Journal of Computational Physics 444 (2021) 110551

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Mutual information for explainable deep learning of multiscale systems

Søren Taverniers [a,1], Eric J. Hall [b,*,1], Markos A. Katsoulakis [c], Daniel M. Tartakovsky [a,*]

[a] Department of Energy Resources Engineering, Stanford University, Stanford, CA 94305, USA
[b] Division of Mathematics, University of Dundee, Dundee, DD1 4HN, UK
[c] Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003, USA

ARTICLE INFO

Article history:
Available online 9 July 2021

Keywords:
Surrogate model
Mutual information
Global sensitivity analysis
Black box
Probabilistic graphical model
Electrical double-layer capacitor

ABSTRACT

Timely completion of design cycles for complex systems ranging from consumer electronics to hypersonic vehicles relies on rapid simulation-based prototyping. The latter typically involves high-dimensional spaces of possibly correlated control variables (CVs) and quantities of interest (QoIs) with non-Gaussian and possibly multimodal distributions. We develop a model-agnostic, moment-independent global sensitivity analysis (GSA) that relies on differential mutual information to rank the effects of CVs on QoIs. The data requirements of this information-theoretic approach to GSA are met by replacing computationally intensive components of the physics-based model with a deep neural network surrogate. Subsequently, the GSA is used to explain the surrogate predictions, and the surrogate-driven GSA is deployed as an uncertainty quantification emulator to close design loops. Viewed as an uncertainty quantification method for interrogating the surrogate, this framework is compatible with a wide variety of black-box models. We demonstrate that the surrogate-driven mutual information GSA provides useful and distinguishable rankings via a validation step for applications of interest in energy storage. Consequently, our information-theoretic GSA provides an "outer loop" for accelerated product design by identifying the most and least sensitive input directions and performing subsequent optimization over appropriately reduced parameter subspaces.

1. Introduction: GSA and deep learning for simulation-aided design

Simulations are a key component of product design as they enable rapid prototyping by guiding costly laboratory tests and investigating regions of the parameter space that are difficult to explore experimentally. To optimize design under uncertainty, an "outer loop" can be included to predict the impact of tunable inputs or control variables (CVs) on a system's quantities of interest (QoIs) [1]. In this approach, CVs are treated as random quantities whose distributions are derived from available experimental data, manufacturing constraints, design criteria, engineering judgment, and/or other domain knowledge. Statistical post-processing of repeated solves of a physics-based model for multiple samples of CVs yields the distributions of QoIs. In the context of optimal and robust design and uncertainty quantification (UQ), this outer loop

* Corresponding authors.
E-mail addresses: ehall001@dundee.ac.uk (E.J. Hall), tartakovsky@stanford.edu (D.M. Tartakovsky).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.jcp.2021.110551
0021-9991/

constitutes a *many query* problem that becomes prohibitively expensive when queries rely solely on direct simulation of physics-based models.

Data-driven surrogate modeling seeks to alleviate this computational cost by constructing a statistical model or emulator for QoIs. Off-the-shelf software such as `TensorFlow` and `PyTorch` facilitates the construction of deep learning surrogates, e.g., deep neural networks (DNNs), from data generated by the underlying physics-based model. This process, which typically involves supervised learning, makes very few assumptions about the nature of the data or data-generating process. This agnosticism makes DNNs suitable for dependent/correlated inputs and non-Gaussian, skewed, multimodal, and/or mutually correlated output QoIs, typically observed in complex real-world systems. While DNNs (e.g., see [2–6]) can tremendously speed up the design pipeline by accelerating and fully automating the prediction of QoIs, they represent *black boxes* that do not shed any light on the form of the function they are approximating. They provide no clear link between this function and the network weights. Moreover, they are non-identifiable, since two DNNs with the same topology, but different weights, can yield very similar outputs for a given set of input data [7].

Global sensitivity analysis (GSA) [8] provides an opportunity to "peek" inside a black-box deep neural networks surrogate and to interpret its predictions by identifying constellations of input parameters that are likely to yield a targeted model response. GSA facilitates exploration of the entire parameter space and quantifies both first-order (individual) and higher-order (interaction) effects that characterize the contribution of variations in CVs to changes in QoIs. Variance-based GSA methods rank input parameters by their contributions to the total variance of a QoI (e.g., Sobol' indices [9] and total effects [10]). Their interpretation is ambiguous when spaces of correlated CVs are large [11,12] and QoIs are highly non-Gaussian [13,14], a situation representative of complex multiscale/multiphysics systems. In contrast, moment-independent GSA approaches are easy to interpret regardless of the nature of the data or data-generating process [15–17]; however, they require knowledge of the CV and QoI distributions or availability of sufficient data to approximate them. DNNs provide sufficient data to capitalize on moment-independent approaches.

A major goal of this study is to harness the synergy between moment-independent GSA and black-box surrogates and to exploit their shared agnosticism to the nature of data and a data-generating process. To this end, we develop UQ emulators for moment-independent GSA, based on information-theoretic quantities, that utilize a DNN surrogate. Information theory has been used to carry out both local sensitivity analysis using Fisher information matrix methods [18–22] and global sensitivity analysis using mutual information [23–28]. Furthermore, information inequalities can be used for robust UQ of QoIs in problems with model uncertainty arising, e.g., when models reflect both physics and sparse or multi-sourced data [29,30].

Our GSA utilizes the concept of *differential mutual information* (MI) [31,32] to compute Mutual Information Sensitivity Indices (MISIs). The latter allow one to interrogate a black-box model by quantifying the dependence of the output QoI with respect to tunable CVs, and to ascertain and rank which CVs are the most influential. As such, MISIs are rigorously the GSA of the black-box surrogate model but correspond indirectly to the original physics-based model and are expected to coincide in the asymptotic limit of "infinite" data. The UQ emulator workflow presented in this work, combining surrogate modeling, global effect rankings, and tools for distinguishing effect rankings, enables an "outer loop" for accelerated design. Moreover, this approach addresses the twin challenges of correlated/dependent CVs and non-Gaussian, skewed, multimodal, and/or mutually correlated QoIs encountered in the modeling of complex systems. These features make MISIs an ideal decision-making tool for simulation-aided design.

Viewed as an uncertainty quantification method for interrogating surrogates, our MI-based GSA is compatible with any black-box model such as physics-informed neural networks [5,33–35] and "data-free" physics-constrained neural networks [4,36–38]. We leverage a Graph-Informed Neural Network (GINN) [6] that is tailored for multiscale physics and systems with correlated CVs. The GINN's ability to generate "big data" allows us to consider higher-order effects due to interactions between the CVs. In turn, the MI-based rankings help to explain the GINN's black-box predictions by placing them in context. We validate these rankings by evaluating response curves along sensitive and insensitive directions and comparing these to their counterparts computed with a physics-based model. This comparison provides a clear interpretation of the GINN's predictions in terms of the physics-based model and enables the use of the GINN to close engineering design loops by deploying it to estimate effect rankings in parameter subspaces yielding optimal QoI values.

Gradient-based methods are often used to interpret DNN predictions. Typically, these methods compute derivatives of the loss function of outputs with respect to inputs using back propagation. Large derivatives with respect to the input, in a suitable metric, are considered influential to the output and, correspondingly, small derivatives are considered less influential. Focusing on a particular class of machine learning tasks related to image classification and pattern recognition and on particular convolutional neural network architectures, these methods use gradients to, e.g., construct saliency maps via local perturbations of the input image [39,40], learn importance features through propagation of activation differences [41], and construct importance scores that accumulate (integrate) gradients over subsets of perturbations [42]. A different approach [43] is to consider interpretability in terms of the filters of a convolutional neural network by adding an MI term in the loss function in order to retain only mutually independent parts of the DNN during training. In contrast, our work focuses on applying GSA to scientific machine learning tasks and uses MI (gradient-free) to interrogate surrogate models where the interpretability is understood in relation to the physics-based model.

In the context of multiscale design [28], we employ the MISI rankings to interpret the surrogate model's predictions by identifying parameter regions that elicit targeted responses and then using new empirical response data predicted by the GINN for those parameter subspaces to refine an existing prototype. We illustrate how MI-based GSA for explainable

DNN predictions enables outer-loop tasks, such as uncertainty quantification and optimal design, to benefit from scientific machine learning. These rankings play a role similar to Shapley values [44], partial dependence plots [45], and individual conditional expectation plots [46] found in the statistical learning and data mining literature. Similar to MISIs, these metrics elucidate the relationship between predicted responses and one or more features in regression models and classifiers based on changes in certain conditional expectations. Unlike these metrics, MISIs depend on distributional—as opposed to moment—information and provide a framework for estimating and ranking higher-order effects, or interactions, among correlated CVs that prove to be crucial for design of complex systems (cf. Fig. 7).

In Section 2, we develop a GSA framework that includes both first- and higher-order MISIs. In Section 3, this methodology is combined with a GINN in the context of a testbed problem related to the design of a supercapacitor. Validation of the MI-based rankings and closure of design loops through subsequent rankings in the reduced parameter space with optimal QoI values are performed in Section 4. In Section 5, we summarize the main conclusions drawn from this study and discuss future work.

2. MISIs for model-agnostic GSA

We consider a model $\mathcal{M}$ of a complex physical system,

$$\boldsymbol{Y} = \mathcal{M}(\boldsymbol{X}), \tag{1}$$

that predicts the response of a collection of QoIs $\boldsymbol{Y} \in \mathbb{R}^q$ to a collection of tunable CVs $\boldsymbol{X} \in \mathbb{R}^p$. The model $\mathcal{M}$ propagates distributions on CVs to distributions on QoIs query-by-query, i.e., it generates one sample response $\boldsymbol{Y}^{(m)}$, $m = 1, \ldots, M$, for each input sample or observation $\boldsymbol{X}^{(m)}$. We place no restrictions on the dependence structure of the CVs or on the nature of their functional relationship to the response.

2.1. First-order effects described by differential MI

Differential MI is a pseudo-distance used in machine learning [47,48] and model selection [49], among others. Quantifying the amount of shared information between $\boldsymbol{V} \in \mathbb{R}^{d_1}$ and $\boldsymbol{W} \in \mathbb{R}^{d_2}$, the differential MI is defined as [31]

$$I(\boldsymbol{V};\boldsymbol{W}) := \iint\limits_{\mathcal{V}\otimes\mathcal{W}} \log\left(\frac{f_{\boldsymbol{V},\boldsymbol{W}}(\boldsymbol{v},\boldsymbol{w})}{f_{\boldsymbol{V}}(\boldsymbol{v}) f_{\boldsymbol{W}}(\boldsymbol{w})}\right) f_{\boldsymbol{V},\boldsymbol{W}}(\boldsymbol{v},\boldsymbol{w}) \mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{w}, \tag{2}$$

where $f_{\boldsymbol{V}}$, $f_{\boldsymbol{W}}$, and $f_{\boldsymbol{V},\boldsymbol{W}}$ denote marginal and joint probability density functions (PDFs) with support $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{V}\otimes\mathcal{W}$, respectively. The differential MI possesses many of the same properties as the discrete MI, including symmetry $I(\boldsymbol{V};\boldsymbol{W}) = I(\boldsymbol{W};\boldsymbol{V})$ and non-negativity $I(\boldsymbol{V};\boldsymbol{W}) \geq 0$ (with equality if and only if $\boldsymbol{V}$ and $\boldsymbol{W}$ are independent). Unlike its discrete counterpart, the differential MI can take on infinite values, e.g., if $\boldsymbol{V} = \boldsymbol{W}$. The following features make the differential MI appropriate for GSA in multiscale design:

(i) its interpretation does not rely on the dependence structure of the CVs,
(ii) its moment independence makes it suitable for a wide range of CV and QoI PDFs, and
(iii) its continuous nature is suitable for analysis of continuous systems.

The first two features enable a model-agnostic implementation, while the last one facilitates uncertainty quantification for downstream computations relying on continuous QoIs.

To describe the first-order effect of a CV $X \in \boldsymbol{X}$ on a target QoI $Y \in \boldsymbol{Y}$, we define a MISI,

$$S_Y(X) := I(X;Y), \tag{3}$$

and interpret it as a measure of the strength of the association between X and Y. A large score indicates that X is a globally influential CV in the design of Y relative to the PDF of $\boldsymbol{X}$. In complex systems, Y is unlikely to be completely described by a single CV X, so the value of $S_Y(X)$ in (3) is likely to remain finite. Since $S_Y(X)$ places equal importance on linear and nonlinear relationships due to the self-equitability of the MI [50], it recovers the rankings of Sobol' indices in the setting of independent CVs $\boldsymbol{X}$, i.e., when Sobol' rankings are justified.

The MISI $S_Y(X)$ in (3) can be estimated from empirical data generated by querying the model $\mathcal{M}$. A plug-in Monte Carlo estimator for $S_Y(X)$,

$$\widehat{S}_Y(X) := \frac{1}{M}\sum_{m=1}^{M} \log\left(\frac{\widehat{f}_{X,Y}(X^{(m)},Y^{(m)};b_X,b_Y)}{\widehat{f}_X(X^{(m)};b_X)\widehat{f}_Y(Y^{(m)};b_Y)}\right), \tag{4}$$

can be computed via joint and marginal kernel density estimators (KDEs) $\widehat{f}$ at input-output data pairs $\{X^{(m)}, Y^{(m)}\}$, $m = 1, \ldots, M$ [51,52]. A Gaussian kernel KDE $\widehat{f}_{\boldsymbol{Z}}$ for an unknown PDF $f_{\boldsymbol{Z}}$ based on M' identically distributed observations $\boldsymbol{Z}^{(1)}, \ldots, \boldsymbol{Z}^{(M')}$ of $\boldsymbol{Z} \in \mathbb{R}^d$ is given by [53]

$$\widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z};\mathbf{b}) := \frac{(2\pi)^{-d/2}}{M'\prod_{j=1}^{d} b_j}\sum_{m=1}^{M'}\prod_{j=1}^{d}\exp\left[-\frac{\left(z_j - Z_j^{(m)}\right)^2}{2b_j^2}\right], \quad \boldsymbol{z}\in\mathbb{R}^d. \tag{5}$$

Among many algorithms for the automated computation of the positive bandwidth parameters $\mathbf{b} = (b_1,\ldots,b_d)^\top$, we chose a direct plug-in bandwidth selector called the improved Sheather–Jones method [54]. To ensure that the joint and marginal KDEs in (4) are defined consistently, i.e., that

$$\int_{\mathcal{X}} \widehat{f}_{X,Y}(x,y;b_X,b_Y)\mathrm{d}x = \widehat{f}_Y(y;b_Y) \quad \text{and} \quad \int_{\mathcal{Y}} \widehat{f}_{X,Y}(x,y;b_X,b_Y)\mathrm{d}y = \widehat{f}_X(x;b_X)\,,$$

we require the smoothing bandwidths for the joint and marginal PDFs to be equal. That is, the bandwidths related to X in $\widehat{f}_X$ and $\widehat{f}_{X,Y}$ must be the same.

The KDE-based direct plug-in estimator $\widehat{S}_Y(X)$ in (4) is easy to implement. However, its computation is not sample-efficient and, hence, unfeasible in the absence of an efficient surrogate; moreover, KDEs are anticipated to fail in high dimensions [51]. In such circumstances, one can deploy alternative strategies for estimating MI, such as a non-parametric k-nearest neighbor algorithm [55] and a non-parametric neural estimation approach suitable for high-dimensional PDFs [56]. Contrary to the discrete MI indices [23,25], non-parametric density estimators, such as (4), introduce no bias associated with a quantization of the QoIs, whose continuous nature may need to be preserved for the purpose of downstream computations.

Remark 1 *(Independent input-output).* The plug-in estimator (4) involves the joint PDFs, i.e., is useful when input-output sample pairs are available. A change of measure in (2) yields an equivalent estimator,

$$\widehat{S}_Y^{\perp}(X) := \frac{1}{M}\sum_{m=1}^{M}\log\left(\frac{\widehat{f}_{X,Y}(X^{(m)},Y^{(m)};b_X,b_Y)}{\widehat{f}_X(X^{(m)};b_X)\widehat{f}_Y(Y^{(m)};b_Y)}\right)\frac{\widehat{f}_{X,Y}(X^{(m)},Y^{(m)};b_X,b_Y)}{\widehat{f}_X(X^{(m)};b_X)\widehat{f}_Y(Y^{(m)};b_Y)}\,, \tag{6}$$

that is suitable for independent samples from the input and output distributions as the expectation is no longer with respect to the joint PDF but with respect to each marginal PDF (cf. [28]).

2.2. Higher-order effects described by conditional differential MI

For large spaces of possibly correlated CVs, it is of interest to also consider the impact of interactions among subsets of CVs on a given QoI. To describe the effects of pairwise interactions between $X_1, X_2 \in \boldsymbol{X}$ on a target QoI Y, we define a second-order MISI,

$$S_Y(X_1,X_2) := I(X_1;X_2\mid Y), \tag{7}$$

in terms of the conditional differential MI,

$$I(\boldsymbol{V};\boldsymbol{W}\mid\boldsymbol{U}) := \iiint_{\mathcal{U}\otimes\mathcal{V}\otimes\mathcal{W}} \log\left(\frac{f(\boldsymbol{u})f(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})}{f(\boldsymbol{u},\boldsymbol{w})f(\boldsymbol{u},\boldsymbol{v})}\right) f(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\mathrm{d}\boldsymbol{w}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{u}. \tag{8}$$

The latter represents the MI between $\boldsymbol{V}$ and $\boldsymbol{W}$ conditioned on $\boldsymbol{U}$ that we express in terms of joint and marginal PDFs.[2] The conditional MI in (8) is related to the MI in (2) through the chain rule,

$$I(V_1,V_2,\ldots,V_k;W) = \sum_{i=1}^{k} I(V_i;W\mid V_{i-1},V_{i-2},\ldots,V_1), \tag{9}$$

for $V_1,V_2,\ldots,V_k\in\boldsymbol{V}$ and $W\in\boldsymbol{W}$ where zero-indexed sets in the conditioning are empty. To see that (7) captures only the second-order effects, we note that $I(X_1,X_2;Y)$ describes the full effect of the pair (X_1,X_2) on Y. According to (9), the full second-order effect is expressed as

$$I(X_1,X_2;Y) = I(X_1;Y) + I(X_2;Y) - I(X_1;X_2) + I(X_1;X_2\mid Y), \tag{10}$$

which includes first-order effects $I(X_1;Y)$ and $I(X_2;Y)$, while $I(X_1;X_2)$ captures the interaction between X_1 and X_2 (the latter term vanishes if X_1 and X_2 are independent). The remaining conditional differential MI in (10) describes the desired second-order effect.

[2] Here and in the sequel we suppress the labels on densities when the distribution is clear from the context.

A plug-in Monte Carlo estimator for the second-order index (7) is

$$\widehat{S}_Y(X_1, X_2) := \frac{1}{M}\sum_{m=1}^{M} \log\left(\frac{\widehat{f}(Y^{(m)}; b_Y)\widehat{f}(X_1^{(m)}, X_2^{(m)}, Y^{(m)}; b_{X_1}, b_{X_2}, b_Y)}{\widehat{f}(X_1^{(m)}, Y^{(m)}; b_{X_1}, b_Y)\widehat{f}(X_2^{(m)}, Y^{(m)}; b_{X_2}, b_Y)}\right), \tag{11}$$

based on input-output triples $(X_1^{(m)}, X_2^{(m)}, Y^{(m)})$, $m = 1, \ldots, M$. The plug-in estimator is justified in the context of surrogate modeling and is easy to implement using KDEs (5) with suitably equalized bandwidths. It can be built from the same sample data used to evaluate (11) and comes with the same caveats.

A kth-order MISI (with $k > 2$) is defined as

$$S_Y(X_1, \ldots, X_k) := I(X_1; X_2; \ldots; X_k \mid Y). \tag{12}$$

It quantifies the impact of the interaction among the collection of variables $\{X_1, \ldots, X_k\} \subset \boldsymbol{X}$ on Y. The conditional multivariate differential MI is defined inductively,

$$I(X_1; X_2; \ldots; X_k \mid Y) = I(X_2; \ldots; X_k \mid X_1, Y) - I(X_2; \ldots; X_k \mid Y). \tag{13}$$

It is symmetric with respect to permutation of the variables X_j, with $1 \le j \le k$. For example, the third-order effect of the interactions among the triple X_1, X_2, and X_3 on Y is given by the third-order MISI,

$$S_Y(X_1, X_2, X_3) = I(X_1; X_2; X_3 \mid Y) = I(X_2; X_3|X_1, Y) - I(X_2; X_3 \mid Y). \tag{14}$$

The conditional multivariate MI in (13) and, hence, S_Y in (12) can be either positive or negative. They are related to a conditional form of the "interaction information" [57] and the "co-information" [58]. Instead of interpreting such higher-order ($k > 2$) effects, we focus on algorithms for ranking the first- and second-order effects with appropriate confidence intervals (cf. Fig. 4 for first-order and Fig. 5 for second-order effect rankings).

2.3. Algorithms for MISI ranking with confidence

We assume the availability of a surrogate model $\mathcal{M}$ for generating large amounts of response data. While based on slightly different theoretical approaches, the two algorithms described below enable the construction of first- and higher-order effect rankings for a given QoI with a focus on the generation of associated confidence intervals. When considering effect rankings, confidence intervals are an important tool for distinguishing closely-ranked pairs of CVs, particularly if some automation of the process is desirable. True to form, each confidence interval is a random interval, estimated from a set of observed responses, that may or may not contain the true (theoretical) response value.

2.3.1. Algorithm 1: compute MISIs with confidence intervals adjusted for ranking

Algorithm 1 (see the pseudocode) constructs the plug-in estimators $\widehat{S}_Y(X_j)$ in (4) with confidence intervals selected such that pairwise comparisons of the MISIs and their accompanying intervals determine the effect ranks. For fixed Y, we order the $\widehat{S}_Y(X_j)$ as

$$\widehat{S}_Y(X_{j_1}) > \widehat{S}_Y(X_{j_2}) > \cdots > \widehat{S}_Y(X_{j_p}). \tag{15}$$

Then, the ranked first-order MISI estimators are

$$(\widehat{\theta}_1, \ldots, \widehat{\theta}_p) := \left(\widehat{S}_Y(X_{j_1}), \ldots, \widehat{S}_Y(X_{j_p})\right), \tag{16}$$

where the individual (additive) effects of the CVs are arranged in order of importance, from the greatest ($\widehat{\theta}_1$) to the least ($\widehat{\theta}_p$). The (true) rank r_j of CV X_j is estimated by the plug-in quantity,

$$\widehat{r}_j := p - \#\{\widehat{S}_Y(X_i) < \widehat{S}_Y(X_j), \quad i = 1, \ldots, p\}, \tag{17}$$

where $\#\{\cdot\}$ denotes the cardinality of the set, such that

$$\widehat{\theta}_{\widehat{r}_j} = \widehat{S}_Y(X_j). \tag{18}$$

Since the MISI $S_Y(X)$ in (3) is a global measure of sensitivity, (16) represents a global ranking of the first-order effect of each CV relative to the distribution of $\boldsymbol{X}$.

One could approximate the standard $100 \cdot (1-\alpha)\%$ confidence interval for θ_k with

$$(\widehat{\theta}_k - z[\alpha/2]\widehat{\sigma}_k, \quad \widehat{\theta}_k + z[\alpha/2]\widehat{\sigma}_k), \qquad \widehat{\sigma}_k := \sqrt{\mathrm{Var}\left[\widehat{\theta}_k\right]} = \sqrt{\mathrm{Var}\left[\widehat{S}_Y(X_{j_k})\right]}, \tag{19}$$

Algorithm 1: Compute first-order MISIs with confidence intervals adjusted for ranking.

input : $\mathcal{M}$ ▷ Surrogate model (1)
input : $(X_1^{(1)},\dots,X_p^{(1)}),\dots,(X_1^{(M)},\dots,X_p^{(M)})$ ▷ M independent samples of CVs
input : $\bar{\gamma}$, TOL ▷ Non-overlap sig. and tolerance
output: $(\widehat{\theta}_1 \pm z[\beta/2]\widehat{\sigma}_1,\dots,\widehat{\theta}_p \pm z[\beta/2]\widehat{\sigma}_p)$ ▷ Rankings with adjusted intervals (23)
begin
Compute target QoI observations
for $m \in \{1,\dots,M\}$ **do**
$Y^{(m)} \leftarrow \mathcal{M}(X_1^{(m)},\dots,X_p^{(m)})$
Compute and rank first-order MISI estimators
for $j \in \{1,\dots,p\}$ **do**
$G(x,y) \leftarrow \log\big[\widehat{f}_{X_j,Y}(x,y)\big] - \log\big[\widehat{f}_{X_j}(x)\widehat{f}_Y(y)\big]$
$\widehat{S}_Y(X_j) \leftarrow \frac{1}{M}\sum_{1\le m\le M} G(X_j^{(m)}, Y^{(m)})$ ▷ First-order MISI (4)
for $j \in \{1,\dots,p\}$ **do**
$\widehat{r}_j \leftarrow p - \#\{\widehat{S}_Y(X_i) < \widehat{S}_Y(X_j),\ i=1,\dots,p\}$ ▷ Rank of jth CV (17)
$\widehat{\theta}_{\widehat{r}_j} \leftarrow \widehat{S}_Y(X_j)$ ▷ Ranked MISI (18)
$\widehat{\sigma}_{\widehat{r}_j} \leftarrow \big(\mathrm{Var}[\widehat{S}_Y(X_j)]\big)^{1/2}$ ▷ MISI standard error (19)
Compute comparison-adjusted confidence intervals with average type I error $\bar{\gamma}$
for $k,l \in \{1,\dots,p\}$ **do**
$s_{kl} \leftarrow (\widehat{\sigma}_k + \widehat{\sigma}_l)/(\widehat{\sigma}_k^2 + \widehat{\sigma}_l^2)^{1/2}$
$f(z) \leftarrow \bar{\gamma} - \frac{4}{p(p-1)}\sum_{1\le k<l\le p}(1-\Phi(z s_{kl}))$ ▷ Average non-overlap sig. (22)
$f'(z) \leftarrow \frac{4}{p(p-1)}\sum_{1\le k<l\le p}\varphi(z s_{kl})\cdot s_{kl}$ ▷ Normal PDF $\varphi = \Phi'$
$z_0 \leftarrow \Phi^{-1}(1-\bar{\gamma}/2)\cdot(\widehat{\sigma}_1^2+\widehat{\sigma}_2^2)^{1/2}/(\widehat{\sigma}_1+\widehat{\sigma}_2)$
$i \leftarrow 0$
while $\mathrm{err}_i < \mathrm{TOL}$ **do**
$z_{i+1} \leftarrow z_i - f(z_i)/f'(z_i)$ ▷ Newton-Raphson iterations
$\mathrm{err}_i \leftarrow |z_{i+1} - z_i|/z_i$
$i \leftarrow i+1$
$z[\beta/2] \leftarrow z_{i+1}$ ▷ Difference level (23)
return $(\widehat{\theta}_1,\dots,\widehat{\theta}_p)$, $(\widehat{\sigma}_1,\dots,\widehat{\sigma}_p)$, $z[\beta/2]$

where $\Phi(z[\alpha/2]) = 1-\alpha/2$ for the standard normal cumulative distribution function Φ, and $\widehat{\sigma}_k$ is the standard error.[3] However, this confidence interval does not readily distinguish the rankings. The fact that the $100\cdot(1-\alpha)\%$ confidence intervals for the ranked effects θ_k and θ_l, with $k \neq l$, fail to overlap does not necessarily mean that the difference in the rankings is statistically significant at the α level. Following [59], Algorithm 1 reports confidence intervals with comparison-adjusted widths, such that the non-overlap significance level meets a given threshold on average. Assuming normality and independence of $\widehat{\theta}_k$ and $\widehat{\theta}_l$, the confidence intervals at level β do not overlap if

$$|\widehat{\theta}_k - \widehat{\theta}_l| > z[\beta/2](\sigma_k + \sigma_l). \tag{20}$$

Inequality (20) holds with probability $1-\gamma_{kl}$, where the pairwise non-overlap significance level γ_{kl} is given by

$$\gamma_{kl} := 2 - 2\Phi\Big(z[\beta/2]\frac{\sigma_k+\sigma_l}{\sigma_{kl}}\Big), \qquad \sigma_{kl} := \sqrt{\sigma_k^2+\sigma_l^2}. \tag{21}$$

We select the level β to ensure that the average of the pairwise errors γ_{kl} over all $1 \le k < l \le p$ is at a predefined level $\bar{\gamma}$ (set to $\bar{\gamma} = 0.01$), i.e.,

$$\bar{\gamma} - \frac{2}{p(p-1)}\sum_{1\le k<l\le p}\gamma_{kl} = 0. \tag{22}$$

The level β, for which (22) holds, is found via Newton–Raphson iteration (cf. Algorithm 1) using sample estimates for the standard errors. For each ranked effect θ_k, the approximate confidence interval at level β is

$$\big(\widehat{\theta}_k - z[\beta/2]\widehat{\sigma}_k,\ \widehat{\theta}_k + z[\beta/2]\widehat{\sigma}_k\big), \qquad 1 \le k \le p; \tag{23}$$

the error bars indicate that the non-overlap significance level is $\bar{\gamma}$ on average. The intervals (23) provide the visual comparison of pairwise effects with clear interpretation: overlapping/non-overlapping intervals imply that the associated ranks are indistinguishable/distinguishable.

[3] Here, $z[\alpha]$ denotes the value such that $\Phi(z[\alpha]) = 1-\alpha$, i.e., the value such that the standard normal PDF in the right-hand tail is equal to α.

With a slight modification, Algorithm 1 can be used to rank the second-order MISI estimators (7),

$$(\widehat{\zeta}_1, \ldots, \widehat{\zeta}_{p'}) := \left(\widehat{S}_Y(X_{j_1}, X_{k_1}), \ldots, \widehat{S}_Y(X_{j_{p'}}, X_{k_{p'}})\right), \tag{24}$$

where $p' := p!/(2!(p-2)!)$ is the number of pairs (X_j, X_k) with $j < k$ that can be formed from $\boldsymbol{X}$. The computation of (24) replaces that of the first-order indices in Algorithm 1; the computation of the comparison-adjusted confidence intervals for ζ_i, $i = 1, \ldots, p'$, proceeds analogously to that of the first-order confidence intervals (23) with the non-overlap significance averaged over all pairs $(\widehat{\zeta}_k, \widehat{\zeta}_l)$ with $1 \leq k \leq l \leq p'$.

2.3.2. Algorithm 2: rank MISIs with percentile confidence intervals

Algorithm 2 (see the pseudocode) constructs non-parametric estimates with confidence intervals for the unknown rankings r_k directly from the sampling distribution. In contrast to Algorithm 1 which uses normality theory, the present method builds a distribution for each rank by repeated observation of the MISIs using the surrogate model. For each $X_j \in \boldsymbol{X}$, we compute MISI replications,

$$\widehat{s}_j^{(n)} := \widehat{S}_Y^{(n)}(X_j), \quad n = 1, \ldots, N, \tag{25}$$

from input-output sample pairs $(X_j^{(m)}, Y^{(m)})$, $m = 1, \ldots, M$. That is, for each replication, we generate M new input-output pairs $(\boldsymbol{X}, Y)$ using the surrogate model and compute (25) for every $j = 1, \ldots, p$ with these observations. From these replications, we use (17) to compute rank replications $\widehat{r}_k^{(n)}$, $k = 1, \ldots, p$. The rank estimators are

$$\widehat{r}_k = \frac{1}{N}\sum_{n=1}^{N} \widehat{r}_k^{(n)}, \tag{26}$$

and the corresponding δ percentile confidence intervals are

$$(\widehat{r}_k[\delta/2], \widehat{r}_k[1-\delta/2]), \qquad 1 \leq k \leq p. \tag{27}$$

The equal-tail percentiles $\widehat{r}_k[\delta/2]$ and $\widehat{r}_k[1-\delta/2]$ are estimated from the replications in the spirit of the bootstrap percentile confidence intervals [60].

The computational burden of Algorithm 2 is greater than that of Algorithm 1, since the estimator is computed for each of the N replications. Yet, this method is non-parametric and its results are anticipated to be more easily interpretable for large numbers of CVs and higher-order effect calculations.

3. Interrogation of black-box surrogates using MISIs

As highlighted in the introduction, our MI-based approach to GSA is applicable to any black-box surrogate model. To illustrate the ability of MI-based GSA to deal with correlated CVs, for which variance-based GSA approaches are of limited value, we combine it with a GINN, a domain-aware DNN surrogate introduced in [6] to overcome computational bottlenecks in complex multiscale and multiphysics systems. We consider two applications of interest in energy storage: the Langmuir adsorption model and a multiscale formulation of the dynamics of electrical double-layer supercapacitors with nanoporous electrodes.

3.1. Langmuir adsorption model

To highlight the application of the MISI rankings, we first consider a simplified Langmuir adsorption model. This setting allows us to compare GINN-based rankings to ground truth rankings computed from the physics-based model. The Langmuir biomolecular adsorption model is widely used to describe competitive dissociative adsorption of two species on a catalyst surface [61], e.g., hydrogen oxidation in fuel cells [62–65]. The coverage dynamics ϑ_A and ϑ_B of species A and B, respectively, are found by long-time integration of the system of nonlinear ODEs,

$$\frac{\mathrm{d}\vartheta_A}{\mathrm{d}t} = k_A^{\mathrm{ads}} P_A (1 - \vartheta_A - \vartheta_B)^2 - k_A^{\mathrm{des}} \vartheta_A^2, \qquad \vartheta_A(0) = \vartheta_A^0, \tag{28a}$$

$$\frac{\mathrm{d}\vartheta_B}{\mathrm{d}t} = k_B^{\mathrm{ads}} P_B (1 - \vartheta_A - \vartheta_B)^2 - k_B^{\mathrm{des}} \vartheta_B^2, \qquad \vartheta_B(0) = \vartheta_B^0, \tag{28b}$$

where P denotes partial pressure, k^{ads} denotes the adsorption rate constant, and k^{des} denotes the desorption rate constant. The system (28) has a steady-state solution

$$\vartheta_A = \frac{(K_A P_A)^{1/2}}{1 + (K_A P_A)^{1/2} + (K_B P_B)^{1/2}}, \tag{29a}$$

$$\vartheta_B = \frac{(K_B P_B)^{1/2}}{1 + (K_A P_A)^{1/2} + (K_B P_B)^{1/2}}, \tag{29b}$$

Algorithm 2: Rank first-order MISIs with percentile confidence intervals.

input : $\mathcal{M}$ ▷ Surrogate model (1)
input : N ▷ Number of replications
input : M ▷ Number of observations per replication
input : $\delta \quad (\delta_l := \delta/2, \delta_u := 1-\delta_l)$ ▷ Equal tail percentile level
output: $(\widehat{r}_1, \dots, \widehat{r}_p)$ ▷ Ranks (26)
output: $(\widehat{r}_1[\delta_l], \widehat{r}_1[\delta_u]), \dots, (\widehat{r}_p[\delta_l], \widehat{r}_p[\delta_u])$ ▷ Percentile confidence intervals (27)

begin
  Compute N replications of MISI first-order effect ranks
  for $n \in \{1, \dots, N\}$ **do**
    Generate M input-output samples
    for $m \in \{1, \dots, M\}$ **do**
      sample $(X_1^{(m)}, \dots, X_p^{(m)})$
      $Y^{(m)} \leftarrow \mathcal{M}(X_1^{(m)}, \dots, X_p^{(m)})$
    Calculate one replication of each MISI
    for $j \in \{1, \dots, p\}$ **do**
      $G(x, y) \leftarrow \log\left[\widehat{f}_{X_j,Y}(x, y)\right] - \log\left[\widehat{f}_{X_j}(x)\widehat{f}_Y(y)\right]$
      $\widehat{s}_j^{(n)} \leftarrow \frac{1}{M}\sum_{1\le m\le M} G(X_j^{(m)}, Y^{(m)})$ ▷ nth replication MISI (25)
    Calculate one replication of each rank
    for $j \in \{1, \dots, p\}$ **do**
      $\widehat{r}_j^{(n)} \leftarrow p - \#\{\widehat{s}_i^{(n)} < \widehat{s}_j^{(n)},\ i = 1, \dots, p\}$ ▷ nth replication rank (17)
  Compute rank statistics from replications
  for $j \in \{1, \dots, p\}$ **do**
    $\widehat{r}_j \leftarrow \frac{1}{N}\sum_{1\le n\le N} \widehat{r}_j^{(n)}$
    $\widehat{r}_j[\delta_l] \leftarrow$ **quantile** $(\{\widehat{r}_j^{(1)}, \dots, \widehat{r}_j^{(N)}\}, \delta_l)$
    $\widehat{r}_j[\delta_u] \leftarrow$ **quantile** $(\{\widehat{r}_j^{(1)}, \dots, \widehat{r}_j^{(N)}\}, \delta_u)$
  return $(\widehat{r}_1, \dots, \widehat{r}_p), (\widehat{r}_1[\delta_l], \widehat{r}_1[\delta_u]), \dots, (\widehat{r}_p[\delta_l], \widehat{r}_p[\delta_u])$ ▷ cf. Algorithm 1 **output**

where $K = k^{\text{ads}}/k^{\text{des}}$ denotes an equilibrium constant. Our goal is to quantify the uncertainty in predicted coverages at equilibrium as functions of E_A and E_B, the changes in adsorption energy of each species,

$$\vartheta_A = \vartheta_A(E_A, E_B) \quad \text{and} \quad \vartheta_B = \vartheta_B(E_A, E_B)\,. \tag{30}$$

These QoIs are highly nonlinear functions of E_A and E_B; the equilibrium constants are expressed by the Arrhenius law,

$$K_A = \exp\left(-\frac{G_A}{k_B T}\right)(P_A + P_B)^{-1}\,, \tag{31a}$$

$$K_B = \exp\left(-\frac{G_B}{k_B T}\right)(P_A + P_B)^{-1}\,, \tag{31b}$$

where k_B is the Boltzmann constant; T is the temperature; and the Gibbs free energy of adsorption,

$$G_A \propto -2E_A \quad \text{and} \quad G_B \propto -2E_B, \tag{32}$$

are, according to density functional theory, given by changes in adsorption energy plus additional terms that are not a function of the catalyst surface [66]. The Gibbs free energy of adsorption is measured "experimentally" for a variety of metal catalyst surfaces via density functional theory, that is, quantum computations for actual metals. In [66], where competitive dissociative adsorption of H_2 and O_2 is considered, the changes in adsorption energy are observed to be correlated and are described by a linear model.

We generate synthetic data, given in Fig. 1, according to the following program. We assume that E_A is drawn from a Gamma distribution (with shape α and scale β)

$$E_A \sim \text{Gamma}(\alpha = 33, \beta = 0.0870)\,. \tag{33a}$$

We then sample correlated E_B according to the model

$$E_B = -2 + 2.5E_A + \epsilon\,, \tag{33b}$$

where each $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is a mean-zero noise. Assuming unit partial pressure, we consider the simplified statistical-mechanics approximation

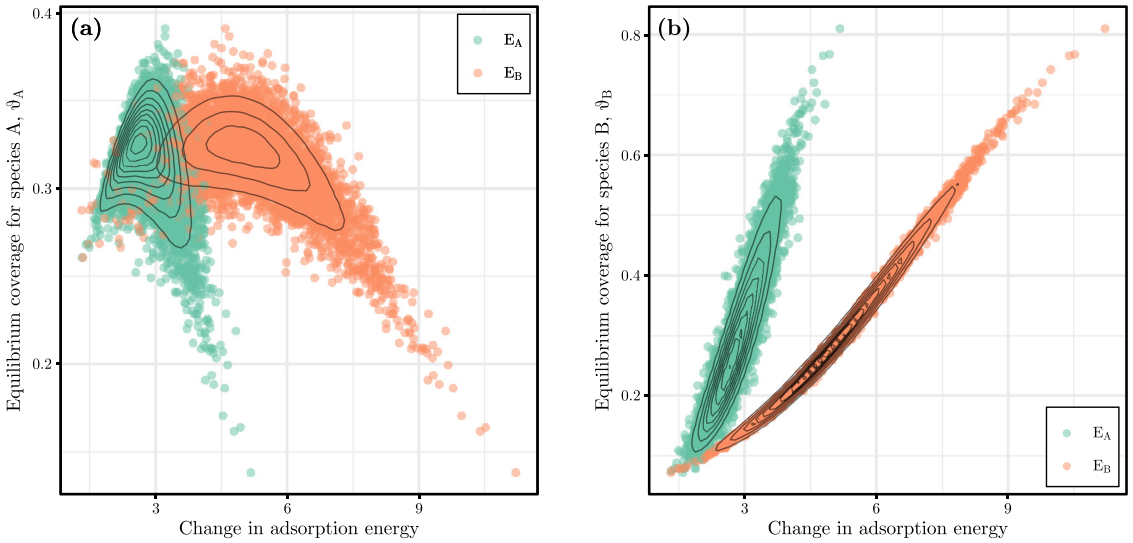

Fig. 1. Response surfaces for the QoIs (a) ϑ_A and (b) ϑ_B (equilibrium coverages (30)) based on inputs E_A and E_B (changes in adsorption energy (33)) simulated using a physics-based model, i.e., the simplified statistical-mechanics approximation (34) and the steady-state solution (29) for the Langmuir model (28). The equilibrium coverages are a nonlinear function of the changes in adsorption energy arising from Arrhenius law (31) and the calculation of the Gibbs free energy of adsorption (32).

$$G_A = -2E_A + 5\,, \qquad G_B = -2E_B + 10\,, \tag{34a}$$

$$K_A = \exp(-G_A/2)\,, \qquad K_B = \exp(-G_B/2)\,. \tag{34b}$$

Taking advantage of the closed form equilibrium solution (29), we generate a corpus of $N_{\text{sam}} = 1.25 \times 10^6$ input-output data $(X_{E_A}, X_{E_B}, Y_{\vartheta_A}, Y_{\vartheta_B})$ that allows us to construct "ground truth" rankings. As the input layer contains correlations, we consider a GINN [6] surrogate and learn two models: a fully resolved GINN that is trained and tested on the full corpus and a sparse GINN that is trained on a subset of 1.25×10^4 data points. For simplicity of presentation we omit the full details on training and testing (see e.g. Section 3.3.1).[4]

In Fig. 2, we demonstrate the first-order MISI obtained using Algorithm 1 and associated ranks calculated according to Algorithm 2, both with 95% confidence intervals. In panels (a) and (c), estimated MISIs (4) based on $M = 10^6$ samples from the physics-based model form a "ground truth". These MISIs are compared to the surrogate indices estimated using $M = 10^6$ predictions from both the resolved and sparse GINN surrogate models. Although there is some bias in the learned surrogate models, the estimates based on large numbers of predictions are largely consistent with the ground truth rankings. In contrast, the estimate based on limited samples from the sparse GINN surrogate has larger confidence intervals (Figs. 2a and 2c), and the bias in the sparse GINN surrogate severely underestimates the MISI value $I(E_B; \vartheta_B)$ (Fig. 2c). All the estimated rankings resulting from Algorithm 2 are consistent and have vanishingly small confidence intervals (Figs. 2b and 2d). The ground truth ranking and the resolved GINN estimates were computed with $N = 10^2$ replications of $M = 10^3$ samples and the sparse GINN model (trained on restricted data) was computed with $N = 10$ replications of $M = 10^2$ samples. Taken together, this experiment demonstrates that, although bias or error in the trained GINN can impact predictions, the surrogate-driven mutual information-based GSA provides useful rankings. Although evaluating the information flow between the physics-based model and its surrogate is beyond the scope of the present work, we point to an analysis [67] of the information-theoretic limits of DNNs, assuming the inputs have a known model and are conditionally independent.

3.2. Multiscale supercapacitor dynamics

We consider an electrical double-layer capacitor [68], whose electrodes are made of a conductive hierarchical nanoporous carbon structure [69]. Electrolyte (an ionized fluid) fills the nanopores and contributes to the formation of the electrical double layer at the electrolyte-electrode interface (see, e.g., Fig. A8 in [6]). Identification of an optimal pore structure of the carbon electrodes holds the promise of manufacturing supercapacitors which boast high power and high energy density [70,71]. This and other advancements, such as lower self-discharge electrolytes [72] for enhanced long-term energy storage, position electrical double-layer capacitors as a viable replacement of Li-ion batteries in electric vehicles or personal electronic devices. Attractive features of electrical double-layer capacitors are their shorter charging times, longer service life, and reduced reliance on hazardous materials [73].

[4] We consider an architecture comprised of two hidden layers, each with 50 neurons and ReLU (Rectified Linear Unit) activations, and a linear output layer. In both cases, the training is set 80% of the available data with remaining 20% reserved for testing and the training and testing tolerance is $O(10^{-5})$.

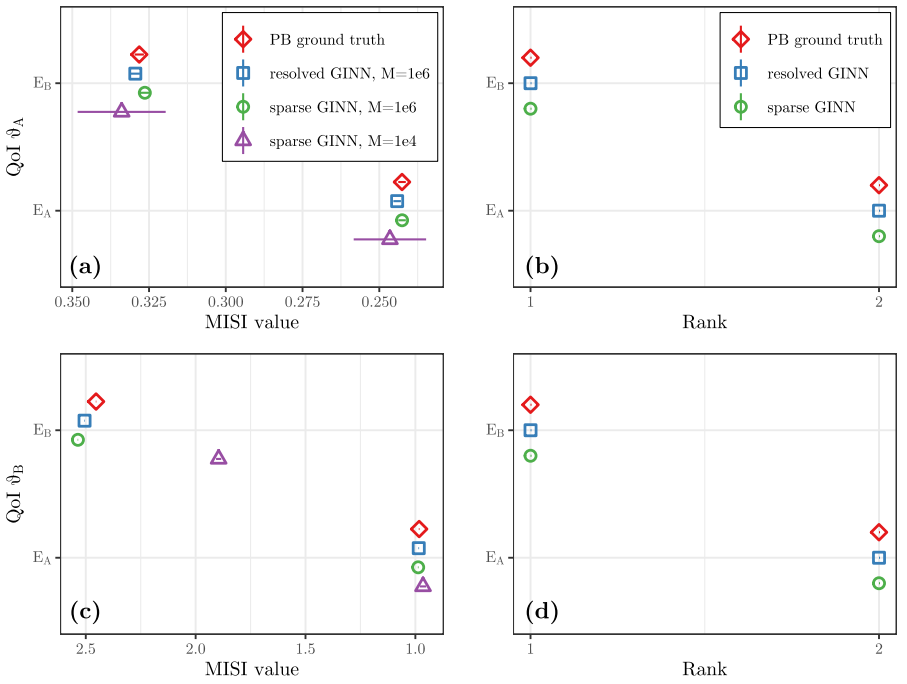

Fig. 2. Plug-in Monte Carlo estimators of the first-order MISIs (4) for both a fully resolved GINN surrogate (learned from all available data) and a sparse GINN surrogate (learned from a restricted data set) are compared to a physics-based ground truth. For Algorithm 1 (left column), we observe that the large-sample estimates based on $M = 10^6$ samples are largely consistent. For Algorithm 2 (right column), we observe that surrogate-driven mutual information-based GSA provides useful and consistent rankings. Note in the (b), (c), and (d) that the 95% confidence intervals are vanishingly small.

Two macroscopic QoIs affect electrical double-layer capacitor performance: effective electrolyte conductivity κ^{eff} and transference number t_+ (fraction of the current carried by the cations), such that $\boldsymbol{Y} := \{Y_{\kappa^{\text{eff}}}, Y_{t_+}\}$. These QoIs are influenced by seven tunable CVs: the electrode surface (fluid-solid interface) potential φ_Γ, initial ion concentration c_{in}, temperature T, porosity ω, (half) pore throat size l_{por}, solid radius r, and Debye length λ_D, such that $\boldsymbol{X} := \{X_{\varphi_\Gamma}, X_{c_{\text{in}}}, X_T, X_\omega, X_{l_{\text{por}}}, X_r, X_{\lambda_D}\}$. A physics-based model $\mathcal{M}$, derived in [74] via homogenization, relates the inputs $\boldsymbol{X}$ to the outputs $\boldsymbol{Y}$. This model involves closure variables (second-order tensors) $\boldsymbol{\chi}_\pm$ and electrical double-layer potential φ_{EDL}, whose determination is expensive and constitutes computational bottlenecks $\boldsymbol{Z} := \{Z_{\boldsymbol{\chi}_\pm}, Z_{\varphi_{\text{EDL}}}\}$. Optimal design of the nanoporous electrodes in electrical double-layer capacitors involves the tuning of the CVs $\boldsymbol{X}$ to elicit changes in the QoIs $\boldsymbol{Y}$.

The complex nonlinear and multiscale relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ makes this a challenging engineering design problem and allows us to highlight the features (i)–(iii) of the MI-based GSA. The joint PDF of the random CVs $\boldsymbol{X}$ systematically quantifies uncertainties and errors arising in the physics-based representation. This key quantity for decision support is captured by a Bayesian Network (BN) [6,28], which encodes both physical relationships and available domain knowledge (Fig. 3). The resulting probabilistic physics-based model $\mathcal{M}$, referred to as a BN PDE, propagates the joint PDF of $\boldsymbol{X}$, i.e., a structured prior, via $\boldsymbol{Z}$ to $\boldsymbol{Y}$ following the conditional relationships in the BN. As in [6], we assume both the CVs X_T, $X_{c_{\text{in}}}$, X_r, and X_ω to be independent and their prior PDFs to be uniform on an interval of $\pm 35\%$ (for X_T and $X_{c_{\text{in}}}$) or $\pm 25\%$ (for X_r and X_ω) around their respective baseline values (see Table 1 reproduced from [6]),

$$X_i \mid \theta_i \sim \text{Uniform}([\theta_i^{\min}, \theta_i^{\max}]), \qquad i = T, c_{\text{in}}, r, \omega; \tag{35}$$

where the hyper-parameters $\theta_i = \{\theta_i^{\min}, \theta_i^{\max}\}$ represent the left and right endpoints of the support intervals. The remaining CVs, X_{λ_D}, X_{φ_Γ} and $X_{l_{\text{por}}}$, are related conditionally (Fig. 3) to these independent inputs through the physical relations,

$$\lambda_D = \sqrt{\frac{RT\mathcal{E}}{2F^2 z^2 \nu c_{\text{in}}}} \quad [\text{nm}], \tag{36a}$$

$$\varphi_\Gamma = \frac{V}{2} - \varphi_{\text{ecm}} - \frac{1}{C_{\text{H}}}\sqrt{4\mathcal{E}RTz^2 c_{\text{in}}}\sqrt{\cosh\left(\frac{e\varphi_\Gamma}{k_B T}\right) - \cosh\left(\frac{e\varphi_{\min}}{k_B T}\right)} \quad [\text{V}], \tag{36b}$$

$$l_{\text{por}} = -r + 0.5\sqrt{4r^2 + 4r^2\left[\frac{\pi}{4\cdot(1-\omega)} - 1\right]} \quad [\text{nm}]\,. \tag{36c}$$

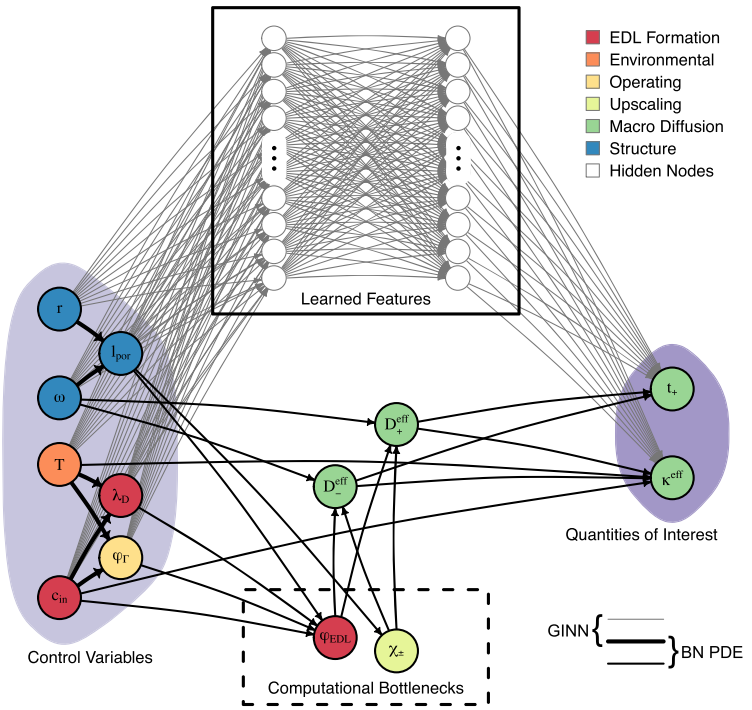

Fig. 3. Visualization of the BN PDE (lower route) and GINN surrogate (upper route) for a multiscale model of EDL supercapacitor dynamics. The BN encodes conditional relationships between the model variables (both inter- and intrascale) and systematically includes domain knowledge into the physics-based model, ensuring the resulting BN PDE makes physically sound predictions. The GINN takes identical inputs $\boldsymbol{X}$ (i.e., structured priors on CVs) to those of the BN PDE, but overcomes the latter's computational bottlenecks $\boldsymbol{Z}$ (dashed box) by replacing them with learned features (solid box) in a DNN to predict the QoIs $\boldsymbol{Y}$. The nodes in the hidden layers of the GINN make it a black box.

Table 1
Statistics of the uniform PDFs of the independent CVs in (35) (from [6]).

Variable label	$\theta^{\min}$	$\theta^{\max}$	Mean/Baseline	Variation	Units
T	208	432	320	±35%	K
c_{in}	0.52	1.08	0.80	±35%	mol/l
r	1.05	1.75	1.40	±25%	nm
ω	0.5025	0.8375	0.6700	±25%	–

Here, $F = 96485$ C/mol is the Faraday constant, R [J/(mol K)] is the gas constant, k_{B} [J/K] is the Boltzmann constant, T [K] is the temperature, $\mathcal{E}$ [F/m] is the absolute permittivity of the solvent, z [-] is the ion charge (valence), ν [-] is the dissociation constant, V is the external voltage, e [C] is the elementary charge, φ_{ecm} is the electrocapillary maximum, c_{in} [mol/l] is the initial ion concentration, and $\varphi_{\min}$ is the mid-plane potential (see [6] for further details). The PDFs of these dependent CVs are estimated by sampling the uniform distributions (35) and computing a corresponding observation via (36). Hence, the physics of the problem induces the correlations between the CVs represented by the conditional relationships in Fig. 3.

3.3. GINNs: surrogate models for multiscale physics

GINNs [6] are domain-aware surrogates for a broad range of complex physics-based models. In the context of electrical double-layer capacitors, a GINN can be used to accelerate the propagation of uncertainty from structured priors on CVs $\boldsymbol{X}$ to distributions of QoIs $\boldsymbol{Y}$ by replacing the computational bottlenecks $\boldsymbol{Z}$ in the BN PDE (the dashed boxed nodes of the BN in Fig. 3) with the GINN's hidden layers. In so doing, it alleviates the cost of computing the QoIs $\boldsymbol{Y}$, which includes bypassing the need to compute the effective diffusion coefficients of the cations (D_+^{eff}) and anions (D_-^{eff}) according to

$$\kappa^{\text{eff}} := \nu z^2 \frac{F^2 c_{\text{in}}}{RT}(D_+^{\text{eff}} + D_-^{\text{eff}}) \quad [\text{mS/cm}]\,, \tag{37a}$$

$$t_+ := \frac{D_+^{\text{eff}}}{D_+^{\text{eff}} + D_-^{\text{eff}}} \quad [-]\,. \tag{37b}$$

3.3.1. GINN construction

The workflow for building the GINN surrogate for the supercapacitor dynamics is summarized as follows. Details, including the procedures for training and testing the GINN using the BN PDE, can be found in [6].

1. **Data generation (BN PDE):** Generate $N_{\text{sam}} = 4 \times 10^3$ input-output samples by drawing the inputs from the structured priors on $\boldsymbol{X}$ and computing the corresponding responses $\boldsymbol{Y}$ with the BN PDE, and select $N_{\text{train}} = 0.75N_{\text{sam}}$ training samples and $N_{\text{test}} = 0.25N_{\text{sam}}$ test samples from this data set.
2. **Training:** Using the N_{train} input-output pairs and `TensorFlow 2`, train with 100 epochs a fully connected NN comprising:
   a) an input layer consisting of the seven CVs $\boldsymbol{X}$,
   b) two hidden layers each consisting of 100 neurons,
   c) an output layer consisting of the two QoIs $\boldsymbol{Y}$,
   d) application of the ReLU (Rectified Linear Unit) activation function, and
   e) a given training error tolerance of 10^{-4}.
3. **Testing:** Test the trained GINN on the N_{test} input-output pairs to analyze its generalization capability for unseen data for a given test error tolerance of 10^{-4}.
4. **Prediction:** Sample $N_{\text{sam}}^{\text{pred}}$ inputs from the structured priors on the CVs, and predict the corresponding responses with the trained GINN.

3.3.2. Computational efficiency of the physics- and GINN-based models

For complex numerical simulations, the cost of step 1 outweighs, by orders of magnitude, the combined cost of steps 2–4 [6]. The GSA results reported below require 3×10^3 samples of the BN PDE (physics-based model) to train the GINN that satisfies both the training and test error tolerances.[5] Since the generation of new samples with the GINN carries a negligible expense compared to the generation of training data with the BN PDE, the computational costs of the physics- and GINN-based GSAs are virtually identical when carrying out the former using 3×10^3 samples; this allows us to investigate the performance of both approaches for a fixed computational budget.

3.4. GINN-based MISI rankings for supercapacitor dynamics

Fig. 4 exhibits the first-order MISI values for $Y_{\kappa^{\text{eff}}}$ and Y_{t_+} (left column) and the corresponding ranks of the CVs $\boldsymbol{X}$ (right column), estimated respectively with Algorithm 1 and Algorithm 2. All of these quantities are computed, alternatively, with the physics- and GINN-based models. The MISI values are equipped with the adjusted confidence intervals indicating a pairwise non-overlap significance $\bar{\gamma} = 0.01$ (on average, at level $\beta \approx 0.05$). The 95% percentile confidence intervals for the CV ranks in (27), i.e., with $\delta = 0.05$, are based on $N = 10^3$ replications; samples for the estimator in each replication are either predicted using the GINN or are bootstrap resampled from a corpus of 3×10^3 physics-based simulations. For both Algorithm 1 and Algorithm 2, the physics- and GINN-based estimators are largely consistent, which is to be expected since the GINN surrogate satisfies both a preset training and test error tolerance. The highlighted gaps between clusters of MISI values in Fig. 4a,c indicate the groupings of various CVs $\boldsymbol{X}$ by their relative importance. In Fig. 4b,d dashed lines correspond to these highlighted gaps; although the clarity of the rankings in Fig. 4b,d facilitates the automation of decisions in outer-loop tasks, the ranks themselves do not contain information about the relative importance of each parameter as in Fig. 4a,c.

For both QoIs, the MISI estimators obtained with the physics-based model lead to indeterminate rankings. For $Y_{\kappa^{\text{eff}}}$, the confidence intervals for $\{X_{l_{\text{por}}}, X_\omega\}$ and $\{X_{c_{\text{in}}}, X_{\varphi_\Gamma}\}$ overlap, and therefore the difference between their MISI values (i.e., their ranking) does not differ at an average significance level $\bar{\gamma} = 0.01$. Similarly, for Y_{t_+} the rankings for $\{X_{l_{\text{por}}}, X_r, X_{c_{\text{in}}}\}$ are not resolved. In contrast, the differences between the corresponding estimators derived from 5×10^4 (for $Y_{\kappa^{\text{eff}}}$) or 10^5 (for Y_{t_+}) GINN-based predictions are pairwise significant at the level $\bar{\gamma} = 0.01$. Likewise, we observe that the ranks generated using $N = 10^3$ replications with $M = 5 \times 10^4$ observations for $Y_{\kappa^{\text{eff}}}$ or $M = 10^5$ observations for Y_{t_+} predicted with the GINN are fully resolved (indeed, the 95% percentile confidence intervals are vanishingly small on the plots). Moreover, these ranks are consistent with the resolved rankings of the MISI values for $Y_{\kappa^{\text{eff}}}$ and Y_{t_+} deduced from the GINN-based estimators obtained with Algorithm 1. These findings demonstrate the benefit of a GINN, since resolving the rankings with the physics-based model is considerably more expensive given the high cost of generating additional response samples with the BN PDE.

Fig. 4 also compares the direct distributional method of Algorithm 2 to bootstrapped estimators with percentile confidence intervals computed using the physics-based model. Using $N = 10^3$ bootstrap (i.e., resampling [53]) replications of $M = 3 \times 10^3$ observations (M is constrained by the fixed computational budget) yields clusters of indeterminate ranks: the ranking of $\{X_{\lambda_D}, X_{l_{\text{por}}}, X_\omega\}$ and $\{X_{c_{\text{in}}}, X_{\varphi_\Gamma}\}$ for $Y_{\kappa^{\text{eff}}}$, and the ranking of $\{X_{l_{\text{por}}}, X_r, X_{c_{\text{in}}}\}$ and $\{X_{\varphi_\Gamma}, X_{\lambda_D}\}$ for Y_{t_+}, cannot be resolved at the 0.05 level. In both cases, this is over 70% of the CVs. Again, that ranking of the first-order effects of the CVs

[5] While testing requires the generation of 10^3 additional input-output samples with the BN PDE, it is not strictly required and hence not taken into account when comparing the rankings generated with the physics-based model and the GINN.

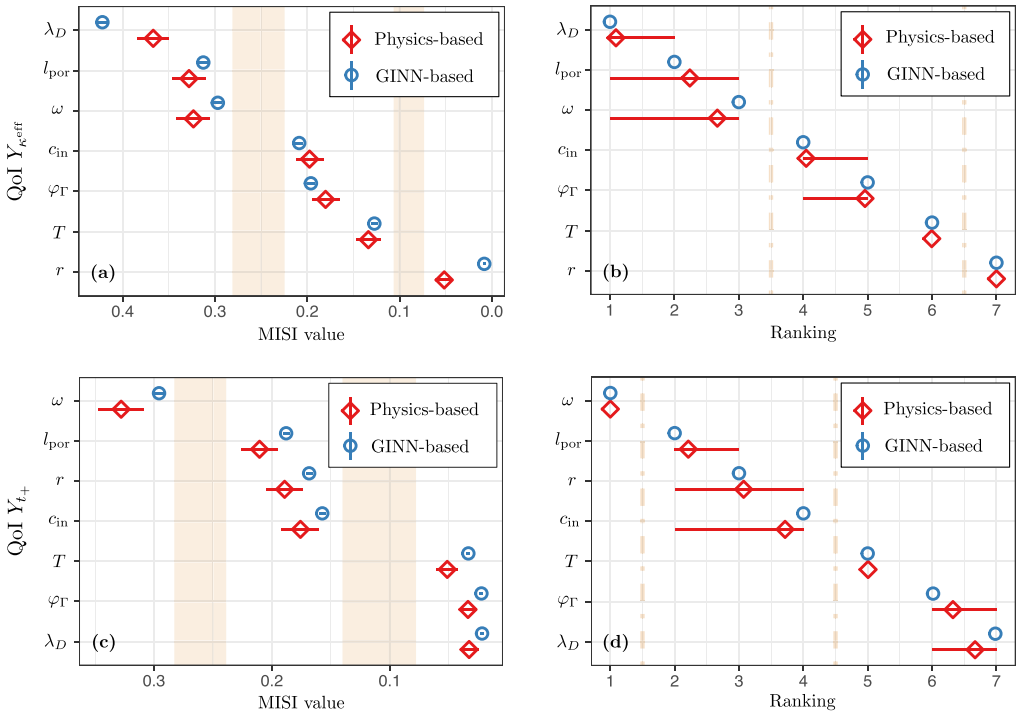

Fig. 4. Plug-in Monte Carlo estimators of the first-order MISIs in (4) indicate the most impactful CVs $\boldsymbol{X}$ for tuning the QoIs $Y_{\kappa^{\text{eff}}}$ (top row) and Y_{t_+} (bottom row). For Algorithm 1 (left column), the width of the confidence intervals (23) is chosen to achieve a non-overlap significance level $\bar{\gamma} = 0.01$ pairwise on average. The highlighted gaps in (a) and (c) indicate clusters of CVs with similar relative importance. For Algorithm 2 (right column), the GINN-based estimators for CV ranking with percentile confidence intervals (27) are consistent with the rankings in (a) and (c). These ranks are resolved, which is not feasible through bootstrapping the 3×10^3 samples from the physics-based model. In both algorithms, the GINN surrogate enables querying sufficiently large amounts of data to distinguish closely-ranked CVs with a high degree of confidence (relative to bootstrapping). In (b) and (d) the dashed lines correspond to the gaps identified in (a) and (c), respectively, and we observe that some of the confidence intervals related to the GINN-based estimates are vanishingly small.

with a high degree of confidence (relative to bootstrapping) within a constrained computational budget critically depends on the availability of a GINN (or, more generally, a surrogate model) to cheaply generate additional response samples.

Since the CVs $\boldsymbol{X}$ are correlated, it is natural to expect higher-order effects due to interactions between the CVs. Fig. 5 displays the second-order MISIs and their ranks estimated using Algorithms 1 and 2, respectively, for the QoI $Y_{\kappa^{\text{eff}}}$. The pairwise comparison of the adjusted confidence intervals with non-overlap significance $\bar{\gamma} = 0.01$ reveals that approximately 80% of the estimators obtained with $M = 3 \times 10^3$ physics-based observations are indistinguishable (Fig. 5a). In contrast, the rankings deduced from the estimators obtained with $M = 5 \times 10^4$ GINN-based observations are fully resolved. The estimated second-order effect ranks based on $N = 30$ replications of $M = 10^5$ GINN-based observations are nearly identical, emphasizing the robustness and consistency of Algorithms 1 and 2. A large proportion of the MISI values is clustered and equally important. The GINN-based GSA resolves $\approx 40\%$ of the ranks compared to $\approx 15\%$ of the ranks distinguishable from $N = 10^3$ bootstrap replications of $M = 3 \times 10^3$ physics-based observations. Here, the number of replications for the GINN-based ranks is chosen such that the total computational time of implementing Algorithm 2 is similar to that of computing all the physics-based bootstrap replications, which is approximately the case when the product $N \cdot M$ is the same for both models.

To summarize the key findings from the numerical experiments presented in Figs. 4 and 5: the GINN surrogate enables the fast generation of useful distinguishable rankings. These rankings are largely consistent with the budget-constrained predictions of the physics-based model. Hence, Algorithms 1 and 2 facilitate the deployment of GINN for the acceleration and future automation of outer-loop decision-support tasks.

4. Design with explainable black-box surrogates

Like all deep neural networks, GINNs are black boxes that lack a clear functional relationship between inputs and outputs. MI-based GSA aids in interpreting and explaining their predictions, thereby enabling the use of black-box surrogates in simulation-based decision-making, including the closure of engineering design loops to facilitate rapid prototyping.

We validate the first-order MISI rankings discussed in Section 3.4, and then use these rankings to explore subregions of the original parameter space that deliver high values of the effective electrolyte conductivity, κ^{eff}. Subsequent effect

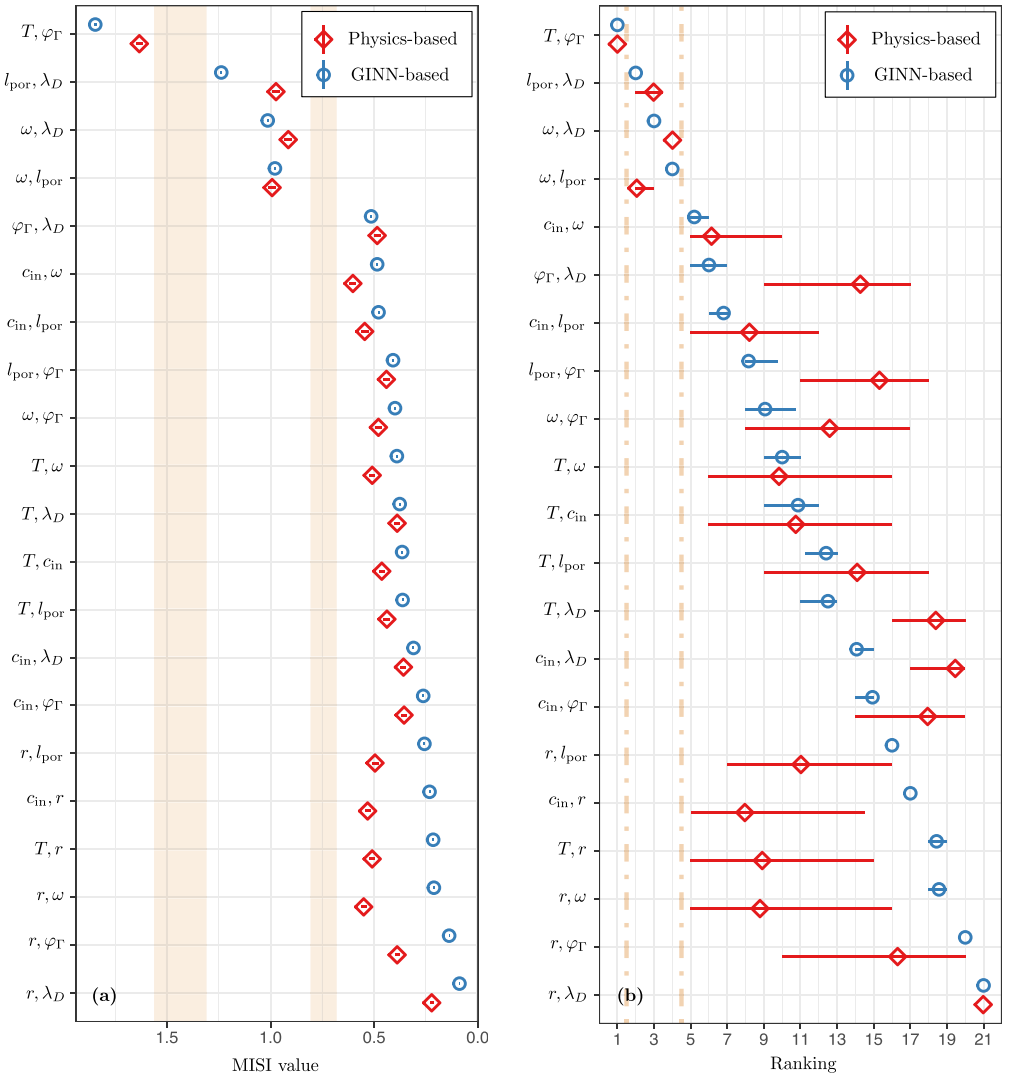

Fig. 5. Plug-in Monte Carlo estimators of the second-order MISIs in (11) indicate the most impactful interactions between any two CVs in $\boldsymbol{X}$ for tuning the QoI $Y_{\kappa^{\text{eff}}}$. The GINN surrogate improves the resolution of the second-order MISI rankings (relative to bootstrapping) computed with (a) Algorithm 1 or (b) Algorithm 2. Although the first-order effect of X_T and X_{φ_Γ} on $Y_{\kappa^{\text{eff}}}$ are not top-ranked (see Fig. 4 (a) and (b)), we observe above that $(X_T, X_{\varphi_\Gamma})$ has the most important second-order effect on $Y_{\kappa^{\text{eff}}}$.

rankings within this parameter subspace suggest follow-up simulations or novel laboratory experiments, resulting in further refinements to the design of nanoporous electrodes for electrical double-layer capacitors.

4.1. Validation of MISI rankings for supercapacitor dynamics

Fig. 6 shows normalized response surfaces, in the form of scatter plots and cubic regression splines based on 10^3 observations, for the QoIs κ^{eff} and t_+ along sensitive and insensitive parameter directions identified by the first-order MISI rankings in Fig. 4. The most sensitive parameter directions are λ_D for κ^{eff} and ω for t_+, and the least sensitive are r for κ^{eff} and λ_D for t_+. The response surfaces for κ^{eff} and t_+ (top and bottom rows in Fig. 6, respectively) demonstrate nonlinear relationships with respect to the most sensitive CV directions (Figs. 6a,c). In contrast, the random scatter for the least sensitive directions (Figs. 6b,d) suggests the lack of a clear relationship between these CVs and the QoIs. The quality and strength of these functional relationships validate the assigned rankings.

The MISI effect ranking and above validation step lend interpretability to the black-box predictions. This enables the use of GINNs in design iterations by predicting new response samples in reduced parameter spaces that optimize certain QoIs. The next section illustrates this procedure.

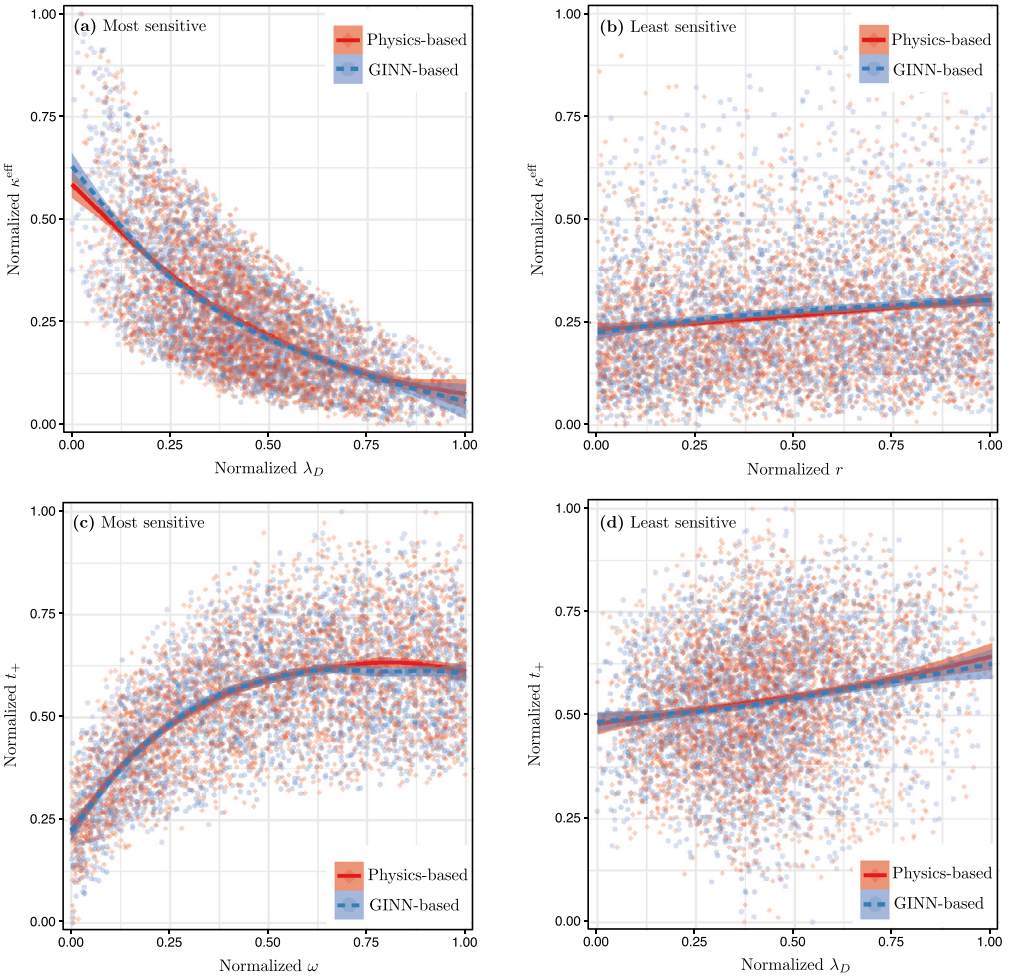

Fig. 6. Response surfaces of the QoIs κ^{eff} (top row) and t_+ (bottom row) with respect to the respective most (left column) and least (right column) sensitive parameters. The plots represent 3×10^3 observations which are fitted with cubic regression splines. These results indicate nonlinear response surfaces for the most sensitive parameter directions, (a) and (c), in contrast to the random dispersion of observations for the least sensitive parameter directions, (b) and (d). This validates the first-order effect rankings in Fig. 4.

4.2. Design of multiscale systems under uncertainty

The first- and second-order MISI rankings suggest that the CVs λ_D, l_{por}, and ω have the largest individual contributions to changes in κ^{eff} (Fig. 4), and the CV pairs (T, φ_Γ), $(l_{\text{por}}, \lambda_D)$, (ω, λ_D) and (ω, l_{por}) (Fig. 5) have the largest pairwise interaction effect. The GINN-generated response surfaces of κ^{eff} for these CV pairs (Fig. 7) identify the parameter subspaces in which a targeted range of the QoI κ^{eff}, e.g., its maximal value, is likely to be achieved. Computation of new MISI rankings in these restricted subregions determines which CVs to retain in the next design cycle.

Fig. 7a,b shows a clear gradient in the response surfaces, with κ^{eff} being largest when λ_D is small and either ω or l_{por} is large; this observation follows directly from (36a) and (37a). For the former case, we zoom in on the region $T \in [208, 360]$ K and $c_{\text{in}} \in [0.9, 1.08]$ mol/l, such that $\lambda_D \in [0.0771, 0.1109]$ nm and $\omega \in [0.7, 0.8375]$. For the latter case, we consider the region $T \in [208, 350]$ K and $c_{\text{in}} \in [0.95, 1.08]$ mol/l (such that $\lambda_D \in [0.0771, 0.1065]$ nm) and $r \in [1.5, 1.75]$ nm and $\omega \in [0.79, 0.8375]$ (such that $l_{\text{por}} \in [1.4030, 2.0965]$ nm). Fig. 8 visualizes the new first-order MISI rankings in the reduced parameter space suggested by the most relevant parameter directions in Fig. 7a,b that were informed by the first-order (Fig. 4) and second-order (Fig. 5) effect rankings. While λ_D still has the biggest impact, φ_Γ and T are now the second- and third-most important CVs in both cases, while r remains the least important CV. Repeating this process informs subsequent decision tasks, yielding a procedure to iteratively refine the materials design. A similar reasoning can be followed based on the response surface of κ^{eff} for variations in ω and l_{por} in Fig. 7c. The narrower shape of this surface reflects the correlation between these CVs, in accordance with (36c).

Comparison of the first-order rankings in Fig. 4 with the second-order rankings in Fig. 5 reveals that even though the individual contributions of φ_Γ and T to changes in κ^{eff} are smaller than those of λ_D, l_{por} and ω, the effect of their interactions dominates that of the pairwise interactions between the latter. This can be explained by the different degrees of

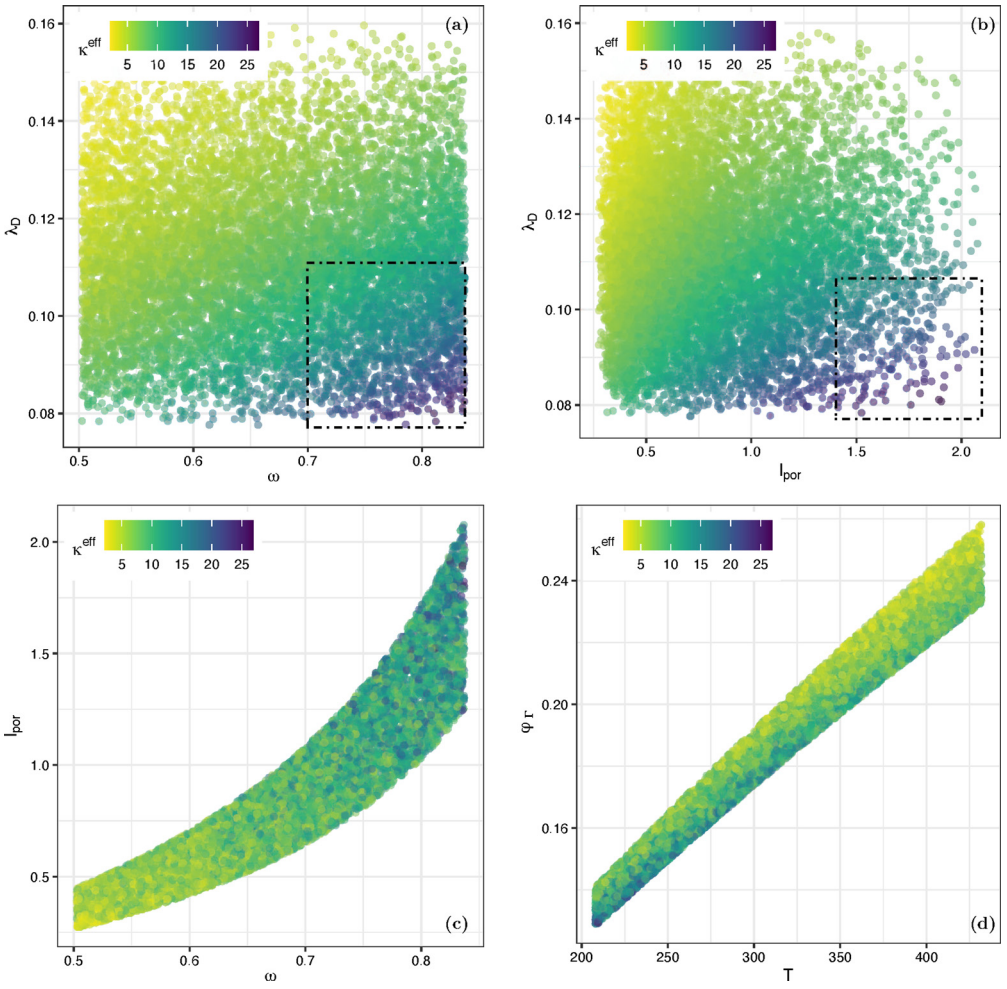

Fig. 7. GINN-predicted response surfaces of the effective electrolyte conductivity κ^{eff} based on 3×10^3 observations identify reduced parameter ranges corresponding to a targeted response which can be explored to close design loops. The response surfaces are plotted over two-dimensional subspaces of the full input space; in (a), (b), and (c) the CVs correspond to the top-ranked first-order MISIs in Fig. 4 and in (d) they correspond to the highest second-order MISI in Fig. 5. In (a) and (b) the subspaces enclosed by dashed boxes correspond to high values of κ^{eff} and are the focus of the further investigations in Fig. 8.

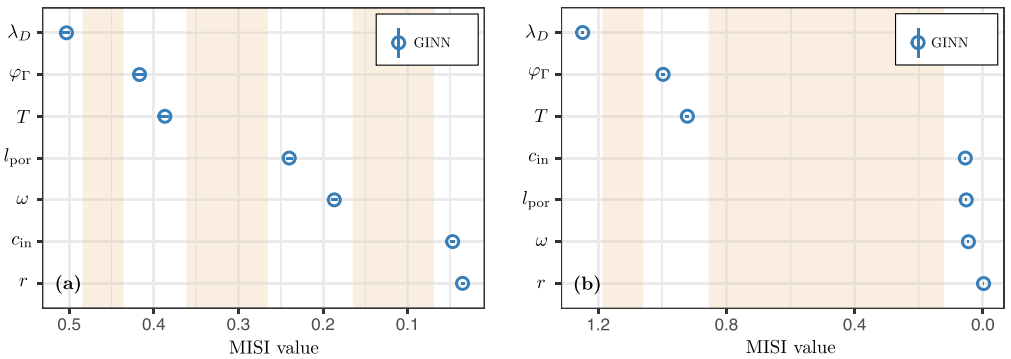

Fig. 8. Plug-in Monte Carlo estimators of the first-order MISI rankings in the large κ^{eff} regime. The latter corresponds to the reduced ranges of (a) λ_D, ω and (b) λ_D, l_{por} indicated by the dashed boxes in Fig. 7a and b, respectively. These new rankings are based on 5×10^4 samples generated by the GINN surrogate. They differ from those in Fig. 4 and inform subsequent decision tasks in the multiscale design process.

dependence among these CVs: λ_D is not directly related to l_{por} or ω, while the pairs (T, φ_Γ) and (ω, l_{por}) are combinations of CVs that depend on each other through physical relations (see (36b) and (36c), respectively). This is reflected by the various shapes of the response surfaces in Fig. 7. This result enables one to reduce parameter ranges for φ_Γ and T to close

design loops and demonstrates the importance of the higher-order effects (i.e., due to interactions between CVs) introduced in Section 2.2 for design loop closure in complex systems.

Since the new parameter ranges considered for predicting the effect rankings in Fig. 8 lie inside the original parameter space, we used the existing trained GINN to predict the new response samples. Alternatively, the MISI rankings might suggest the exploration of new parameter ranges outside the original space, thereby guiding the generation of (a limited amount of) new physics-based data corresponding to those ranges and retraining the GINN with this new training dataset. For deep neural networks with many hidden layers, transfer learning [75] can be employed to reuse most of the existing network and thereby reduce the cost of training the new surrogate.

5. Conclusions and outlook

We developed a UQ emulator for moment-independent global sensitivity analysis (GSA) based on differential mutual information (MI). Mutual Information Sensitivity Indices (MISIs) provide a model-agnostic mechanism for ranking the impact of correlated tunable control variables (CVs) on quantities of interest (QoIs). The high computational cost of querying physics-based models, typically a barrier to the use of such data-driven methods, is ameliorated by leveraging deep learning-based surrogate models that enable fast generation of response data. Although these black boxes do not generate engineering insights, our MI-based GSA allows one to interrogate surrogates and explain surrogate predictions in the context of the physics-based model to close engineering design loops.

We utilized our UQ emulators in conjunction with a recently developed Graph-Informed Neural Network (GINN) [6], capable of handling correlated CVs, and tested our approach on two models of interest in energy storage including a multiscale model of an electrical double-layer capacitor. We presented two algorithms for estimating and ranking MISIs and, for the applications of interest, calculated first-order MISIs to capture individual CV effects and second-order MISIs to capture the effects of pairwise interactions between CVs. We validated the first-order MISI rankings against both physics- and GINN-based QoI response surfaces. Finally, we closed the engineering design loop by considering the most sensitive input directions and investigating effect rankings in a reduced parameter space corresponding to large effective conductivity values.

Our analysis leads to the following major conclusions.

1. Viewed as uncertainty quantification for surrogate models, our MI-based GSA works seamlessly with the GINN—a promising result that encourages its application to interrogate other black-box surrogate models.
2. UQ emulators deliver well-resolved first-order effect rankings, facilitated by either comparison-adjusted (Algorithm 1) or percentile (Algorithm 2) confidence intervals. While not as rigorous as UQ for physics-based models, they can nevertheless be used to characterize interventions that enable outer loop control and optimization.
3. For the applications of interest, the resolved first-order effect rankings produced with the GINN-driven UQ emulators are consistent between Algorithms 1 and 2 (see Figs. 2 and 4).
4. The most/least sensitive CVs identified through the first-order MISI effect rankings show a nonlinear/nearly flat response curve for the QoIs, supporting the validity of the MI-based approach. The relative magnitudes of the largest first-order MISIs for both QoIs produced by Algorithm 1 are also in line with the differences between the normalized response curves; this holds true for the smallest first-order MISIs as well.
5. The impact of mutual interactions between correlated CVs on the QoIs needs to be taken into account via higher-order MISIs. Pairwise interactions between CVs with small individual contributions to the QoIs can dominate those between CVs with larger additive effects.
6. Within a reduced parameter space leading to optimal QoI values, the relative importance of the various CVs is different from that in the original parameter space (which could be nonlinear, see Fig. 7c,d). This new ranking informs subsequent design cycles, spawning an iterative procedure that enables rapid prototyping and reduces time to market.

Motivated by its successful application to GINNs, we aim to pair our MI-based GSA with other deep neural network surrogates including physics-informed and physics-constrained neural networks. In particular, our MISI rankings could help simplify the custom loss functions of those deep neural networks by filtering out less important parameters.

CRediT authorship contribution statement

Søren Taverniers: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Eric J. Hall:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Markos A. Katsoulakis:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. **Daniel M. Tartakovsky:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

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.

Acknowledgements

The research of S.T. and D.T. was partially supported by the Air Force Office of Scientific Research (AFOSR) under grant FA9550-18-1-0474 and by a gift from Total, both awarded to D.T. A portion of this research was undertaken when E.H. was a postdoctoral research scientist in the Chair of Mathematics for Uncertainty Quantification at RWTH Aachen University, Germany and was partially supported by the Alexander von Humboldt Foundation. The research of M.K. was partially supported by the Air Force Office of Scientific Research (AFOSR) under grant FA-9550-18-1-0214 and by the National Science Foundation (NSF) under grants DMS-2008970 and CISE-1934846.

References

[1] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Rev. 60 (3) (2018) 550–591, https://doi.org/10.1137/16M1082469.
[2] G. Balokas, S. Czichon, R. Rolfes, Neural network assisted multiscale analysis for the elastic properties prediction of 3D braided composites under uncertainty, Compos. Struct. 183 (2018) 550–562, https://doi.org/10.1016/j.compstruct.2017.06.037.
[3] R.K. Tripathy, I. Bilionis Deep UQ, Learning deep neural network surrogate models for high dimensional uncertainty quantification, J. Comput. Phys. 375 (2018) 565–588, https://doi.org/10.1016/j.jcp.2018.08.036.
[4] 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, https://doi.org/10.1016/j.jcp.2019.05.024.
[5] M. Raissi, P. Perdikaris, G. 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, https://doi.org/10.1016/j.jcp.2018.10.045.
[6] E.J. Hall, S. Taverniers, M.A. Katsoulakis, D.M. Tartakovsky Ginns, Graph-informed neural networks for multiscale physics, J. Comput. Phys. 433 (2021) 110192, https://doi.org/10.1016/j.jcp.2021.110192.
[7] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Boston, MA, 2016, https://www.deeplearningbook.org/.
[8] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis: The Primer, John Wiley & Sons Ltd., 2008.
[9] I.M. Sobol, Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp. 1 (4) (1993) 407–414.
[10] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, Reliab. Eng. Syst. Saf. 52 (1) (1996) 1–17, https://doi.org/10.1016/0951-8320(96)00002-6.
[11] T.A. Mara, S. Tarantola, P. Annoni, Non-parametric methods for global sensitivity analysis of model output with dependent inputs, Environ. Model. Softw. 72 (2015) 173–183, https://doi.org/10.1016/j.envsoft.2015.07.010.
[12] B. Iooss, C. Prieur, Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications, Int. J. Uncertain. Quantificat. 9 (5) (2019) 493–514.
[13] E. Borgonovo, Measuring uncertainty importance: investigation and comparison of alternative approaches, Risk Anal. 26 (5) (2006) 1349–1361, https://doi.org/10.1111/j.1539-6924.2006.00806.x.
[14] E. Borgonovo, A new uncertainty importance measure, Reliab. Eng. Syst. Saf. 92 (6) (2007) 771–784, https://doi.org/10.1016/j.ress.2006.04.015.
[15] V. Ciriello, I. Lauriola, D.M. Tartakovsky, Distribution-based global sensitivity analysis in hydrology, Water Resour. Res. 55 (11) (2019) 8708–8720, https://doi.org/10.1029/2019WR025844.
[16] W. Castaings, E. Borgonovo, M. Morris, S. Tarantola, Sampling strategies in density-based sensitivity analysis, Environ. Model. Softw. 38 (2012) 13–26, https://doi.org/10.1016/j.envsoft.2012.04.017.
[17] C. Vetter, A.A. Taflanidis, Global sensitivity analysis for stochastic ground motion modeling in seismic-risk assessment, Soil Dyn. Earthq. Eng. 38 (2012) 128–143, https://doi.org/10.1016/j.soildyn.2012.01.004.
[18] A.J. Majda, B. Gershgorin, Quantifying uncertainty in climate change science through empirical information theory, Proc. Natl. Acad. Sci. USA 107 (34) (2010) 14958–14963, https://doi.org/10.1073/pnas.1007009107.
[19] M. Komorowski, M.J. Costa, D.A. Rand, M.P.H. Stumpf, Sensitivity, robustness, and identifiability in stochastic chemical kinetics models, Proc. Natl. Acad. Sci. USA 108 (21) (2011) 8645–8650, https://doi.org/10.1073/pnas.1015814108.
[20] A.J. Majda, B. Gershgorin, Improving model fidelity and sensitivity for complex systems through empirical information theory, Proc. Natl. Acad. Sci. USA 108 (25) (2011) 10044–10049, https://doi.org/10.1073/pnas.1105174108.
[21] Y. Pantazis, M.A. Katsoulakis, D.G. Vlachos, Parametric sensitivity analysis for biochemical reaction networks based on pathwise information theory, BMC Bioinform. 14 (1) (2013) 1, https://doi.org/10.1186/1471-2105-14-311.
[22] Y. Pantazis, M.A. Katsoulakis, A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics, J. Chem. Phys. 138 (5) (2013) 054115, https://doi.org/10.1063/1.4789612.
[23] G.C. Critchfield, K.E. Willard, D.P. Connelly, Probabilistic sensitivity analysis methods for general decision models, Comput. Biomed. Res. 19 (3) (1986) 254–265, https://doi.org/10.1016/0010-4809(86)90020-0.
[24] H. Liu, W. Chen, A. Sudjianto, Relative entropy based method for probabilistic sensitivity analysis in engineering design, J. Mech. Des. 128 (2) (2006) 326–336, https://doi.org/10.1115/1.2159025.
[25] N. Lüdtke, S. Panzeri, M. Brown, D.S. Broomhead, J. Knowles, M.A. Montemurro, D.B. Kell, Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks, J. R. Soc. Interface 5 (19) (2008) 223–235, https://doi.org/10.1098/rsif.2007.1079.
[26] Q. Liu, T. Homma, A new computational method of a moment-independent uncertainty importance measure, Reliab. Eng. Syst. Saf. 94 (7) (2009) 1205–1211, https://doi.org/10.1016/j.ress.2008.10.005.
[27] S. Rahman, The f-sensitivity index, SIAM/ASA J. Uncertain. Quantificat. 4 (1) (2016) 130–162, https://doi.org/10.1137/140997774.
[28] K. Um, E.J. Hall, M.A. Katsoulakis, D.M. Tartakovsky, Causality and Bayesian network PDEs for multiscale representations of porous media, J. Comput. Phys. 394 (2019) 658–678, https://doi.org/10.1016/j.jcp.2019.06.007.
[29] E.J. Hall, M.A. Katsoulakis, Robust information divergences for model-form uncertainty arising from sparse data in random PDE, SIAM/ASA J. Uncertain. Quantificat. 6 (4) (2018) 1364–1394, https://doi.org/10.1137/17M1143344.
[30] J. Feng, J.L. Lansford, M.A. Katsoulakis, D.G. Vlachos, Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences, Sci. Adv. 6 (42) (2020), https://doi.org/10.1126/sciadv.abc3204, https://advances.sciencemag.org/content/6/42/eabc3204.
[31] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd edition, Wiley-Interscience, Hoboken, NJ, 2006.
[32] E.S. Soofi, Capturing the intangible concept of information, J. Am. Stat. Assoc. 89 (428) (1994) 1243–1254, https://doi.org/10.1080/01621459.1994.10476865.
[33] D. Zhang, L. Lu, L. Guo, G.E. Karniadakis, Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems, J. Comput. Phys. 397 (2019) 108850, https://doi.org/10.1016/j.jcp.2019.07.048.

[34] Y. Yang, P. Perdikaris, Adversarial uncertainty quantification in physics-informed neural networks, J. Comput. Phys. 394 (2019) 136–152, https://doi.org/10.1016/j.jcp.2019.05.027.
[35] X. Meng, G.E. Karniadakis, A composite neural network that learns from multi-fidelity data: application to function approximation and inverse PDE problems, J. Comput. Phys. 401 (2020) 109020, https://doi.org/10.1016/j.jcp.2019.109020.
[36] J. Sirignano, K. Spiliopoulos, DGM: a deep learning algorithm for solving partial differential equations, J. Comput. Phys. 375 (2018) 1339–1364, https://doi.org/10.1016/j.jcp.2018.08.029.
[37] J. Berg, K. Nyström, A unified deep artificial neural network approach to partial differential equations in complex geometries, Neurocomputing 317 (2018) 28–41, https://doi.org/10.1016/j.neucom.2018.06.056.
[38] L. Sun, H. Gao, S. Pan, J.-X. Wang, Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data, Comput. Methods Appl. Mech. Eng. 361 (2020) 112732, https://doi.org/10.1016/j.cma.2019.112732.
[39] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: visualising image classification models and saliency maps, in: ICLR, 2014, pp. 1–8.
[40] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, B. Kim, Sanity checks for saliency maps, published in the Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada. Arxiv eprint version contains: "Updating Guided Backprop experiments due to bug. The results and conclusions remain the same" (November 2020), arXiv:1810.03292.
[41] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences, in: D. Precup, Y.W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning, vol. 70, International Conference on Machine Learning, PMLR, 2017, pp. 3145–3153.
[42] M. Sundararajan, A. Taly, Q. Yan, Gradients of counterfactuals, arXiv:1611.02639, 2016.
[43] Q. Zhang, Y.N. Wu, S. Zhu, Interpretable convolutional neural networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8827–8836.
[44] E. Štrumbelj, I. Kononenko, A general method for visualizing and explaining black-box regression models, in: A. Dobnikar, U. Lotrič, B. Šter (Eds.), Adaptive and Natural Computing Algorithms, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 21–30.
[45] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[46] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation, J. Comput. Graph. Stat. 24 (1) (2015) 44–65, https://doi.org/10.1080/10618600.2014.907095.
[47] C.M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, 2006.
[48] D. Koller, N. Friedman, Probabilistic Graphical Models, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2009.
[49] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference, 2nd edition, Springer-Verlag, New York, 2002.
[50] J.B. Kinney, G.S. Atwal, Equitability, mutual information, and the maximal information coefficient, Proc. Natl. Acad. Sci. 111 (9) (2014) 3354–3359, https://doi.org/10.1073/pnas.1309933111.
[51] A. Krishnamurthy, K. Kandasamy, B. Poczos, L. Wasserman, Nonparametric estimation of Renyi divergence and friends, in: E.P. Xing, T. Jebara (Eds.), Proceedings of the 31st International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 32, PMLR, Bejing, China, 2014, pp. 919–927.
[52] K. Kandasamy, A. Krishnamurthy, B. Poczos, L. Wasserman, et al., Nonparametric von Mises estimators for entropies, divergences and mutual informations, in: Advances in Neural Information Processing Systems, 2015, pp. 397–405.
[53] L.A. Wasserman, All of Nonparametric Statistics: With 52 Illustrations, Springer, 2006.
[54] Z.I. Botev, J.F. Grotowski, D.P. Kroese, Kernel density estimation via diffusion, Ann. Stat. 38 (5) (2010) 2916–2957, https://doi.org/10.1214/10-AOS799.
[55] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Phys. Rev. E 69 (6) (2004) 066138, https://doi.org/10.1103/PhysRevE.69.066138.
[56] M.I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, D. Hjelm, Mutual information neural estimation, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, vol. 80 of Proceedings of Machine Learning Research, PMLR, Stockholmsmässan, Stockholm Sweden, 2018, pp. 531–540.
[57] W. McGill, Multivariate information transmission, Psychometrika 19 (1954) 97–116, https://doi.org/10.1007/BF02289159.
[58] A.J. Bell, The co-information lattice, in: Proc. 4th Int. Symp. Independent Component Analysis and Blind Source Separation, 2003, pp. 921–926.
[59] H. Goldstein, M.J.R. Healy, The graphical presentation of a collection of means, J. R. Stat. Soc. A 158 (1) (1995) 175–177, https://doi.org/10.2307/2983411.
[60] B. Efron, Nonparametric standard errors and confidence intervals, Can. J. Stat. 9 (2) (1981) 139–158, https://doi.org/10.2307/3314608.
[61] M.E. Davis, R.J. Davis, Fundamentals of Chemical Reaction Engineering, McGraw-Hill Chemical Engineering Series, McGraw-Hill Higher Education, New York, NY, 2003, https://resolver.caltech.edu/CaltechBOOK:2003.001.
[62] S. Gu, B. Xu, Y. Yan, Electrochemical energy engineering: a new frontier of chemical engineering innovation, Annu. Rev. Chem. Biomol. Eng. 5 (1) (2014) 429–454, https://doi.org/10.1146/annurev-chembioeng-060713-040114, pMID: 24702299.
[63] S. Liu, T. Ishimoto, D.S. Monder, M. Koyama, First-principles study of oxygen transfer and hydrogen oxidation processes at the ni-ysz-gas triple phase boundaries in a solid oxide fuel cell anode, J. Phys. Chem. C 119 (49) (2015) 27603–27608, https://doi.org/10.1021/acs.jpcc.5b10878.
[64] S. Lee, S. Mukerjee, E. Ticianelli, J. McBreen, Electrocatalysis of co tolerance in hydrogen oxidation reaction in pem fuel cells, Electrochim. Acta 44 (19) (1999) 3283–3293, https://doi.org/10.1016/S0013-4686(99)00052-3, http://www.sciencedirect.com/science/article/pii/S0013468699000523.
[65] T. Nagasawa, K. Hanamura, Theoretical analysis of hydrogen oxidation reaction in solid oxide fuel cell anode based on species territory adsorption model, J. Power Sources 290 (2015) 168–182, https://doi.org/10.1016/j.jpowsour.2015.04.160, http://www.sciencedirect.com/science/article/pii/S0378775315008289.
[66] J. Feng, J. Lansford, A. Mironenko, D.B. Pourkargar, D.G. Vlachos, M.A. Katsoulakis, Non-parametric correlative uncertainty quantification and sensitivity analysis: application to a Langmuir bimolecular adsorption model, AIP Adv. 8 (3) (2018) 035021, https://doi.org/10.1063/1.5021351, https://advances.sciencemag.org/content/6/42/eabc3204.
[67] N. Tishby, N. Zaslavsky, Deep learning and the information bottleneck principle, in: 2015 IEEE Information Theory Workshop (ITW), 2015, pp. 1–5.
[68] A. Soffer, M. Folman, The electrical double layer of high surface porous carbon electrode, J. Electroanal. Chem. Interfacial Electrochem. 38 (1) (1972) 25, https://doi.org/10.1016/S0022-0728(72)80087-1.
[69] R. Narayanan, H. Vijwani, S.M. Mukhopadhyay, P.R. Bandaru, Electrochemical charge storage in hierarchical carbon manifolds, Carbon 99 (2016) 267, https://doi.org/10.1016/j.carbon.2015.11.078.
[70] K. Nomura, H. Nishihara, N. Kobayashi, T. Asadab, T. Kyotani, 4.4 V supercapacitors based on super-stable mesoporous carbon sheet made of edge-free graphene walls, Energy Environ. Sci. 12 (2019) 1542, https://doi.org/10.1039/C8EE03184C.
[71] Z. Li, S. Gadipelli, H. Li, C.A. Howard, D.J. Brett, P.R. Shearing, Z. Guo, I.P. Parkin, F. Li, Tuning the interlayer spacing of graphene laminate films for efficient pore utilization towards compact capacitive energy storage, Nat. Energy 5 (2) (2020) 160–168, https://doi.org/10.1038/s41560-020-0560-6.
[72] Z. Wang, et al., Extremely low self-discharge solid-state supercapacitors via the confinement effect of ion transfer, J. Mater. Chem. A 7 (2019) 8633, https://doi.org/10.1039/C9TA01028A.
[73] F. Béguin, E. Frąckowiak, G.Q. Max Lu, Supercapacitors: Materials, Systems, and Applications, 1st edition, Wiley-VCH, 2013.
[74] X. Zhang, D.M. Tartakovsky, Effective ion diffusion in charged nanoporous materials, J. Electrochem. Soc. 164 (4) (2017) E53–E61, https://doi.org/10.1149/2.0491704jes.
[75] L.Y. Pratt, Discriminability-based transfer between neural networks, in: NIPS Conference, Adv. Neural Inf. Process. Syst. 5 (1993) 204–211.