	0.59 ± 0.01	0.21 ± 0.01
	PGNN_1	0.59 ± 0.01	1.04 ± 0.03	0.59 ± 0.01	2.19 ± 0.29	0.59 ± 0.01	0.20 ± 0.01
	PGNN_2	0.59 ± 0.01	1.02 ± 0.02	0.59 ± 0.01	1.96 ± 0.24	0.59 ± 0.01	0.20 ± 0.01
1000	NN−100	0.59 ± 0.00	1.01 ± 0.02	0.59 ± 0.00	1.76 ± 0.27	0.59 ± 0.00	0.18 ± 0.02
	PGNN−30	0.59 ± 0.01	0.99 ± 0.04	0.59 ± 0.01	2.55 ± 0.24	0.59 ± 0.01	0.18 ± 0.02
	PGNN−65	0.59 ± 0.01	0.98 ± 0.02	0.59 ± 0.01	2.08 ± 0.24	0.59 ± 0.01	0.18 ± 0.02
	PGNN−100	0.59 ± 0.01	1.00 ± 0.02	0.59 ± 0.01	1.79 ± 0.31	0.59 ± 0.01	0.19 ± 0.02
	PGNN_1	0.58 ± 0.00	0.99 ± 0.02	0.58 ± 0.00	1.99 ± 0.21	0.58 ± 0.00	0.18 ± 0.01
	PGNN_2	0.58 ± 0.00	0.97 ± 0.02	0.59 ± 0.01	1.71 ± 0.27	0.58 ± 0.00	0.17 ± 0.01
2000	NN−100	0.57 ± 0.01	0.92 ± 0.05	0.57 ± 0.01	1.91 ± 0.22	0.57 ± 0.01	0.15 ± 0.02
	PGNN−30	0.59 ± 0.01	0.92 ± 0.03	0.59 ± 0.01	2.55 ± 0.20	0.59 ± 0.01	0.15 ± 0.03
	PGNN−65	0.58 ± 0.01	0.91 ± 0.03	0.58 ± 0.01	2.19 ± 0.22	0.58 ± 0.01	0.14 ± 0.02
	PGNN−100	0.57 ± 0.01	0.92 ± 0.03	0.57 ± 0.01	1.91 ± 0.28	0.57 ± 0.01	0.14 ± 0.01
	PGNN_1	0.57 ± 0.01	0.91 ± 0.05	0.57 ± 0.01	1.98 ± 0.37	0.57 ± 0.01	0.13 ± 0.01

(continued on next page)

Table C.18 (*continued*)

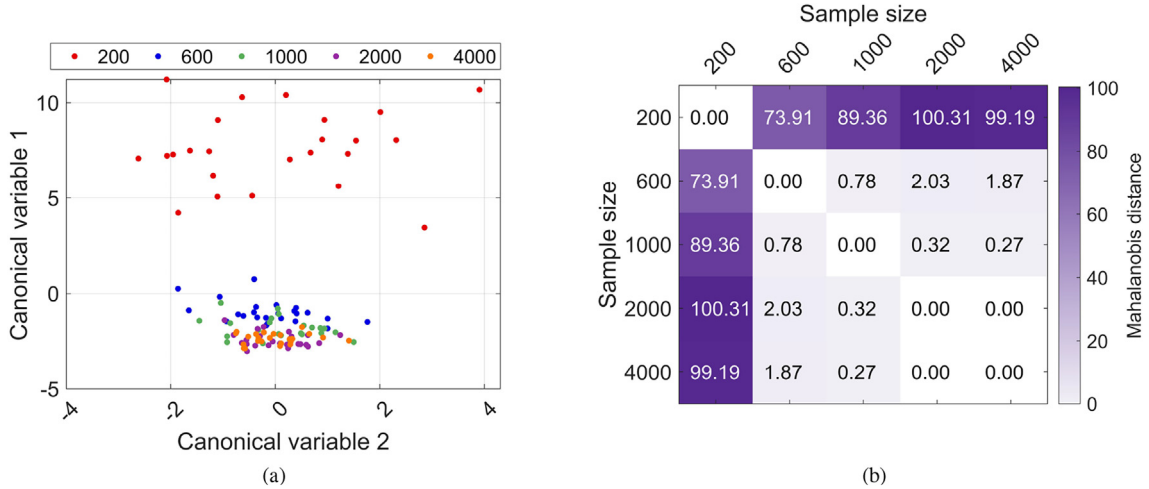
Model Type	Prior	Scenario 1		Scenario 2		Scenario 3	
	Weight	Empirical Error	Prior_1	Empirical	Prior_2	Empirical	Prior_3
	PGNN_2	0.57 ± 0.01	0.91 ± 0.05	0.57 ± 0.01	1.88 ± 0.32	0.57 ± 0.01	0.12 ± 0.01
	NN−100	0.57 ± 0.02	0.93 ± 0.06	0.57 ± 0.02	1.69 ± 0.35	0.57 ± 0.02	0.14 ± 0.03
	PGNN−30	0.58 ± 0.00	0.91 ± 0.03	0.58 ± 0.00	2.29 ± 0.18	0.58 ± 0.00	0.14 ± 0.02
	PGNN−65	0.57 ± 0.01	0.91 ± 0.03	0.57 ± 0.01	1.88 ± 0.24	0.57 ± 0.01	0.13 ± 0.02
4000	PGNN−100	0.57 ± 0.02	0.93 ± 0.06	0.57 ± 0.02	1.63 ± 0.33	0.57 ± 0.01	0.14 ± 0.02
	PGNN_1	0.56 ± 0.01	0.89 ± 0.02	0.56 ± 0.01	1.59 ± 0.27	0.56 ± 0.01	0.13 ± 0.02
	PGNN_2	0.56 ± 0.01	0.88 ± 0.02	0.56 ± 0.01	1.48 ± 0.24	0.56 ± 0.01	0.12 ± 0.01

Table C.19
MANOVA results across the five data sizes for the Bohachevsky case study using a regularization weight of $1e-5$ and a prior weight of $1e-6$. Scenarios 1, 2, and 3 correspond to the different priors. P values are shaded to indicate a statistically significant difference (critical P value 0.05).

Model Type	Scenario 1			Scenario 2			Scenario 3		
	Dimension	P_1	P_2	Dimension	P_1	P_2	Dimension	P_1	P_2
PGNN_1	1	0.00	0.90	1	0.00	0.99	1	0.00	0.35
PGNN_2	1	0.00	0.82	1	0.00	0.35	1	0.00	0.52

Table C.20
MANOVA results across the five data sizes for the Ackley case study using a regularization weight of $1e-5$ and a prior weight of $1e-6$. Scenarios 1, 2, and 3 correspond to the different priors. P values are shaded to indicate a statistically significant difference between the means (critical P value 0.05).

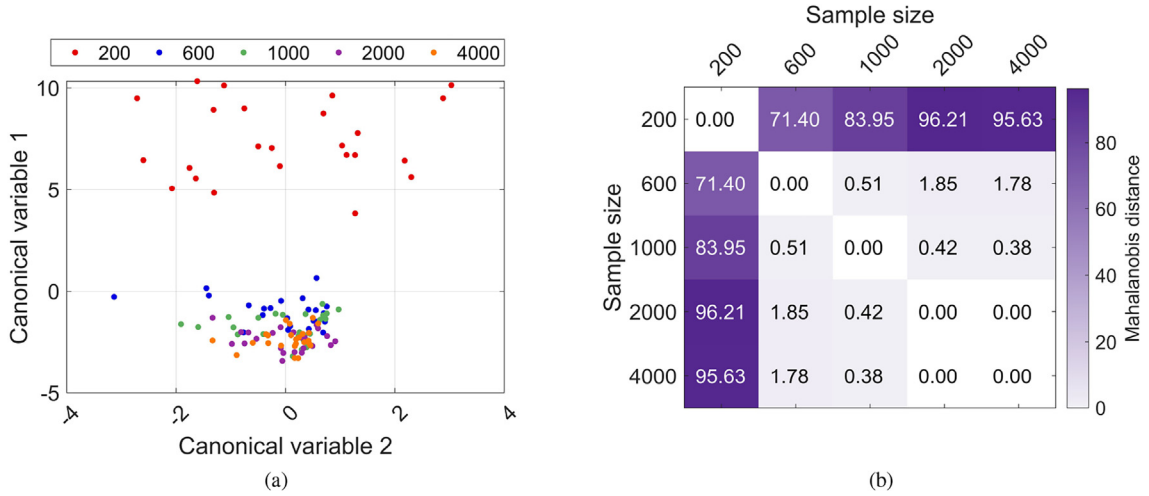
Model Type	Scenario 1			Scenario 2			Scenario 3		
	Dimension	P_1	P_2	Dimension	P_1	P_2	Dimension	P_1	P_2
PGNN_1	2	0.00	0.00	2	0.00	0.00	2	0.00	0.00
PGNN_2	2	0.00	0.00	2	0.00	0.00	2	0.00	0.00



(a) (b)

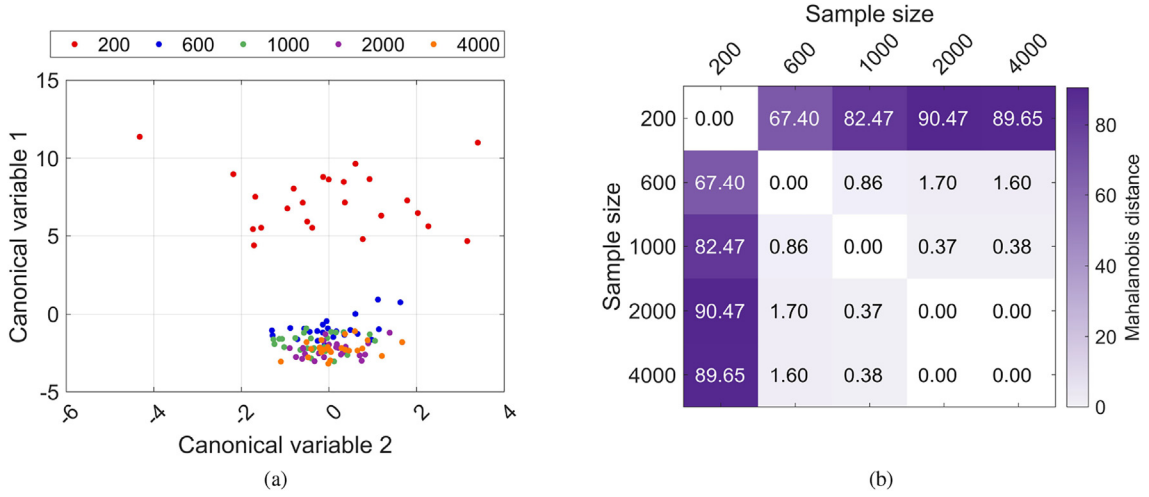
Fig. C.1. MANOVA study results for PGNN_1 using the Bohachevsky function and the boundary prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

of the Bohachevsky case study. Figs. C.1b, C.2b, C.3b, C.4b, C.5b, and C.6b demonstrate the canonical analysis of the Ackley case study. Figs. C.1b, C.2b, C.3b, C.4b, C.5b, and C.6b demonstrate the Mahalanobis distance matrix for the Bohachevsky case study. Figs. C.7b, C.8b, C.9b, C.10b, C.11b, and C.12b demonstrate the Mahalanobis distance matrix of the Ackley case study.



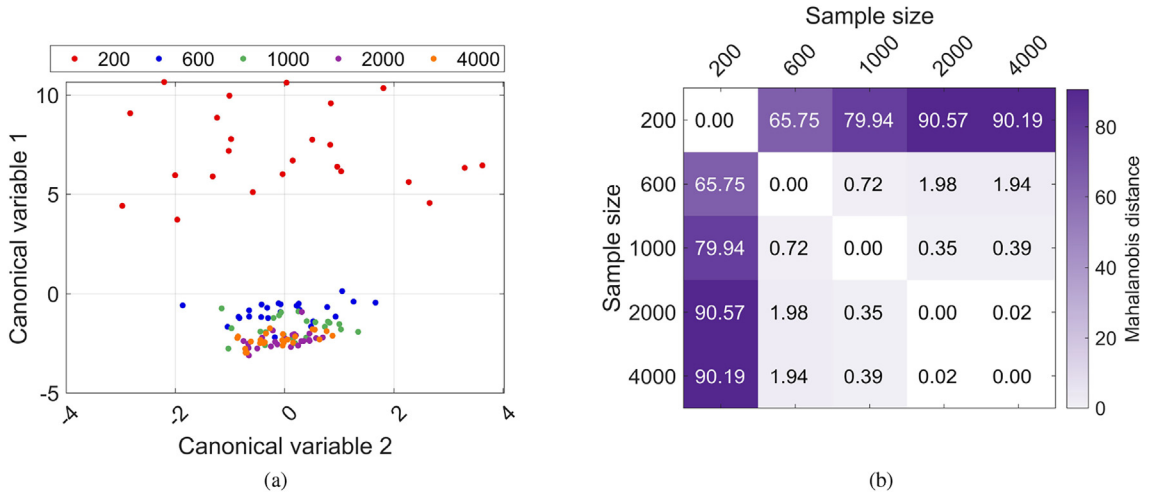
(a) (b)

Fig. C.2. MANOVA study results for $PGNN_1$ using the Bohachevsky function and the initial prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.



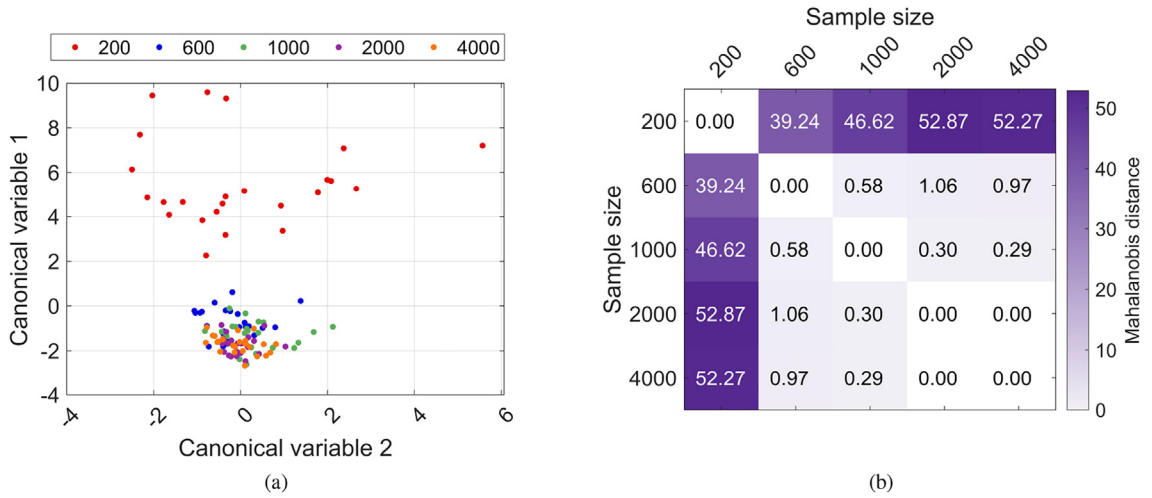
(a) (b)

Fig. C.3. MANOVA study results for $PGNN_1$ using the Bohachevsky function and the symmetry prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.



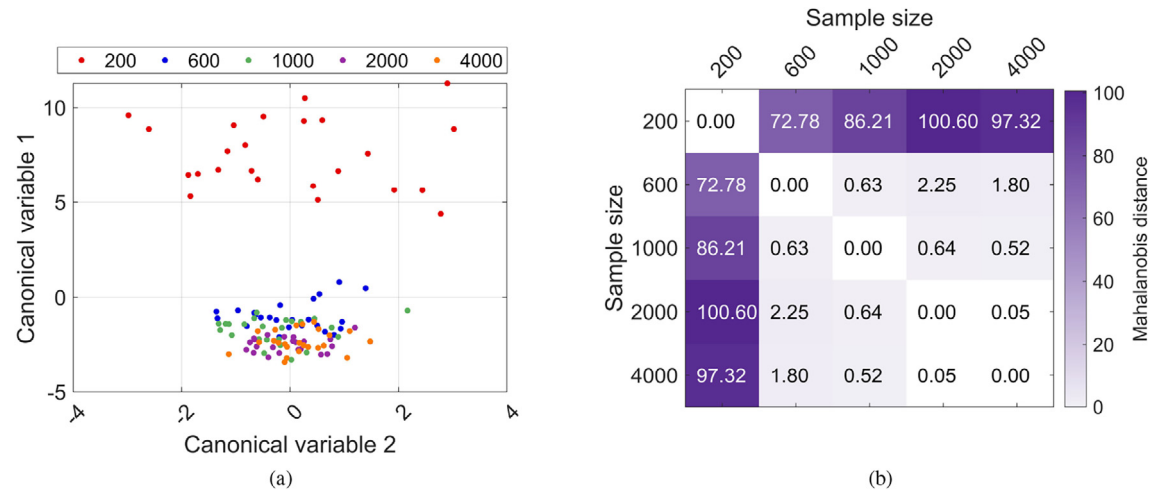
(a) (b)

Fig. C.4. MANOVA study results for $PGNN_2$ using the Bohachevsky function and the boundary prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.



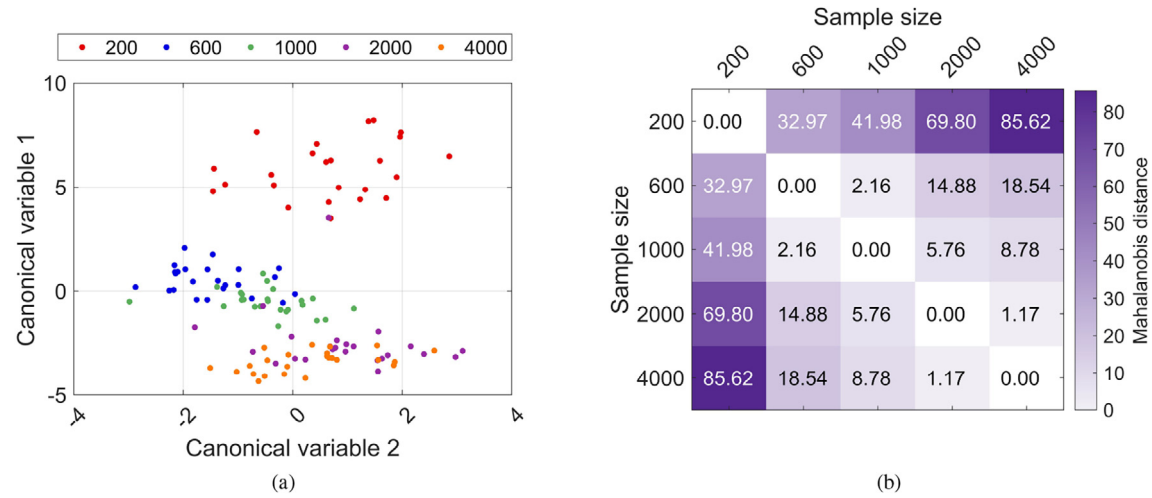
(a) (b)

Fig. C.5. MANOVA study results for $PGNN_2$ using the Bohachevsky function and the initial prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.



(a) (b)

Fig. C.6. MANOVA study results for $PGNN_2$ using the Bohachevsky function and the symmetry prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.



(a) (b)

Fig. C.7. MANOVA study results for $PGNN_1$ using the Ackley function and the boundary prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

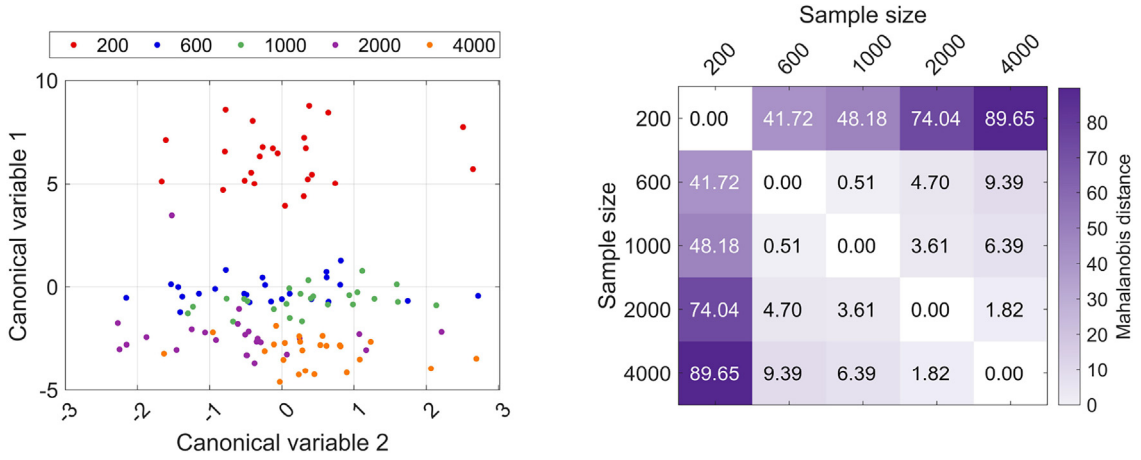


Fig. C.8. MANOVA study results for $PGNN_1$ using the Ackley function and the initial prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

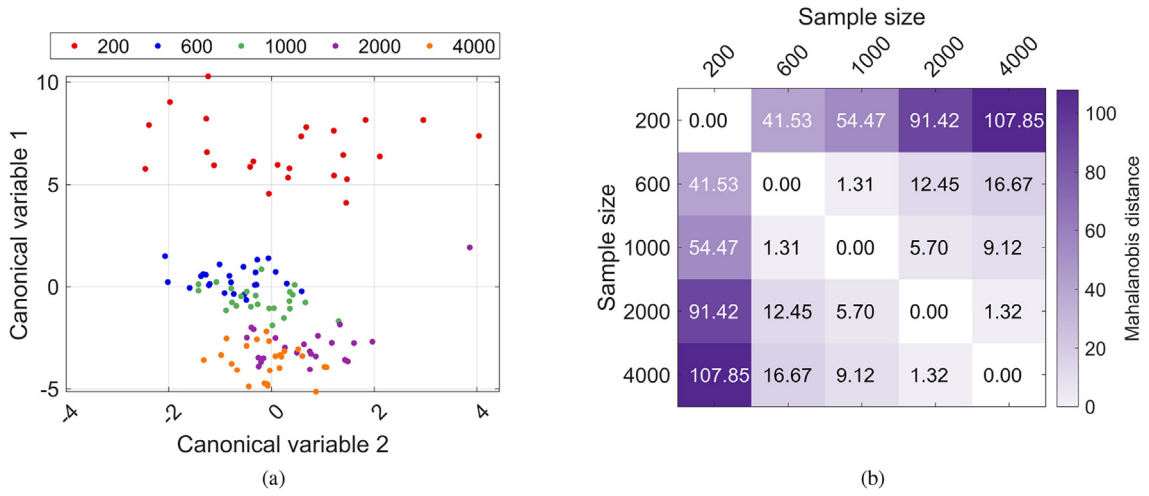


Fig. C.9. MANOVA study results for $PGNN_1$ using the Ackley function and the symmetry prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

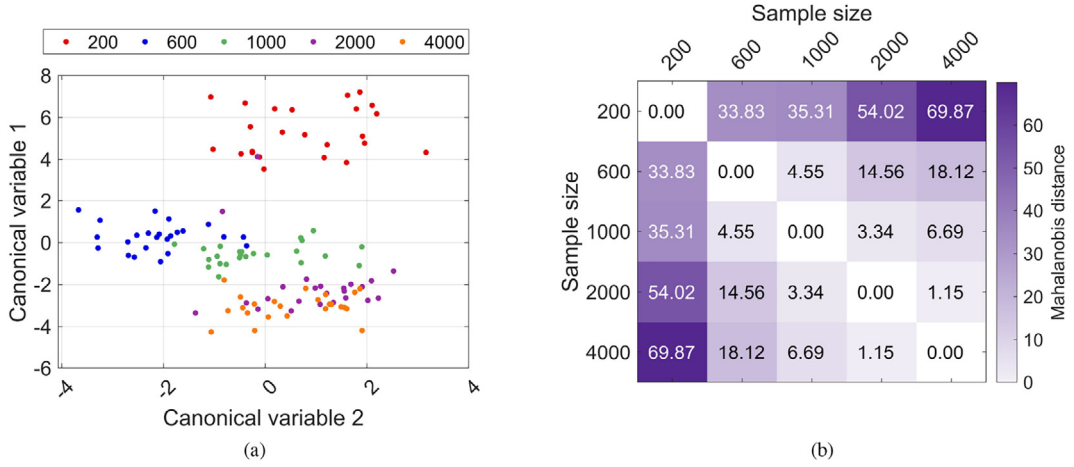


Fig. C.10. MANOVA study results for $PGNN_2$ using the Ackley function and the boundary prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

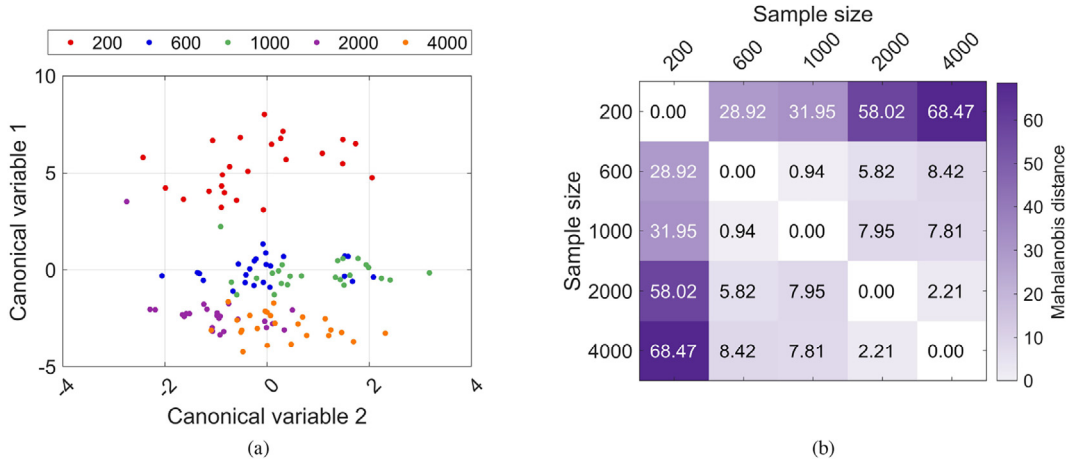


Fig. C.11. MANOVA study results for $PGNN_2$ using the Ackley function and the initial prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

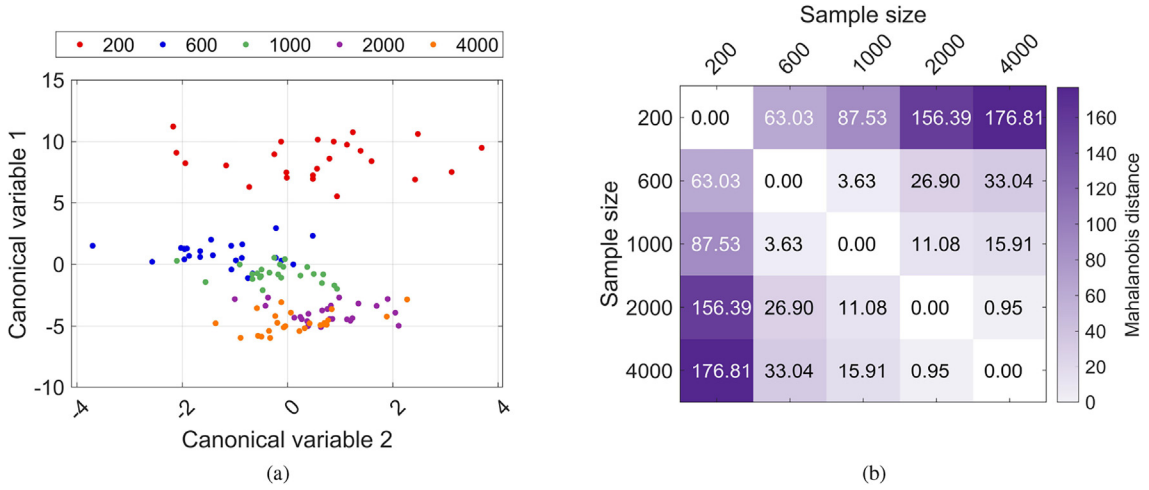


Fig. C.12. MANOVA study results for $PGNN_2$ using the Ackley function and the symmetry prior: (a) Canonical analysis scatter plot and (b) Mahalanobis distances between sample size means.

Appendix D. Data set size ANOVA and multiple comparison test study

Table D.21 presents the multiple comparison test results using Tukey's honestly significant difference procedure on the ANOVA test results (omitted but available upon request).

Appendix E. Data set size and RMSE performance correlation study

Tables E.22 and E.23 present the strength and direction of correlation between the data set size and the empirical and prior-consistency errors using Superman's rank correlation and the results from the data set size sensitivity study in SubSection 4.2.

Table D.21
Multiple comparison test of the empirical and prior-consistency RMSEs across the five data sizes for the three prior scenarios of the Bohachevsky and Ackley case studies.The model hyperparameters for this test include a regularization weight of $1e-5$ and a prior weight of $1e-6$. P values are shaded to indicate a statistically significant difference between the means (critical P value 0.05).

Model Type	Data set size compared	Empirical RMSE P values						Prior-consistency RMSE P values					
		Bohachevsky Scenario			Ackley Scenario			Bohachevsky Scenario			Ackley Scenario		
		1	2	3	1	2	3	1	2	3	1	2	3
$PGNN_1$	[200, 600]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[200, 1000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[200, 2000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[200, 4000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[600, 1000]	0.27	0.10	0.13	0.22	0.30	0.28	0.04	0.97	0.49	0.00	0.12	0.00
	[600, 2000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.75	0.00	0.00	0.09	0.00
	[600, 4000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72	0.00	0.00	0.00	0.00
	[1000, 2000]	0.40	0.17	0.30	0.00	0.00	0.00	0.29	0.98	0.22	0.00	1.00	0.00
	[1000, 4000]	0.45	0.21	0.34	0.00	0.00	0.00	0.38	0.98	0.23	0.00	0.00	0.00
	[2000, 4000]	1.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	0.21	0.00	1.00
$PGNN_2$	[200, 600]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[200, 1000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[200, 2000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[200, 4000]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[600, 1000]	0.09	0.34	0.13	0.43	0.00	0.29	0.03	0.72	0.47	0.00	0.02	0.00
	[600, 2000]	0.00	0.01	0.00	0.00	0.89	0.00	0.00	1.00	0.00	0.00	0.87	0.00
	[600, 4000]	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	[1000, 2000]	0.25	0.50	0.06	0.00	0.00	0.00	0.36	0.60	0.06	0.00	0.20	0.00
	[1000, 4000]	0.21	0.59	0.17	0.00	0.00	0.00	0.46	0.56	0.10	0.00	0.03	0.00
	[2000, 4000]	1.00	1.00	0.99	0.00	0.04	0.01	1.00	1.00	1.00	0.12	0.00	0.99

Table E.22
Spearman's rho correlation and P value between the empirical RMSE and the data size for the three prior scenarios of the Bohachevsky and Ackley case studies.The model hyperparameters for this test include a regularization weight of $1e-5$ and a prior weight of $1e-6$. P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Model Type	Bohachevsky Case Study						Ackley Case Study					
	Scenario 1		Scenario 2		Scenario 3		Scenario 1		Scenario 2		Scenario 3	
	ρ	P value	ρ	P value	ρ	P value	ρ	P value	ρ	P value	ρ	P value
$PGNN_1$	−0.83	0.00	−0.82	0.00	−0.83	0.00	−0.92	0.00	−0.92	0.00	−0.93	0.00
$PGNN_2$	−0.83	0.00	−0.76	0.00	−0.82	0.00	−0.90	0.00	−0.91	0.00	−0.92	0.00

Table E.23
Spearman's rho correlation and P value between the prior-consistency RMSE and the data size for the three prior scenarios of the Bohachevsky and Ackley case studies.The model hyperparameters for this test include a regularization weight of $1e-5$ and a prior weight of $1e-6$. P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Model Type	Bohachevsky Case Study						Ackley Case Study					
	Scenario 1		Scenario 2		Scenario 3		Scenario 1		Scenario 2		Scenario 3	
	ρ	P value	ρ	P value	ρ	P value	ρ	P value	ρ	P value	ρ	P value
$PGNN_1$	−0.85	0.00	−0.34	0.00	−0.82	0.00	−0.92	0.00	−0.79	0.00	−0.91	0.00
$PGNN_2$	−0.84	0.00	−0.34	0.00	−0.80	0.00	−0.92	0.00	−0.72	0.00	−0.93	0.00

Appendix F. Empirical and prior-consistency performance correlation study

Table F.24 presents the strength and direction of correlation between the empirical and prior-consistency errors using Superman's rank correlation and the results from the data set size sensitivity study in SubSection 4.2.

Appendix G. Hyperparameter sensitivity investigation

Tables G.25 and G.27 present the results of training the PGNN−30, PGNN−65, PGNN−100, $PGNN_1$, and $PGNN_2$ models with the following $[\rho_r, \rho_p]$ loss weights: $[1e-3, 1e-4], [1e-3, 1e-6], [1e-3, 1e-6], [1e-5, 1e-4]$, and $[1e-7, 1e-4]$. The NN−100 reference model was trained with a

Table F.24
Spearman's rho correlation and P value between the empirical RMSE and the prior-consistency RMSE for the three prior scenarios of the Bohachevsky and Ackley case studies.The model hyperparameters for this test include a regularization weight of $1e-5$ and a prior weight of $1e-6$. P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Model Type	Bohachevsky Case Study						Ackley Case Study					
	Scenario 1		Scenario 2		Scenario 3		Scenario 1		Scenario 2		Scenario 3	
	ρ	P value	ρ	P value	ρ	P value	ρ	P value	ρ	P value	ρ	P value
NN	0.94	0.00	0.34	0.00	0.98	0.00	0.95	0.00	0.74	0.00	0.90	0.00
PGNN-30	0.95	0.00	0.11	0.24	0.89	0.00	0.92	0.00	0.85	0.00	0.85	0.00
PGNN-65	0.93	0.00	0.26	0.00	0.95	0.00	0.93	0.00	0.77	0.00	0.88	0.00
PGNN-100	0.96	0.00	0.36	0.00	0.98	0.00	0.94	0.00	0.75	0.00	0.89	0.00
$PGNN_1$	0.94	0.00	0.43	0.00	0.98	0.00	0.94	0.00	0.85	0.00	0.90	0.00
$PGNN_2$	0.95	0.00	0.37	0.00	0.97	0.00	0.93	0.00	0.76	0.00	0.88	0.00

Table G.25
Prior weight sensitivity study on the empirical and prior-consistency RMSE for the Bohachevsky case study using a 1000 sample data set and a regularization weight of $1e-3$.

Model Type	Prior	Scenario 1		Scenario 2		Scenario 3	
	Weight	Empirical Error	$Prior_1$	Empirical	$Prior_2$	Empirical	$Prior_3$
NN−100	–	21.11 ± 3.67	40.42 ± 8.17	21.11 ± 3.67	23.78 ± 16.67	21.11 ± 3.67	27.86 ± 5.55
PGNN−30	$1e-4$	61.36 ± 15.32	111.34 ± 29.26	61.41 ± 15.41	69.33 ± 40.54	61.41 ± 15.41	61.82 ± 24.14
	$1e-6$	61.44 ± 15.49	111.52 ± 29.67	61.41 ± 15.41	69.33 ± 40.54	61.41 ± 15.41	61.82 ± 24.14
	$1e-8$	61.41 ± 15.41	111.44 ± 29.43	61.41 ± 15.41	69.33 ± 40.54	61.41 ± 15.41	61.82 ± 24.14
PGNN−65	$1e-4$	27.19 ± 3.77	50.62 ± 9.72	27.17 ± 3.78	29.95 ± 16.87	27.17 ± 3.78	33.15 ± 5.88
	$1e-6$	27.17 ± 3.78	50.42 ± 9.95	27.17 ± 3.78	29.95 ± 16.87	27.17 ± 3.78	33.15 ± 5.88
	$1e-8$	27.17 ± 3.78	50.43 ± 9.95	27.17 ± 3.78	29.95 ± 16.87	27.17 ± 3.78	33.15 ± 5.88
PGNN−100	$1e-4$	21.12 ± 3.68	40.23 ± 8.33	21.11 ± 3.67	23.78 ± 16.76	21.11 ± 3.67	27.86 ± 5.55
	$1e-6$	21.11 ± 3.67	40.42 ± 8.17	21.11 ± 3.67	23.78 ± 16.76	21.11 ± 3.67	27.86 ± 5.55
	$1e-8$	21.11 ± 3.67	40.42 ± 8.17	21.11 ± 3.67	23.78 ± 16.76	21.11 ± 3.67	27.86 ± 5.55
$PGNN_1$	$1e-4$	19.34 ± 2.81	35.09 ± 5.49	19.38 ± 2.98	26.06 ± 17.04	19.38 ± 2.98	25.10 ± 4.71
	$1e-6$	19.38 ± 2.98	35.08 ± 6.05	19.38 ± 2.98	26.06 ± 17.04	19.38 ± 2.98	25.10 ± 4.71
	$1e-8$	19.38 ± 2.98	35.08 ± 6.05	19.38 ± 2.98	26.06 ± 17.04	19.38 ± 2.98	25.10 ± 4.71
$PGNN_2$	$1e-4$	19.66 ± 2.26	35.71 ± 5.53	20.97 ± 3.67	24.07 ± 13.09	19.65 ± 2.94	25.24 ± 4.48
	$1e-6$	19.85 ± 2.50	35.66 ± 5.71	20.97 ± 3.67	24.07 ± 13.09	19.65 ± 2.94	25.24 ± 4.48
	$1e-8$	19.85 ± 2.50	35.66 ± 5.71	20.97 ± 3.67	24.07 ± 13.09	19.65 ± 2.94	25.24 ± 4.48

Table G.26
T-test of the prior weight sensitivty study on the multivariate empirical and prior-consistency RMSEs for the Bohachevsky case study using a 1000 sample data set and a regularization weight of $1e-3$. P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Model Type	Prior weights	Scenario 1				Scenario 2				Scenario 3			
	compared	F	P	T_2	P	F	P	T_2	P	F	P	T_2	P
$PGNN_1$	$[1e-4, 1e-6]$	0.07	0.98	0.01	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
	$[1e-4, 1e-8]$	0.07	0.98	0.01	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
	$[1e-6, 1e-8]$	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
$PGNN_2$	$[1e-4, 1e-6]$	0.08	0.97	0.16	0.92	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
	$[1e-4, 1e-8]$	0.08	0.97	0.16	0.92	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
	$[1e-6, 1e-8]$	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Table G.27
Regularisation weight sensitivity study on the empirical and prior-consistency RMSE for the Bohachevsky case study using a 1000 sample data set and a prior weight of $1e-4$.

Model Type	$\mathscr{L}_2$	Scenario 1		Scenario 2		Scenario 3	
	Weight	Empirical Error	Prior_1	Empirical	Prior_2	Empirical	Prior_3
NN−100	–	21.11 ± 3.67	40.42 ± 8.17	21.11 ± 3.67	23.78 ± 16.67	21.11 ± 3.67	27.86 ± 5.55
PGNN−30	$1e-3$	61.36 ± 15.32	111.34 ± 29.26	61.41 ± 15.41	69.33 ± 40.54	61.41 ± 15.41	61.82 ± 24.14
	$1e-5$	61.78 ± 15.67	111.04 ± 29.82	61.79 ± 15.67	63.19 ± 44.10	61.79 ± 15.67	63.71 ± 24.56
	$1e-7$	61.47 ± 16.25	111.42 ± 27.77	61.33 ± 16.30	57.06 ± 44.01	61.33 ± 16.29	62.93 ± 24.72
PGNN−65	$1e-3$	27.19 ± 3.77	50.62 ± 9.72	27.17 ± 3.78	29.95 ± 16.87	27.17 ± 3.78	33.15 ± 5.88
	$1e-5$	27.84 ± 4.40	50.99 ± 10.20	27.83 ± 4.40	28.12 ± 19.00	27.83 ± 4.40	34.48 ± 6.80
	$1e-7$	28.03 ± 4.02	51.99 ± 10.54	28.02 ± 4.01	28.36 ± 17.94	28.02 ± 4.01	34.45 ± 6.07
PGNN−100	$1e-3$	21.12 ± 3.68	20.23 ± 8.33	21.11 ± 3.67	23.78 ± 16.76	21.11 ± 3.67	27.86 ± 5.55
	$1e-5$	21.71 ± 4.10	40.56 ± 7.34	21.74 ± 4.08	23.56 ± 17.01	21.74 ± 4.08	28.35 ± 6.19
	$1e-7$	21.85 ± 3.98	41.95 ± 7.45	21.83 ± 4.00	23.82 ± 12.32	21.83 ± 4.00	28.53 ± 5.82
PGNN_1	$1e-3$	19.34 ± 2.81	35.09 ± 5.49	19.38 ± 2.98	26.06 ± 17.04	19.38 ± 2.98	25.10 ± 4.71
	$1e-5$	19.19 ± 2.96	35.10 ± 7.49	19.06 ± 2.96	26.88 ± 17.88	19.06 ± 2.96	24.52 ± 4.42
	$1e-7$	18.59 ± 2.34	33.60 ± 5.62	18.77 ± 2.58	21.16 ± 11.81	18.50 ± 2.34	23.86 ± 4.11
PGNN_2	$1e-3$	19.66 ± 2.26	35.71 ± 5.53	20.97 ± 3.67	24.07 ± 13.09	19.65 ± 2.94	25.24 ± 4.48
	$1e-5$	19.21 ± 2.65	32.88 ± 6.79	20.95 ± 3.89	27.32 ± 15.57	19.44 ± 3.00	24.58 ± 4.43
	$1e-7$	19.79 ± 3.27	36.37 ± 7.13	21.22 ± 2.89	19.09 ± 13.38	18.70 ± 2.35	23.52 ± 3.95

Table G.28
T-test of the regularisation weight sensitivity study on the multivariate empirical and prior-consistency RMSEs for the Bohachevsky case study using a 1000 sample data set a prior weight of $1e-4$. P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Model Type	Reg. weights	Scenario 1				Scenario 2				Scenario 3			
	compared	F	P	T^2	P	F	P	T^2	P	F	P	T^2	P
PGNN_1	$[1e-3, 1e-5]$	0.92	0.43	0.07	0.97	0.03	0.99	0.19	0.91	0.46	0.71	0.22	0.90
	$[1e-3, 1e-7]$	0.43	0.73	1.10	0.59	1.16	0.32	1.71	0.44	0.82	0.49	1.40	0.51
	$[1e-5, 1e-7]$	0.94	0.42	0.71	0.71	1.38	0.25	1.77	0.43	1.21	0.30	0.69	0.71
PGNN_2	$[1e-3, 1e-5]$	0.44	0.72	2.84	0.26	0.40	0.75	0.62	0.74	3.88	0.01	0.67	0.72
	$[1e-3, 1e-7]$	1.01	0.39	0.16	0.92	0.42	0.74	1.75	0.43	2.99	0.03	2.05	0.39
	$[1e-5, 1e-7]$	0.59	0.62	3.97	0.16	0.85	0.46	4.09	0.15	0.71	0.55	0.90	0.65

Table G.29
T-test for the Bohachevsky case study using the 1000 sample data set with a prior weight of $1e-4$ and the results from the three regularization weights for each model variant (72 samples per model variant, 24 random weight initializations by 3 regularization weights). P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Models	Scenario 1				Scenario 2				Scenario 3			
compared	F	P	T^2	P	F	P	T^2	P	F	P	T^2	P
[PGNN−30,PGNN_1]	66.58	0.00	531.41	0.00	68.96	0.00	537.92	0.00	111.82	0.00	875.43	0.00
[PGNN−30,PGNN_2]	64.11	0.00	520.08	0.00	63.45	0.00	488.84	0.00	101.06	0.00	847.02	0.00
[PGNN−65,PGNN_1]	6.22	0.00	228.95	0.00	3.65	0.01	221.12	0.00	20.65	0.00	262.65	0.00
[PGNN−65,PGNN_2]	5.02	0.00	200.96	0.00	1.68	0.17	115.86	0.00	12.84	0.00	229.74	0.00
[PGNN−100,PGNN_1]	3.03	0.03	29.65	0.00	2.37	0.07	20.68	0.00	2.99	0.03	21.17	0.00
[PGNN−100,PGNN_2]	3.06	0.03	25.13	0.00	0.59	0.62	0.70	0.70	3.37	0.02	20.37	0.00
[PGNN_1,PGNN_2]	0.12	0.95	1.73	0.42	1.37	0.25	14.80	0.00	1.62	0.18	3.79	0.15

Table G.30
T-test for the Bohachevsky case study using the 1000 sample data set for the results form the three prior weights and the results from the three regularization weights for each model variant (144 samples per model variant, 24 random weight initializations by 3 regularization weights and 24 random weight initializations by 3 prior weights). P values are shaded to indicate a statistically significant correlation (critical P value 0.05).

Models	Scenario 1				Scenario 2				Scenario 3			
compared	F	P	T^2	P	F	P	T^2	P	F	P	T^2	P
[PGNN−30,PGNN_1]	131.76	0.00	1069.33	0.00	132.29	0.00	108.37	0.00	225.18	0.00	1898.06	0.00
[PGNN−30,PGNN_2]	134.94	0.00	1052.08	0.00	126.66	0.00	995.91	0.00	216.50	0.00	1850.09	0.00
[PGNN−65,PGNN_1]	11.68	0.00	428.14	0.00	5.35	0.00	416.01	0.00	48.90	0.00	483.94	0.00
[PGNN−65,PGNN_2]	11.70	0.00	402.15	0.00	40.95	0.00	433.49	0.00	40.95	0.00	433.49	0.00
[PGNN−100,PGNN_1]	4.63	0.00	49.71	0.00	3.32	0.02	32.57	0.00	5.52	0.00	30.53	0.00
[PGNN−100,PGNN_2]	6.82	0.00	42.27	0.00	2.20	0.09	0.59	0.74	3.89	0.01	31.37	0.00
[PGNN_1,PGNN_2]	0.63	0.60	2.49	0.29	4.34	0.00	23.90	0.00	0.90	0.44	6.48	0.04

regularization weight $\rho_r = 1e-3$. All the models were trained 24 times with randomly initialized weights for the Bohachevsky function using a training data set of 1000 samples.

Tables G.26 and G.28 present the Box's M and two-sample Hotelling's T_2 test results which assess if changing the prior weight or the regularization weight have an effect on the multivariate empirical and prior-consistency RMSEs of the $PGNN_1$ and $PGNN_2$ models.

Table G.29 presents the Box's M and two-sample Hotelling's T_2 test results comparing the proposed models to the reference models with the prior weight $1e-4$ and the three regularization weights (72 samples per model variant, 24 random weight initializations by 3 regularization weights).

Table G.30 presents the Box's M and two-sample Hotelling's T_2 test results comparing the proposed models to the reference models with the three prior weights and the three regularization weights for each model variant (144 samples per model variant, 24 random weight initializations by 3 regularization weights and 24 random weight initializations by 3 prior weights).

References

[1] A. Karpatne, G. Atluri, J.H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, V. Kumar, Theory-guided data science: A new paradigm for scientific discovery from data, IEEE Trans. Knowl. Data Eng. 29 (10) (2017) 2318–2331.

[2] G. Hautier, C.C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory, Chem. Mater. 22 (12) (2010) 3762–3767.

[3] X. Zhao, K. Shirvan, R.K. Salko, F. Guo, On the prediction of critical heat flux using a physics-informed machine learning-aided framework, Appl. Therm. Eng. 164 (2020) 114540.

[4] V. Schmidt, A. Luccioni, S.K. Mukkavilli, N. Balasooriya, K. Sankaran, J. Chayes, Y. Bengio, Visualizing the consequences of climate change using cycle-consistent adversarial networks, arXiv preprint arXiv:1905.03709.

[5] F. Schrodt, J. Kattge, H. Shan, F. Fazayeli, J. Joswig, A. Banerjee, M. Reichstein, G. Bönisch, S. Díaz, J. Dickie, et al., Bhpmf –a hierarchical bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography, Glob. Ecol. Biogeogr. 24 (12) (2015) 1510–1521.

[6] X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, V. Kumar, Physics guided rnns for modeling dynamical systems: A case study in simulating lake temperature profiles, in: Proceedings of the 2019 SIAM International Conference on Data Mining SIAM, 2019, pp. 558–566.

[7] L. von Rueden, S. Mayer, J. Garcke, C. Bauckhage, J. Schuecker, Informed machine learning-towards a taxonomy of explicit integration of knowledge into machine learning, arXiv preprint arXiv:1903.12394.

[8] H. Denli, N. Subrahmanya, et al., Multi-scale graphical models for spatio-temporal processes, Advances in Neural Information Processing Systems (2014) 316–324.

[9] K.C. Wong, L. Wang, P. Shi, Active model with orthotropic hyperelastic material for cardiac image analysis, in: International Conference on Functional Imaging and Modeling of the Heart Springer, 2009, pp. 229–238.

[10] S. Chatterjee, K. Steinhaeuser, A. Banerjee, S. Chatterjee, A. Ganguly, Sparse group lasso: Consistency and climate applications, in: Proceedings of the 2012 SIAM International Conference on Data Mining SIAM, 2012, pp. 47–58.

[11] J. Liu, K. Wang, S. Ma, J. Huang, Accounting for linkage disequilibrium in genome-wide association studies: A penalized regression method, Stat. Interface 6 (1) (2013) 99.

[12] A. Karpatne, W. Watkins, J. Read, V. Kumar, Physics-guided neural networks (pgnn): An application in lake temperature modeling, arXiv preprint arXiv:1710.11431.

[13] N. Muralidhar, M.R. Islam, M. Marwah, A. Karpatne, N. Ramakrishnan, Incorporating prior domain knowledge into deep neural networks, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 36–45.

[14] J.-L. Wu, K. Kashinath, A. Albert, D. Chirila, H. Xiao, et al., Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems, arXiv preprint arXiv:1905.06841.

[15] D. Liu, Y. Wang, Multi-fidelity physics-constrained neural network and its application in materials modeling, J. Mech. Design 141(12).

[16] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.

[17] S. Goswami, C. Anitescu, S. Chakraborty, T. Rabczuk, Transfer learning enhanced physics informed neural network for phase-field modeling of fracture, Theoret. Appl. Fract. Mech. 106 (2020) 102447.

[18] N. Geneva, N. Zabaras, Modeling the dynamics of pde systems with physics-constrained deep auto-regressive networks, J. Comput. Phys. 403 (2020) 109056.

[19] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, P. Perdikaris, Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data, J. Comput. Phys. 394 (2019) 56–81.

[20] S. Chakraborty, P.P. Chattopadhyay, S.K. Ghosh, S. Datta, Incorporation of prior knowledge in neural network model for continuous cooling of steel using genetic algorithm, Appl. Soft Comput. 58 (2017) 297–306.

[21] A.D. Jagtap, K. Kawaguchi, G.E. Karniadakis, Adaptive activation functions accelerate convergence in deep and physics-informed neural networks, J. Comput. Phys. 404 (2020) 109136.

[22] J. Chen, Y. Liu, Probabilistic physics-guided machine learning for fatigue data analysis, Expert Syst. Appl. 168 (2021) 114316.

[23] S. Roychowdhury, M. Diligenti, M. Gori, Regularizing deep networks with prior knowledge: A constraint-based approach, Knowl.-Based Syst. 222 (2021) 106989.

[24] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer Science & Business Media, 2009.

[25] T. de Wolff, H. Carrillo, L. Martí, N. Sanchez-Pi, Towards optimally weighted physics-informed neural networks in ocean modelling, arXiv preprint arXiv:2106.08747.

[26] T. Dash, S. Chitlangia, A. Ahuja, A. Srinivasan, Incorporating domain knowledge into deep neural networks, arXiv preprint arXiv:2103.00180.

[27] R. Wang, R. Yu, Physics-guided deep learning for dynamical systems: A survey, arXiv preprint arXiv:2107.01272.

[28] J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, Integrating scientific knowledge with machine learning for engineering and environmental systems (2021). arXiv:2003.04919.

[29] I. Nabney, Netlab Algorithms for Pattern Recognition, Springer, 2004.

[30] R.E. Schumacker, Using R with multivariate statistics, Sage Publications, 2015.

[31] J.P. Stevens, Applied multivariate statistics for the social sciences, Routledge, 2012.

[32] Q. Wang, J. Li, M. Gouge, A.R. Nassar, P.P. Michaleris, E.W. Reutzel, Physics-based multivariable modeling and feedback linearization control of melt-pool geometry and temperature in directed energy deposition, J. Manuf. Sci. Eng. 139 (2) (2017) 021013.

Mohamed Atwya received the Integrated Master's of Engineering degree in industry in mechatronic and robotic engineering from the University of Sheffield, Sheffield, U.K., in 2018, where he is currently working toward the Ph.D. degree in automatic control and systems engineering. His research interests include data-driven modeling, physics-based modeling, and control systems design and analysis.

George Panoutsos received his PhD degree in automatic control and systems engineering from the University of Sheffield, Sheffield, U.K, in 2007. He joined the Department of Automatic Control and Systems Engineering (University of Sheffield, UK) as a Lecturer in November 2009, and promoted to Professor of Computational Intelligence in January 2019. George's research is financially supported by the UK EPSRC, Innovate UK, EU Horizon 2020 as well as directly by industry. George has over 60 research publications in theoretical as well as applied contributions in the areas of computational intelligence, data-driven modelling, optimisation, and decision support systems. In terms of applied research, the majority of his work is on advanced manufacturing systems, as well as healthcare applications, while also currently exploring research applications in infrastructure.