

Causality and Bayesian Network PDEs for multiscale representations of porous media



Kimoon Um ^{a,1}, Eric J. Hall ^{b,*,1,2}, Markos A. Katsoulakis ^b,
Daniel M. Tartakovsky ^{a,*}

^a Department of Energy Resources Engineering, Stanford University, 367 Panama Street, Stanford, CA 94305, USA

^b Department of Mathematics and Statistics, University of Massachusetts Amherst, 710 N. Pleasant Street, Amherst, MA 01003, USA

ARTICLE INFO

Article history:

Received 2 October 2018

Received in revised form 28 April 2019

Accepted 5 June 2019

Available online 7 June 2019

Keywords:

Bayesian Networks

Causality

Porous media

Energy storage

Global sensitivity analysis

Mutual Information

ABSTRACT

Microscopic (pore-scale) properties of porous media affect and often determine their macroscopic (continuum- or Darcy-scale) counterparts. Understanding the relationship between processes on these two scales is essential to both the derivation of macroscopic models of, e.g., transport phenomena in natural porous media, and the design of novel materials, e.g., for energy storage. Microscopic properties exhibit complex statistical correlations and geometric constraints that present challenges for the estimation of macroscopic quantities of interest (QoIs), e.g., in the context of global sensitivity analysis (GSA) of macroscopic QoIs with respect to microscopic material properties. We present a systematic way of building correlations into stochastic multiscale models through Bayesian Networks. The proposed framework allows us to construct the joint probability density function (PDF) of model parameters through causal relationships that are informed by domain knowledge and emulate engineering processes, e.g., the design of hierarchical nanoporous materials. These PDFs also serve as input for the forward propagation of parametric uncertainty thereby yielding a Bayesian Network PDE. To assess the impact of causal relationships and microscale correlations on macroscopic material properties, we propose a moment-independent GSA and corresponding effect rankings for Bayesian Network PDEs, based on the differential Mutual Information, that leverage the structure of Bayesian Networks and account for both correlated inputs and complex non-Gaussian (skewed, multimodal) QoIs. Our findings from numerical experiments, which feature a non-intrusive uncertainty quantification workflow, indicate two practical outcomes. First, the inclusion of correlations through structured priors based on causal relationships informed by domain knowledge impacts predictions of QoIs and has important implications for engineering design. Second, structured priors with non-trivial correlations yield different effect rankings than independent priors; these rankings are more consistent with the anticipated physics.

© 2019 Elsevier Inc. All rights reserved.

* Corresponding authors.

E-mail addresses: kimoon.um@gmail.com (K. Um), hall@uq.rwth-aachen.de (E.J. Hall), markos@math.umass.edu (M.A. Katsoulakis), tartakovsky@stanford.edu (D.M. Tartakovsky).

¹ Both authors contributed equally to this work.

² Present address: Chair of Mathematics for Uncertainty Quantification, Department of Mathematics, RWTH Aachen University, 52056 Aachen, Germany.

1. Introduction

Understanding statistical and causal relations between properties/model parameters at various scales is essential for science-based predictions in general and for forecasts of transport phenomena in porous media in particular. For example, the design of materials for energy storage devices aims to optimize macroscopic material properties (quantities of interest or QoIs), such as effective diffusion coefficient and capacitance, through engineered pore structures [1,2]. Quantification of both uncertainty in predictions of these macroscopic QoIs and their sensitivity to variability and uncertainty in microscopic features are crucial for informing decision tasks such as optimal experimental design and reliability engineering.

The simulation-assisted approach to the optimal design of porous metamaterials takes advantage of the availability of microscopic (pore-scale) and macroscopic (continuum or Darcy-scale) models, and of a link between the two provided by various upscaling techniques [1–3]. Such a link also facilitates analyses of both uncertainty propagation from the microscopic scale to the macroscopic scale and sensitivity of macroscopic material properties to the microscopic ones [4]. Predictions resulting from this multiscale approach can be made more robust by incorporating information about correlations and causal relationships between scales and within a single scale. Causal relationships are natural and stem from domain knowledge such as physical, chemical, and/or engineering design constraints, expert knowledge, and available data. Our primary objective is to bring a Bayesian Network perspective to stochastic multiscale modeling thereby providing a systematic framework for incorporating structured priors based on causal relationships informed by domain knowledge.

Bayesian Networks are a special class of hierarchical probabilistic graphical models (PGMs) with a directed acyclic graph structure that represents conditional relationships among random variables [5–7]. Bayesian Networks and, more generally, PGMs provide a rich framework for encoding distributions over large, complex domains of interacting random variables. In our application, they provide a coherent framework for representing causal relationships both between problem scales and among the space of parameters representing pore-scale features. Incorporating Bayesian Networks into random PDEs is a novel approach to the modeling of multiscale porous media; it breaks down the stochastic modeling and statistical inference tasks into smaller, controllable parts enabling us to

- (i) build systematically informed parameter priors that include physical constraints and/or correlations;
- (ii) mirror engineering processes related to the design of hierarchical nanoporous media networks;
- (iii) construct Bayesian Network (random) Darcy-scale PDE models informed by domain knowledge including (possibly uncertain) pore-scale data, parameters, and constraints; and
- (iv) carry out global sensitivity analysis (GSA), effect rankings, and uncertainty quantification (UQ).

Related to this last point, this framework for incorporating structured priors through causal relationships informed by domain knowledge demands GSA tools that differ from the standard variance-based GSA methods. Conventional variance-based methods assume unstructured (i.e., mutually independent) priors and are neither easy to interpret nor cheap to compute for correlated and/or dependent inputs [8–10]. In our application, causal relationships exist not just between parameters but also between scales. The latter is important since our predictive PDFs are not necessarily Gaussian and/or do not have a known analytical form. This suggests moving away from a fixed number of moments to a moment-independent quantity such as Mutual Information. Leveraging the structure of our Bayesian Networks, we employ a moment-independent GSA relying on Mutual Information [11,12] and empirical distributions acquired through simulations. We demonstrate that differential Mutual Information provides a measure of input effects for Bayesian Network PDEs that is suited to tackling the twin challenges of structured or correlated inputs and non-Gaussian (highly skewed, multimodal) QoIs.

Design of nanoporous metamaterials for energy storage provides an ideal setting to illustrate the Bayesian Network PDE approach. Macroscopic material properties are dependent on a set of microscopic parameters characterizing the pore geometry, e.g., pore radius or pore connectivity. These microscopic parameters are typically correlated due to the presence of geometric and topological constraints and uncertain due to natural and/or manufacturing variability. This setting gives rise to a number of theoretical and practical questions: How does uncertainty in microscopic properties (quantified, e.g., in terms of a pore-size distribution) propagate to the macroscopic scale (expressed in terms of the PDF of, e.g., the effective diffusion coefficient)? How sensitive are a material's macroscopic properties to its microscopic counterparts? Etc. Our non-intrusive computational workflow, which makes exhaustive sampling for prediction and uncertainty quantification feasible, has three ingredients: (i) Rosenblatt transforms to decorrelate inputs for non-intrusive scientific computing, (ii) generalized polynomial chaos expansions obtained using the DAKOTA software [13], and (iii) kernel density estimation techniques.

In Section 2, we formulate a macroscopic (Darcy-scale) model of reactive transport in hierarchical nanoporous media. The model parameters are expressed in terms of microscopic (pore-scale) material properties by means of homogenization [4], which facilitates multiscale UQ and GSA. The work [4] considers a similar application problem and utilizes homogenization for forward propagation of uncertainty thereby enabling variance-based GSA. However, in this approach the space of input distributions for pore features is assumed to be independent and consequently simulations that maintain these assumptions are restricted to distributions supported over narrow, non-physical ranges. A major contribution of the present work is to use Bayesian Networks for stochastic modeling and to incorporate domain knowledge using causal relationships. Furthermore, we propose Mutual Information-based GSA as a suitable sensitivity and effect ranking tool for Bayesian Network PDEs that is enabled by a non-intrusive uncertainty quantification workflow. Section 3 contains a description of our Bayesian Network PDE approach for modeling and uncertainty quantification. Section 4 contains an implementation for a

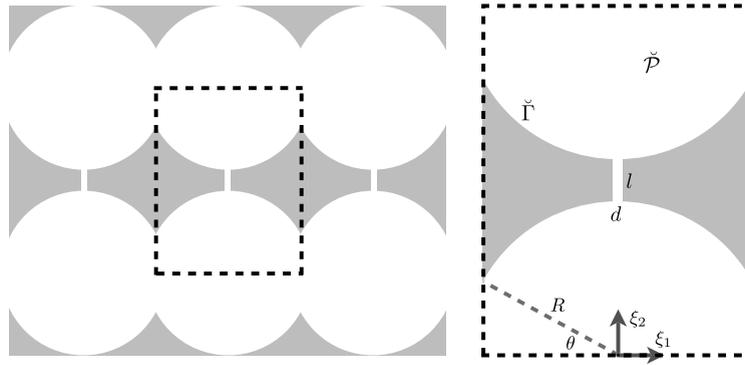


Fig. 1. A hierarchical nanoporous material [4] exhibiting horizontally oriented nanotunnels through mesopores connected by a series of vertically oriented nanotubes. The porous media volume \mathcal{V} (left) consists of a periodic arrangement of unit cells $\check{\mathcal{V}}$ (right) with pore space $\check{\mathcal{P}}$ and fluid-solid interface $\check{\Gamma}$. The parameters $\{R, \theta, d, l\}$ describing the nanopore features are constrained by the geometry of \mathcal{V} .

model problem highlighting our workflow that uses inverse Rosenblatt transforms to non-intrusively interface with existing software, such as DAKOTA. This section also collects the results of our numerical experiments, which demonstrate the impact of causality and non-trivial correlations. In Section 5, we propose differential Mutual Information GSA indices and effect rankings for Bayesian Network PDEs. Our examples illustrate how the inclusion of causal relationships that encode domain knowledge about structural constraints provides rankings more consistent with the physics anticipated for a simple hierarchical nanoporous material. In Section 6, we present an alternative Bayesian Network for the model problem to highlight the method's flexibility and ability to mirror distinct engineering and design processes. Major conclusions drawn from our study are summarized in Section 7.

2. Models of flow and transport in nanoporous materials

A volume $\mathcal{V} = \mathcal{P} \cup \mathcal{S}$ of a hierarchical porous material is comprised of a fluid-filled pore space \mathcal{P} and solid matrix \mathcal{S} , with a (multi-connected) fluid-solid interface denoted by $\Gamma = \mathcal{P} \cap \mathcal{S}$. To mimic a manufacturing process and to make subsequent use of the homogenization theory, we assume that the volume \mathcal{V} consists of a periodic arrangement of unit cells $\check{\mathcal{V}} = \check{\mathcal{P}} \cap \check{\mathcal{S}}$ with pore space $\check{\mathcal{P}} \subset \mathcal{P}$, solid matrix $\check{\mathcal{S}} \subset \mathcal{S}$, and fluid-solid interface $\check{\Gamma} = \check{\mathcal{P}} \cap \check{\mathcal{S}}$. For example, the hierarchical nanoporous material in Fig. 1 consists of mesopores that are connected longitudinally (horizontally) through nanotunnels and transversely (vertically) by a series of nanotubes. These features are described by a set of parameters $\{R, \theta, d, l\}$, where R is the mesopore radius; θ is the angle of overlap between adjacent mesopores in a nanotunnel; and d and l are the diameter and length of the nanotubes, which serve as nano-bridges between adjacent mesopores/nanotunnels.

The design of novel materials calls for a systematic analysis of the sensitivity of desired macroscopic (Darcy-scale) properties to imperfections (natural variability) in microscopic (pore-scale) parameters and/or their distributions. Following [4], we use the homogenization theory to map uncertainty in the microscopic parameters and processes to their macroscopic counterparts. Sources of uncertainty, as well as representations of randomness in the microscopic and macroscopic models, are described below.

2.1. Pore-scale model

At the pore-scale, the evolution of a solute concentration $c(\mathbf{x}, t)$ (mol/ ℓ^3), at point $\mathbf{x} \in \mathcal{P}$ and time $t > 0$, is governed by the evolution equation

$$\frac{\partial c}{\partial t} = \nabla \cdot (D \nabla c), \quad \mathbf{x} \in \mathcal{P}, \quad t > 0, \quad (1)$$

where $D(\mathbf{x})$ (ℓ^2/T), $\mathbf{x} \in \mathcal{P}$, is the pore-scale diffusion coefficient. The spatial variability of D allows for Fickian diffusion through mesopores and Knudsen diffusion through nanotubes. This equation is subject to the uniform initial condition

$$c(\mathbf{x}, 0) = c_{\text{in}}, \quad \mathbf{x} \in \mathcal{P},$$

and the boundary condition

$$-D \mathbf{n} \cdot \nabla c = q_m \frac{\partial s}{\partial t}, \quad \mathbf{x} \in \Gamma, \quad t > 0,$$

where q_m and $s(\mathbf{x}, t)$ are related to the sorption properties of the material surface Γ . Specifically, $q(\mathbf{x}, t) = q_m \cdot s(\mathbf{x}, t)$ is the adsorption amount per unit area of Γ (mol/ ℓ^2), q_m (mol/ ℓ^2) is the maximal adsorption amount, and $s(\mathbf{x}, t)$ is the fractional coverage of Γ . The fractional coverage is assumed to follow Lagergren's pseudo-first-order rate equation,

$$\frac{ds}{dt} = \gamma (s_{eq} - s),$$

where γ ($1/T$) is the adsorption rate constant and the equilibrium adsorption coverage fraction s_{eq} satisfied Langmuir's adsorption isotherm,

$$s_{eq} = \frac{Kc}{1 + Kc}, \tag{2}$$

with the adsorption equilibrium constant K (ℓ^3/mol).

2.2. Darcy-scale model

At the macroscopic scale, the volume-averaged solute concentration u ,

$$u(\mathbf{x}, t) := \frac{1}{\|\check{\mathcal{V}}\|} \int_{\check{\mathcal{V}}(\mathbf{x})} c(\xi, t) d\xi = \frac{1}{\|\check{\mathcal{P}}(\mathbf{x})\|} \int_{\check{\mathcal{P}}(\mathbf{x})} c(\xi, t) d\xi = \frac{\phi}{\|\check{\mathcal{P}}\|} \int_{\check{\mathcal{P}}(\mathbf{x})} c(\xi, t) d\xi, \quad \mathbf{x} \in \mathcal{V},$$

treats a porous material as a continuum. Here, $\|\cdot\|$ denotes the volume of a domain and $\phi := \|\mathcal{P}\|/\|\mathcal{V}\| = \|\check{\mathcal{P}}\|/\|\check{\mathcal{V}}\|$ is the material porosity. Using homogenization via multiple-scale expansions [1], one can show that u satisfies a reaction-diffusion equation

$$\phi \frac{\partial u}{\partial t} = \nabla \cdot (\mathbf{D}_{eff} \nabla u) - \phi q_m \gamma_{eff} \frac{Ku}{1 + Ku}, \quad \mathbf{x} \in \mathcal{V}. \tag{3}$$

The effective diffusion coefficient \mathbf{D}_{eff} and the effective rate constant γ_{eff} are random, stemming from uncertainty in pore-scale structures and processes that are propagated by the homogenization map. Specifically, the effective rate constant γ_{eff} ($1/L$) is computed as

$$\gamma_{eff} = \frac{\|\check{\Gamma}\|}{\|\check{\mathcal{P}}\|}, \tag{4}$$

i.e., is defined solely by the pore geometry, and the effective diffusion coefficient \mathbf{D}_{eff} (L^2/T), a second rank tensor, depends on both the pore geometry and the pore-scale processes. The latter is computed in terms of a closure variable χ , given by

$$\mathbf{D}_{eff} = \frac{1}{\|\check{\mathcal{V}}\|} \int_{\check{\mathcal{P}}} (\mathbf{I} + \nabla_{\xi} \chi^T) d\xi, \tag{5}$$

where \mathbf{I} is the identity matrix. The closure variable $\chi(\xi)$ is a $\check{\mathcal{V}}$ -periodic vector defined on $\check{\mathcal{P}}$, which satisfies the Laplace equation

$$\nabla_{\xi} \cdot (D \nabla_{\xi} \chi) = \mathbf{0}, \quad \xi \in \check{\mathcal{P}}, \tag{6}$$

subject to the normalizing condition

$$\langle \chi \rangle := \frac{1}{\|\check{\mathcal{V}}\|} \int_{\check{\mathcal{P}}} \chi(\xi) d\xi = \mathbf{0}, \tag{7}$$

the boundary condition along the fluid-solid segments $\check{\Gamma}$,

$$\mathbf{n} \cdot \nabla_{\xi} \chi = -\mathbf{n} \cdot \mathbf{I}, \quad \xi \in \check{\Gamma}, \tag{8}$$

and $\check{\mathcal{V}}$ -periodic boundary condition on the remaining fluid segments of the boundary of $\check{\mathcal{P}}$. For the hierarchical nanoporous material in Fig. 1, these auxiliary conditions reduce to

$$\chi_1(-a, \xi_2) = \chi_1(a, \xi_2) = 0, \quad \frac{\partial \chi_1}{\partial \xi_2}(\xi_1, 0) = \frac{\partial \chi_1}{\partial \xi_2}(\xi_1, b) = 0, \tag{9}$$

and

$$\chi_2(\xi_1, 0) = \chi_2(\xi_1, b) = 0, \quad \frac{\partial \chi_2}{\partial \xi_1}(-a, \xi_2) = \frac{\partial \chi_2}{\partial \xi_1}(a, \xi_2) = 0, \tag{10}$$

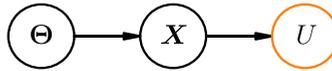


Fig. 2. A Bayesian Network describing the components of the full statistical model P , in (13), for the multiscale porous media system takes into account the joint PDF $P(\Theta)$ of the input parameters, the PDF $P(\mathbf{X} | \Theta)$ of the upscaling variable that maps pore-scale properties to Darcy-scale variables, and the PDF $P(U | \mathbf{X})$ related to Darcy-scale QoIs. This figure and other Bayesian Networks are produced using [14].

where

$$a = R \cos \theta \quad \text{and} \quad b = 2R \cos \left(\sin^{-1} \left(\frac{d}{2R} \right) \right) + l = 2R \sqrt{1 - \frac{d^2}{4R^2}} + l$$

are, respectively, the width and height of the unit cell $\check{\mathcal{V}}$. In the next section, we introduce a representation for uncertainty in the pore- and Darcy-scale models that will enable us to compartmentalize various stochastic modeling and statistical inference tasks.

3. Bayesian Network formulation for random PDE models of multiscale porous media

A Bayesian Network is a directed graph structure in which each node represents a random variable with an associated PDF and each edge encodes a conditional relationship [7]. A Bayesian Network PDE incorporates these structured probabilistic models into a forward physical model and allows one to capture conditional relationships, including correlations, constraints, and causality, in a systematic way. The key components of the full statistical model include:

- (i) inputs $\Theta = \{R, \theta, d, l\}$, a random vector related to the parameters describing pore-scale features in Fig. 1;
- (ii) upscaling variable \mathbf{X} , a random vector related to the closure equations (6)–(8); and
- (iii) QoI U , a random macroscopic quantity, such as macroscopic parameters \mathbf{D}_{eff} and γ_{eff} in (4) and (5) or a functional of u in (3).

In the context of the multiscale model, the relationships between problem scales are naturally causal: the PDF of U depends causally on the PDF of \mathbf{X} since, for example, the closure variable χ determines u and \mathbf{D}_{eff} . In turn, the PDF of \mathbf{X} depends causally on the PDF of Θ since the pore-scale parameter values determine χ , e.g., via (9) and (10). The Bayesian Network in Fig. 2 describes these causal relationships for the full statistical model where the directed network structure indicates that there is only one-way communication between the different scales.

The Bayesian Network in Fig. 2 encodes the relationships among the conditional and marginal PDFs for Θ , \mathbf{X} , and U . We denote the joint PDF of the input parameters by $P(\Theta)$. For simplicity we assume the statistical models for \mathbf{X} and U to be known, that is, uncertainty enters only through the parameters since we have fixed both the form of equations for χ and u in (3)–(8) and the numerical methods for approximating them. Under this assumption, the conditional PDF of \mathbf{X} given a sample Θ is the Dirac delta function

$$P(\mathbf{X} | \Theta) = \delta_{\mathbf{X}}(\mathbf{X} - \chi(\xi; \Theta)). \quad (11)$$

Similarly, the conditional PDF of U given a sample $\mathbf{X} = \chi$ is

$$P(U | \mathbf{X}) = \delta_U(U - u(\mathbf{x}, t; \mathbf{X})). \quad (12)$$

Then the full statistical model P is given by

$$P := P(\Theta, \mathbf{X}, U) = P(U | \mathbf{X}) P(\mathbf{X} | \Theta) P(\Theta) = \delta_U(U - u(\mathbf{x}, t; \mathbf{X})) \delta_{\mathbf{X}}(\mathbf{X} - \chi(\xi, t; \Theta)) P(\Theta). \quad (13)$$

In the remainder of this section, we discuss expanded Bayesian Networks for representing causal relationships between parameters within a single scale that allow us to encode correlations among pore scale features. In the context of the hierarchical material in Fig. 1, causal relationships stemming from domain knowledge, including expert knowledge and geometrical design constraints, provide a basis for positing conditional relationships that in turn give rise to correlations among pore scale features after selecting (uninformed) statistical models for each component of the Bayesian Network. In Section 3.1, we consider the assumption of independent priors for the pore-scale parameters and contrast this in Section 3.2 with causality arising from natural structural constraints encountered in engineering design.

Remark 1. Widely adopted approaches to uncertainty quantification, e.g., Monte Carlo methods, involve strategies for generating surrogates of the model P in (13). The direct application of these traditional UQ methods requires the form of P to be known. In the sequel, we introduce an approach relying on Rosenblatt transformations, generalized polynomial chaos expansions, and the popular UQ software package DAKOTA that makes sampling P feasible.

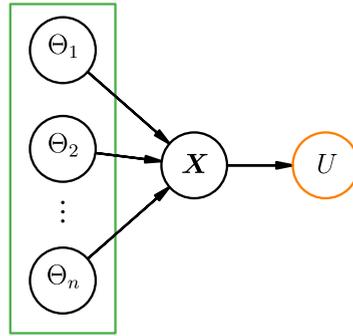


Fig. 3. A Bayesian Network describing the components of the full statistical model under the assumption of independent priors on pore-scale features $\Theta = (\Theta_1, \dots, \Theta_n)$. The flat structure of the Θ component in the model above contrasts with the rich structure of the Bayesian Network in Fig. 4 that captures causal relationships among the pore-scale features in order to ensure sampling geometries consistent with the hierarchical nanoporous material in Fig. 1 over the physically relevant hyperparameter ranges in Table 2.

Table 1

Narrow hyperparameter ranges for uninformed (uniform) priors on pore-scale features used with Bayesian Networks (15) and (20) that allow for comparison with results in [4] (cf. with the extended, physically relevant range in Table 2).

	$p = R$ (nm)	$p = \theta$ (rad)	$p = d$ (nm)	$p = l$ (nm)
(maximum) p_+	60	0.7	8	18
(minimum) p_-	10	0.07	4	8

3.1. Independent uniform priors

As a first simple case we revisit the analysis [4] of the hierarchical nanopore geometry in Fig. 1 to illustrate the Bayesian Network formulation for independent pore-scale features that do not take full advantage of domain knowledge. A naive model for representing uncertainty in each of the pore-scale parameters assumes pairwise independent priors. Recall that for $\Theta = (\Theta_1, \dots, \Theta_n)$, the variables $\Theta_1, \dots, \Theta_n$ are pairwise independent, $\Theta_i \perp \Theta_j$ for all $i, j = 1, \dots, n$ such that $i \neq j$, if and only if

$$P(\Theta_i, \Theta_j) = P(\Theta_i)P(\Theta_j),$$

where $P(\Theta_i)$ denotes the (marginal) PDF of Θ_i and $P(\Theta_i, \Theta_j)$ denotes the joint PDF of (Θ_i, Θ_j) [7]. Under the assumption of independent priors $P(\Theta_i)$, the joint PDF $P(\Theta)$ factors into the product of the priors,

$$P(\Theta) = \prod_{i=1}^n P(\Theta_i). \tag{14}$$

The full statistical model for P in (13) is then given by

$$P_0 := \delta_U(U - u(\mathbf{x}, t; \mathbf{X})) \delta_{\mathbf{X}}(\mathbf{X} - \chi(\xi, t; \Theta_1, \dots, \Theta_n)) \prod_{i=1}^n P(\Theta_i). \tag{15}$$

In the case of (15), the Bayesian Network has the special form in Fig. 3 where the independent priors assumption leads to an overall flatness in the graph structure for $P(\Theta)$. For PDF $P(\Theta_i)$ with finite variance, the pairwise independence assumption implies the variables Θ_i are uncorrelated. In particular, (15) holds with $|\Theta| = n = 4$ for the parameters $\{R, \theta, l, d\}$ that describe pore-scale features in Fig. 1. A model or PDF can then be specified for each Θ_i , for example, the uniform priors considered in [4].

In the absence of additional data that could be used to select a statistical model for each of the pore-scale features, it is reasonable to assume uninformed priors, i.e., that each $P(\Theta_i)$ is uniformly distributed, where the support of each distribution must be specified using expert knowledge. In Tables 1 and 2 two different ranges are given for specifying the hyperparameters for such uniform priors; a narrow range in Table 1 and an extended, physically relevant range in Table 2. However, assuming independent, uniformly distributed priors according to (14) with hyperparameter ranges in Table 2, it is possible to sample geometries that are inconsistent with Fig. 1. The Bayesian Network in Fig. 3 does not take into account geometrical structural constraints between the pore scale parameters that naturally arise when considering a periodic arrangement of unit cells \mathcal{V} as in the hierarchical nanoporous structure in Fig. 1. Next, we build Bayesian Networks based on causal relationships that encode such natural structural constraints thereby incorporating additionally available domain knowledge directly into the Bayesian Network random PDE formulation.

Table 2
Physical hyperparameter ranges for uninformed (uniform) priors on pore-scale features used with Bayesian Networks (15), (20) and (38). Expert knowledge is used to specify the support of each feature distribution in the absence of data for statistical model selection.

	$p = R$ (nm)	$p = \theta$ (rad)	$p = d$ (nm)	$p = l$ (nm)
(maximum) p_+	60	0.4π	60	60
(minimum) p_-	10	0.05π	5	1

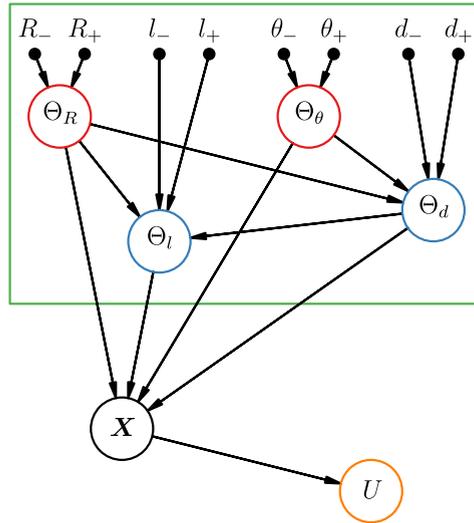


Fig. 4. The rich structures of the Bayesian Network above, representing the probabilistic model P_1 in (20), encodes causal relationships arising from structural constraints (cf. Fig. 5) that are absent in the model P_0 in (15) with independent priors in Fig. 3. In this Bayesian Network, conditional dependencies among the variables Θ induce various correlation structures that depend on the selected hyperparameters (cf. the correlation structure for model P_1 over the narrow range of hyperparameters in Fig. 6b vs. the physical range in Fig. 6c).

3.2. Correlations arising from pore-scale structural constraints

The inclusion of causal relationships suggested by geometrical constraints on the pore-scale parameters adds more complexity to the structure of Θ in the Bayesian Network in Fig. 3. Here and below we use Θ_p with labels $p \in \{R, \theta, l, d\}$ in place of indices where no confusion arises. Moreover, for each Θ_p we fix hyperparameters $\{p_+, p_-\}$ corresponding to upper and lower bounds on Θ_p in Tables 1 and 2; performing inference over the Bayesian Network to infer hyperparameters from relevant data is beyond the scope of this work. There are many different conditional relationships among pore-scale features consistent with Fig. 1 giving rise to a plurality of Bayesian Networks. Different causal relationships mirror distinct engineering or design processes (see Section 6) and yield different correlation structures among pore-scale features (cf. Fig. 6 and Fig. 11b).

Fig. 4 presents one possible Bayesian Network capturing causal relationships that allow one to encode structural constraints among the hierarchical pore-scale parameters $\{R, \theta, l, d\}$. Choosing $\{\Theta_R, \Theta_\theta\}$ as independent parameters to be consistent with our particular design goal and assuming uniform priors,

$$\Theta_R \sim \text{unif}(R_-, R_+) \quad \text{and} \quad \Theta_\theta \sim \text{unif}(\theta_-, \theta_+), \tag{16}$$

constrains both the distribution of Θ_d and of Θ_l . Specifically, in order for the sample $\Theta_d = d$ to be consistent with the features of the hierarchical nanopore structure in Fig. 1, the nanotube diameter cannot exceed the width of the unit cell \check{Y} , i.e., $d < 2R \cos \theta$ (see Fig. 5). That is, we are naturally able to specify the conditional PDF Θ_d given samples $\Theta_R = R$ and $\Theta_\theta = \theta$. Assuming an uninformed or uniform model for this conditional PDF we have

$$\Theta_d | \Theta_R, \Theta_\theta \sim \text{unif}(d_-, \min\{2\Theta_R \cos \Theta_\theta, d_+\}). \tag{17}$$

The length of the nanotube is bounded below by $l > 2R - \sqrt{4R^2 - d^2}$ when the vertical gap between the mesopores is zero (see Fig. 5) and then the PDF of Θ_l given Θ_R and Θ_d is

$$\Theta_l | \Theta_R, \Theta_d \sim \text{unif}(\max\{l_-, 2\Theta_R - \sqrt{4\Theta_R^2 - \Theta_d^2}\}, l_+), \tag{18}$$

where again we assume a uniform model for the conditional PDF of Θ_l .

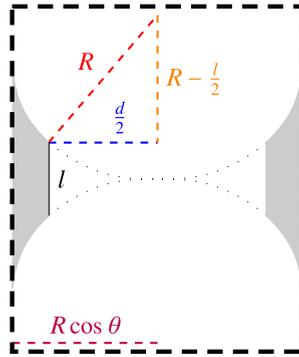
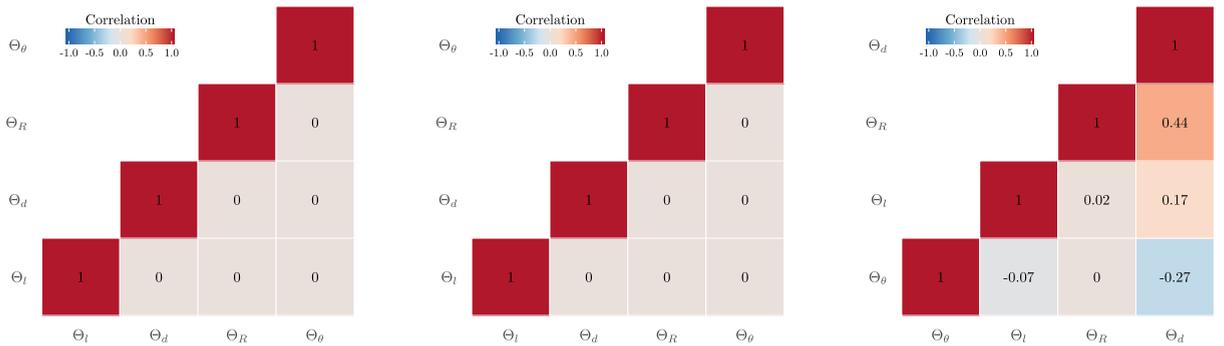


Fig. 5. The conditional distribution of Θ_d and Θ_l in (17) and (18) arises from geometric constraints that arise naturally when considering the hierarchical nanopore structure in Fig. 1. Above, we illustrate that the nanotube radius $d/2$ cannot exceed $R \cos \theta$, half the width of the unit cell \check{V} resulting in (17). Further, to exclude gaps between vertical mesopores that are less than zero, $l > 2R - \sqrt{4R^2 - d^2}$ from the right triangle in the diagram above resulting in (18).



(a) Independent priors (14) over narrow hyperparameter range. (b) Causal priors (19) over narrow hyperparameter range. (c) Causal priors (19) over physical hyperparameter range.

Fig. 6. The empirical correlation structure of $P(\Theta)$, the distribution on pore-scale features, is presented for the probabilistic model P_0 with independent priors (14) and the model P_1 with causal priors (19) over both the narrow and physical hyperparameter ranges in Tables 1 and 2, respectively. In general, the physical hyperparameter range is inaccessible to the model P_0 as the sampling strategy violates structural constraints by producing sample geometries inconsistent with Fig. 1. In contrast, the causal relationships included in P_1 enable sampling over the physical hyperparameter range thereby impacting predictions of QoIs (see Section 4), however the nontrivial correlation structure in Fig. 6c poses a challenge for GSA (see Section 5). (For interpretation of the colors in the figures, the reader is referred to the web version of this article.)

The joint distribution for the pore-scale input parameters is then given by

$$P(\Theta) = P(\Theta_R, \Theta_\theta, \Theta_l, \Theta_d) = P(\Theta_l | \Theta_R, \Theta_d)P(\Theta_d | \Theta_R, \Theta_\theta)P(\Theta_R)P(\Theta_\theta) \tag{19}$$

where each of the conditional and marginal PDFs is specified by (16)–(18) (cf. to the assumption of independent priors in (14)). Once a range of hyperparameters is fixed, one can sample from the PDFs (16) and (18) and hence, using (19), the correlation structure of the variable Θ can be computed empirically. In Fig. 6, we compare the empirical correlation structures obtained by computing the Pearson correlation coefficient for 10^6 samples of Θ for different priors $P(\Theta)$ and hyperparameter ranges. The empirical correlations in Figs. 6b and 6c are both computed assuming priors (19) with distributions (16) to (18) whereas the correlations in Fig. 6a assume independent uniform priors (14). Over the narrow range Table 1, Fig. 6b indicates the variables Θ_R , Θ_θ , Θ_l , and Θ_d are uncorrelated in contrast to the nontrivial correlations observed in Fig. 6c over the physical hyperparameter range Table 2. As expected, Fig. 6a confirms the model with independent, uniform priors are uncorrelated over the narrow hyperparameter range Table 1. Over physical hyperparameter ranges Table 2, independent sampling according to (14) yields geometries inconsistent with the hierarchical nanoporous material Fig. 1.

Incorporating (19) into the full statistical model (13), we obtain a new probabilistic model P_1 ,

$$P_1 := \delta_U(U - u(\mathbf{x}, t; \mathbf{X})) \delta_{\mathbf{X}}(\mathbf{X} - \chi(\xi, t; \Theta)) P(\Theta_l | \Theta_R, \Theta_\theta, \Theta_d)P(\Theta_d | \Theta_R, \Theta_\theta)P(\Theta_R)P(\Theta_\theta), \tag{20}$$

that describes the full statistical model for the multiscale system with the causal relationships in Fig. 4 assuming that uncertainty only enters through the parameters. When the models for U and \mathbf{X} are known and trivial (e.g., have a known analytic form), then sampling (20) is a straightforward task however this is typically not the case. In the next section, we review tools that will enable us to feasibly sample the statistical models for U and \mathbf{X} in order to carry out UQ and GSA.

3.3. Constructing a physics-informed probabilistic model for macroscopic QoIs

We are interested in functionals $g = g(\Theta) = g(\mathbf{D}_{\text{eff}}(\Theta), \gamma_{\text{eff}}(\Theta))$ such as the projections,

$$D_L := \mathbf{D}_{\text{eff}}^{11} \quad \text{and} \quad D_T := \mathbf{D}_{\text{eff}}^{22}, \tag{21}$$

corresponding to the longitudinal and the transverse components of the effective diffusion coefficient tensor. In the hierarchical nanoporous media in Fig. 1, D_L is related to diffusion through nanotunnels/mesopores and D_T through nanotubes. In the numerical experiments presented in the sequel, we report uncertainty in these macroscopic quantities by giving an estimate of the PDF $f_g \sim P(g | \Theta)$ for the PDF of g given knowledge of pore-scale inputs. In the context of the homogenized pore- to Darcy-scale model in Fig. 2, we have the representation

$$P(g | \Theta) = P(g | \mathbf{X})P(\mathbf{X} | \Theta)P(\Theta). \tag{22}$$

For example, the univariate PDF for $D_L = g(\mathbf{D}_{\text{eff}}(\Theta), \gamma_{\text{eff}}(\Theta))$ based on the Bayesian Network in Fig. 4 is

$$P(D_L | \Theta) = P(D_L | \mathbf{X})P(\mathbf{X} | \Theta)P(\Theta_l | \Theta_R, \Theta_\theta, \Theta_d)P(\Theta_d | \Theta_R, \Theta_\theta)P(\Theta_R)P(\Theta_\theta). \tag{23}$$

In general, the form of the statistical model or PDF for $P(g | \mathbf{X})$ and hence $P(g | \Theta)$ in (22) is unknown. The PDF in (22) can be estimated empirically using simulations of the forward model, however, sampling (22) based strictly on input-output pairs may be computationally expensive due to the high cost involved in simulating the forward model. The estimation of QoIs with high fidelity, therefore, poses a computational challenge.

The computation of estimates for QoIs of the form (22) is made feasible using a two-step procedure that relies on first finding a truncated generalized polynomial chaos expansion (gPCE) for the variable g and second using this surrogate to build an appropriate kernel density estimator (KDE) for the desired distribution. A surrogate \hat{g} for $g(\Theta)$ is given by the gPCE [15],

$$g(\Theta) = \sum_{i=0}^{\infty} G_i \Psi_i(\Theta) \approx \sum_{i=0}^{N_{PC}} G_i \Psi_i(\Theta) =: \hat{g}, \tag{24}$$

where $\Psi_i(\Theta)$ are an orthogonal multivariate polynomial basis, G_i are the expansion coefficients, and the expansion is truncated after N_{PC} terms such that

$$N_{PC} - 1 = \prod_{i=1}^{N_p} (1 + \kappa_i), \tag{25}$$

where κ_i is the polynomial order bound for the i th dimension and N_p is the number of parameters (to give a sense of scale, the numerical experiments in Section 4.2 use $N_p = 4$, $\kappa_i = 4$, and $N_{PC} = 626$).

The second step involves producing N samples \hat{g}^k using the gPCE (24) corresponding to realizations Θ^k for $k = 1, \dots, N$, for a fixed number N . These samples are then used to construct a KDE $\bar{f}_{\hat{g}}$ for the desired density f_g , for example, using a Gaussian-kernel,

$$\bar{f}_{\hat{g}}(\eta) = \frac{1}{N\sqrt{2\pi h^2}} \sum_{k=1}^N \exp \left[-\frac{(\eta - \hat{g}^k)^2}{2h^2} \right], \tag{26}$$

where h is the kernel bandwidth; for the multivariate density $\bar{f}_{\hat{g}_1, \hat{g}_2}$ we similarly employ Gaussian-kernels

$$\bar{f}_{\hat{g}_1, \hat{g}_2}(\eta_1, \eta_2) = \frac{1}{N\sqrt{2\pi h_1 h_2}} \sum_{k=1}^N \exp \left[-\frac{(\eta_1 - \hat{g}_1^k)^2}{2h_1^2} - \frac{(\eta_2 - \hat{g}_2^k)^2}{2h_2^2} \right], \tag{27}$$

with bandwidths h_1 and h_2 . To have a high-fidelity representation of the desired density it is desirable to choose $N \gg N_{PC}$.

The software DAKOTA [13] was used to automatize the process of computing the coefficients, basis functions, and truncation parameters appearing in (24). This approach is taken in [4] in the context of independent priors with the aim of computing Sobol' indices for global sensitivity analysis. Although generalizations of gPCE that handle correlated inputs exist (e.g., [10,16]), we instead present a recipe for obtaining the desired gPCE in (24) that non-intrusively uses existing codes and software packages such as DAKOTA that implement methods assuming uncorrelated inputs in the next section.

4. Quantifying uncertainty in Darcy-scale flow variables

The present section deals with uncertainty quantification for the Bayesian Network PDE model for multiscale porous media outlined in Sections 2 and 3. The causal relationships encoded by the Bayesian Network for the full statistical model in Fig. 2 propagate uncertainty from pore-scale parameters to Darcy-scale variables via the homogenization map in (4) to (10). Together, these allow one to study systematically the impact of microscopic structural uncertainty on macroscopic flow variables. At present, we incorporate the Bayesian Networks constructed in Section 3 for informed priors into the random PDE homogenization framework thereby examining through simulations and numerical experiments the role of causality in predictions of Darcy-scale flow variable QoIs. Specifically, we report on numerical experiments concerning the joint and marginal distributions of Darcy-scale flow variables where uncertainty stems from pore-scale features with correlations arising from structural constraints encoded by the Bayesian Network in Fig. 4. As these numerical experiments utilize DAKOTA [13] to compute gPCE surrogates for Darcy-scale variables, we first illustrate in Section 4.1 a technique for decorrelating inputs to allow the non-intrusive use of existing codes and software packages. This workflow, which relies on Rosenblatt transformations, gPCEs, and DAKOTA, enables UQ and GSA by making it feasible to sample from the desired QoI with respect to a given statistical model P in (13).

4.1. Non-intrusive input decorrelation using Rosenblatt transforms

Many variance-based methods for uncertainty quantification and sensitivity analysis, such as Sobol' indices, and hence the popular software packages that implement these methods, require models that assume statistically independent inputs. Bayesian Networks encode correlations, see Fig. 6, through the specification of causal relationships (in the case of Figs. 3 and 4 the relationships between scales and within single scales are also causal). Presently, we highlight how the Rosenblatt transform can be used to decorrelate inputs by mapping a vector of random variables with a specified joint distribution onto a vector of independent uniform random variables when the conditional distributions are known. This procedure, in Algorithm 1 below, enables the non-intrusive use of DAKOTA and existing codes for solving the forward model for our application of interest when the conditional dependencies are represented using Bayesian Networks.

The Rosenblatt transform [17] turns the problem of sampling a general joint distribution into the problem of sampling a vector of independent $\text{unif}(0, 1)$ random variables. Let $\mathbf{X} = (X_1, \dots, X_k)$ be a random vector with a continuous joint cumulative distribution function $F(x_1, \dots, x_k)$. Define a transform, $\mathcal{T}(x_1, \dots, x_k) = (z_1, \dots, z_k)$, given by

$$\begin{aligned} z_1 &= P(X_1 \leq x_1) = F_1(x_1), \\ z_2 &= P(X_2 \leq x_2 \mid X_1 = x_1) = F_{2|1}(x_2 \mid x_1), \\ &\vdots \\ z_k &= P(X_k \leq x_k \mid X_{k-1} = x_{k-1}, \dots, X_1 = x_1) = F_{k|k-1, \dots, 1}(x_k \mid x_{k-1}, \dots, x_1), \end{aligned} \tag{28}$$

where $F_{i|j}$ is the conditional cumulative distribution function of X_i given X_j , i.e., $F_{i|j}(x_i|x_j) = P(X_i < x_i \mid X_j = x_j)$. The Rosenblatt transform, $\mathbf{Z} := \mathcal{T}(\mathbf{X})$, yields $\mathbf{Z} = (Z_1, \dots, Z_k)$ uniformly distributed on the k -dimensional hypercube, that is, Z_1, \dots, Z_k are independent and identically distributed (IID) $\text{unif}(0, 1)$ random variables. Note that this transform depends on the ordering of the elements in the vector \mathbf{X} due to the serial nature of the conditioning; we denote the Rosenblatt transform and the inverse, when it exists, associated with the ordering of a particular vector \mathbf{X} with a subscript, e.g., $\mathcal{T}_{\mathbf{X}}$.

In our application, a target vector Θ with a structure encoded by a Bayesian Network can be obtained by applying the inverse Rosenblatt transform to a vector of independent uniform variables. For example, given the random vector of parameters $\Theta = (\Theta_R, \Theta_\theta, \Theta_l, \Theta_d)$ with joint distribution in (19), the transform (28) simplifies to

$$\begin{aligned} z_1 &= F_R(x_1), \\ z_2 &= F_{\theta|R}(x_2 \mid x_1) = F_\theta(x_2) \quad (\text{since } \Theta_R \perp \Theta_\theta), \\ z_3 &= F_{d|R,\theta}(x_3 \mid x_2, x_1), \\ z_4 &= F_{l|d,R}(x_4 \mid x_3, x_1) \quad (\text{since } \Theta_l \perp \Theta_\theta \mid \Theta_d), \end{aligned}$$

due to the independence and conditional independence of the variables [7], as indicated in Fig. 4 by the absence of causal relationships between Θ_R and Θ_θ and between Θ_l and Θ_d , respectively. Thus, the Rosenblatt transform is given by

$$\mathcal{T}_\Theta(\Theta) = \mathcal{T}_\Theta(\Theta_R, \Theta_\theta, \Theta_l, \Theta_d) = (F_R(\Theta_R), F_\theta(\Theta_\theta), F_{d|R,\theta}(\Theta_d \mid \Theta_R, \Theta_\theta), F_{l|R,d}(\Theta_l \mid \Theta_R, \Theta_d)) =: \mathbf{Z}$$

where, using the statistical models indicated in (16) to (18), the components of \mathbf{Z} are

$$\begin{aligned}
 Z_1 = F_R(\Theta_R) &= \frac{\Theta_R - R_-}{R_+ - R_-}, \\
 Z_2 = F_\theta(\Theta_\theta) &= \frac{\Theta_\theta - \theta_-}{\theta_+ - \theta_-}, \\
 Z_3 = F_d(\Theta_d | \Theta_R, \Theta_\theta) &= \frac{\Theta_d - d_-}{\min\{\Theta_R \cos \Theta_\theta, d_+\} - d_-}, \\
 Z_4 = F_{l_{R,d}}(\Theta_l | \Theta_R, \Theta_d) &= \frac{\Theta_l - \max\{l_-, 2\Theta_R - \sqrt{4\Theta_R^2 - \Theta_d^2}\}}{l_+ - \max\{l_-, 2\Theta_R - \sqrt{4\Theta_R^2 - \Theta_d^2}\}}.
 \end{aligned}$$

The corresponding inverse Rosenblatt transform $\Theta = \mathcal{T}_\Theta^{-1}(\mathbf{Z})$ is, component-wise,

$$\begin{aligned}
 \Theta_R &= Z_1(R_+ - R_-) + R_- \\
 \Theta_\theta &= Z_2(\theta_+ - \theta_-) + \theta_- \\
 \Theta_d &= Z_3(\min\{2\Theta_R \cos \Theta_\theta, d_+\} - d_-) + d_- \\
 \Theta_l &= Z_4(l_+ - \max\{l_-, 2\Theta_R - \sqrt{4\Theta_R^2 - \Theta_d^2}\}) + \max\{l_-, 2\Theta_R - \sqrt{4\Theta_R^2 - \Theta_d^2}\},
 \end{aligned}$$

where the target vector Θ has the distribution $P(\Theta)$ given in (19). Thus, the inverse transform maps a random vector $\mathbf{Z} \sim \text{unif}(0, 1)^k$ into a target distribution $P(\Theta)$ using knowledge of the conditional dependencies associated with the Bayesian Network for $P(\Theta)$, in particular, where the conditional distribution Θ_j given $\Theta_{j-1}, \dots, \Theta_1$ is known for each $j = 1, \dots, k$.

Together with the Bayesian Network PDE, Rosenblatt transforms provide a strategy for non-intrusively incorporating constraints and correlations into existing computational frameworks. Algorithm 1 describes the use of DAKOTA for computing surrogates for Darcy-scale QoIs based on correlated inputs given an inverse Rosenblatt transform and an existing solver for the forward problem. A surrogate \hat{g} for a QoI $g(Y(\Theta))$, e.g., the gPCE coefficients G in (24) and truncation parameter N_{PC} in (25), can be computed with DAKOTA using several forward simulations of the response $Y = \mathcal{M}(\Theta)$ where \mathcal{M} denotes the portion of the forward model solver that maps a random input Θ to Y . For example, if the QoI is $g(Y(\Theta)) = D_L$ then \mathcal{M} would correspond to the projection $\mathbf{D}_{\text{eff}}^{11}$ in (21) of the solution to the coupled system (5) with (6) to (8). The compositional model $Y = (\mathcal{M} \circ \mathcal{T}^{-1})(\mathbf{Z})$ that first employs the inverse Rosenblatt transform provides a non-intrusive means of computing with DAKOTA since the statistics of the output of $\mathcal{M} \circ \mathcal{T}^{-1}$ in response to \mathbf{Z} are identical to the statistics of the output of \mathcal{M} in response to Θ [18]. An important observation is that Algorithm 1 returns a surrogate \hat{g} for $g(Y(\Theta))$ with respect to the input variables \mathbf{Z} , not for Θ directly, and therefore care must be taken in the interpretation of Sobol' indices [8]. In the next section, we observe that the density functions for effective Darcy-scale QoIs exhibit non-Gaussian behavior. Together, these two challenges—correlated inputs and non-Gaussian QoIs—motivate the investigation of moment-independent global sensitivity indices in Section 5.

Algorithm 1: Decorrelate inputs via Rosenblatt transforms for non-intrusive scientific computing.

```

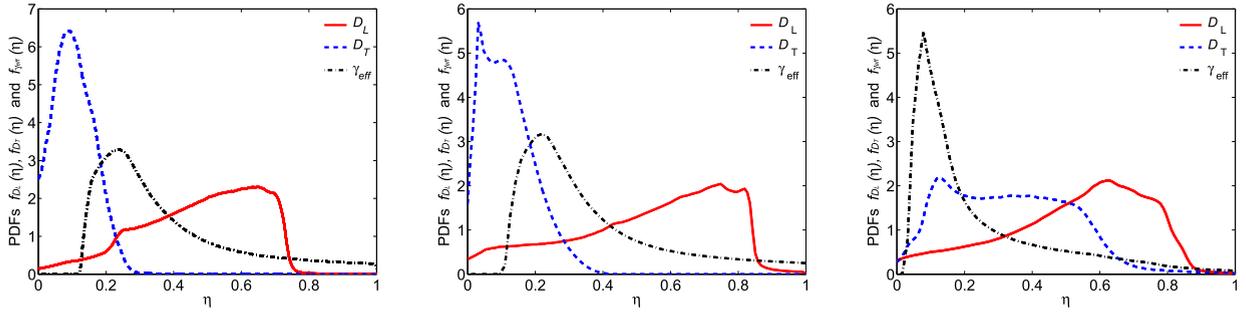
input :  $\mathcal{T}_\Theta^{-1}$                                 ▷ inverse Rosenblatt transform
input :  $\mathcal{M}$                                     ▷ forward model solver
output:  $\hat{g}_M$                                   ▷ surrogate for  $g(Y(\Theta))$  e.g., gPCE in (24)

begin
  DAKOTA as wrapper to produce surrogate using M input-output simulations
  for i ← 1 to M do
    sample IID  $Z_k \sim \text{unif}(0, 1), k = 1, \dots, n$ 
     $\mathbf{Z}_i \leftarrow (Z_1, \dots, Z_n)$ 
     $\Theta \leftarrow \mathcal{T}_\Theta^{-1}(\mathbf{Z}_i)$ 
     $Y_i \leftarrow \mathcal{M}(\Theta)$                                 ▷ DAKOTA maps independent  $\mathbf{Z}_i \mapsto Y_i$ 
   $\hat{g}_M \leftarrow \text{DAKOTA}(Y_1(\mathbf{Z}_1), \dots, Y_M(\mathbf{Z}_M))$ 
  return  $\hat{g}_M$                                             ▷ based on M input-output simulations

```

4.2. Numerical experiments: incorporating causality in predictions of Darcy-scale flow

For the application of interest, generating sufficiently many samples for UQ and GSA is a real challenge. The numerical experiments that follow employ the following common setup that makes sampling the distribution of QoIs computationally feasible. DAKOTA is used as a wrapper to map random inputs on pore-scale parameters to Darcy-scale responses yielding gPCE surrogates for Darcy-scale variables, i.e., effective longitudinal diffusion D_L , effective transverse diffusion D_T , and effective sorption rate constant γ_{eff} .



(a) Independent priors (model P_0) over narrow hyperparameter range. (b) Causal priors (model P_1) over narrow hyperparameter range. (c) Causal priors (model P_1) over physical hyperparameter range.

Fig. 7. A comparison of the marginal densities for Darcy-scale QoIs above highlights the importance of incorporating causal relationships into the modeling process; the distribution of Darcy-scale QoIs in Fig. 7c for model P_1 in (20) with causal priors over the physical hyperparameter range in Table 2 are markedly different from the QoIs in Fig. 7a for model P_2 in (15) with independent priors over the narrow hyperparameter range in Table 1. The QoIs in Fig. 7b for model P_1 are expected to be qualitatively similar to the QoIs in Fig. 7a due to the similarities in the correlation structure for the priors over the narrow range of hyperparameters (cf. Fig. 6).

With regard to sampling input-output pairs for the generation of gPCE surrogates, we follow the decorrelation procedure described in Algorithm 1 for each sample Θ that relates to a possible configuration of pore-scale features consistent with the hierarchical nanoporous material in Fig. 1. In particular, Algorithm 1 allows seamless, non-intrusive integration with existing codes for the numerical solution of the multiscale forward model presented in Section 2. For the numerical solutions of the forward model, we first solve the closure equations (6) to (10) using a finite element code written in COMSOL and then compute the rate constant γ_{eff} in (4) and effective diffusions D_L and D_T in (21) by numerically evaluating the quadrature in (5). For the required gPCE surrogates, given in (24), we select for Ψ_i the Askey scheme of hypergeometric orthogonal polynomials [15] with $N_{PC} = 626$ (i.e., for $N_p = 4$ parameters we consider polynomials of degree $\kappa_i = 4$, for each $i = 1, \dots, N_p$, in (25)). These gPCE surrogates are then used to construct KDEs for the desired Darcy-scale flow variables. The KDE approximations for densities below employ the Gaussian-kernels for univariate and multivariate densities, described in (26) and (27), respectively, with $N = 10^8$ where the kernel bandwidths are estimated using a modified Sheather–Jones method [19].

As a first experiment, we compare simulations of Darcy-scale flow variables based on the causal-priors model P_1 in (20) to simulations based on the independent priors model P_0 in (15) (cf. Bayesian Networks in Figs. 3 and 4). Over the narrow range of hyperparameter values given in Table 1, we anticipate qualitative similarities in the resulting Darcy-scale outputs as both P_0 and P_1 exhibit statistically uncorrelated parameter distributions over this hyperparameter range as demonstrated by the empirical correlations in Figs. 6a and 6b. In Fig. 7, we observe that the marginal distributions for Darcy-scale flow variables based on model P_1 in Fig. 7b are qualitatively consistent with the marginals based on model P_0 in Fig. 7a. Likewise, in Fig. 8 the joint distributions (D_L, D_T) , $(D_T, \gamma_{\text{eff}})$, and $(\gamma_{\text{eff}}, D_L)$ for simulations based on model P_1 in Fig. 8b and model P_0 in Fig. 8a exhibit qualitative similarities. Although the qualitative nature of the simulated distributions suggests no remarkable difference in the physics over the narrow hyperparameter range, in general the simulations based on model P_1 follow a different sampling procedure than the simulations based on model P_0 .

A quantitative comparison suggests that models P_0 and P_1 lead to macroscopic outputs with distinct distributions with statistical significance. To compare the densities in Figs. 7a and 7b and in Figs. 8a and 8b quantitatively, we work directly with samples from the gPCEs, employing a two-way statistical test on the equality of distributions. Since the data in Figs. 7a and 7b and in Figs. 8a and 8b appear commensurable, we select a nonparametric Cramér test [19] indicated to be sensitive against location alternatives that is applicable to both univariate and multivariate distributions so as to have a consistent presentation. Although the estimated densities in Fig. 7b (in Fig. 8b) are superficially similar to the densities in Fig. 7a (respectively, in Fig. 8a), the statistical tests each based on 2000 sample values, summarized in Table 3, reject the hypothesis on equality of distributions with high statistical significance for all but one comparison.

In contrast to P_0 , model P_1 enables numerical experiments over physically relevant hyperparameter ranges, such as Table 2, by embedding domain knowledge through causal relationships thereby ensuring sampled geometries are consistent with Fig. 1. Over the physical range of hyperparameters, the marginal and joint densities in Figs. 7c and 8c, respectively, for the Darcy-scale flow variable are markedly different from the corresponding marginal and joint densities observed in Figs. 7a, 7b, 8a and 8b. In comparing Fig. 7c to Fig. 7b, we observe that density for the effective rate constant γ_{eff} becomes more positively skewed and more peaked suggesting less variance in the estimate of γ_{eff} over the physical hyperparameter range. In contrast, we observe in comparing Fig. 7c to Fig. 7b that the density for D_T becomes more uniform in distribution (and more like the distribution of D_L) suggesting that the simulations over the physical hyperparameter range realize a richer space of transverse diffusions. Similarly, in Fig. 8c the joint distributions for the Darcy-scale flow variables employing model P_1 over the physical hyperparameter range demonstrate similar qualitative changes, with the joint densities involving γ_{eff} narrowing in variability and realizing a more variety in the observed transverse diffusion D_T . Altogether, the differences

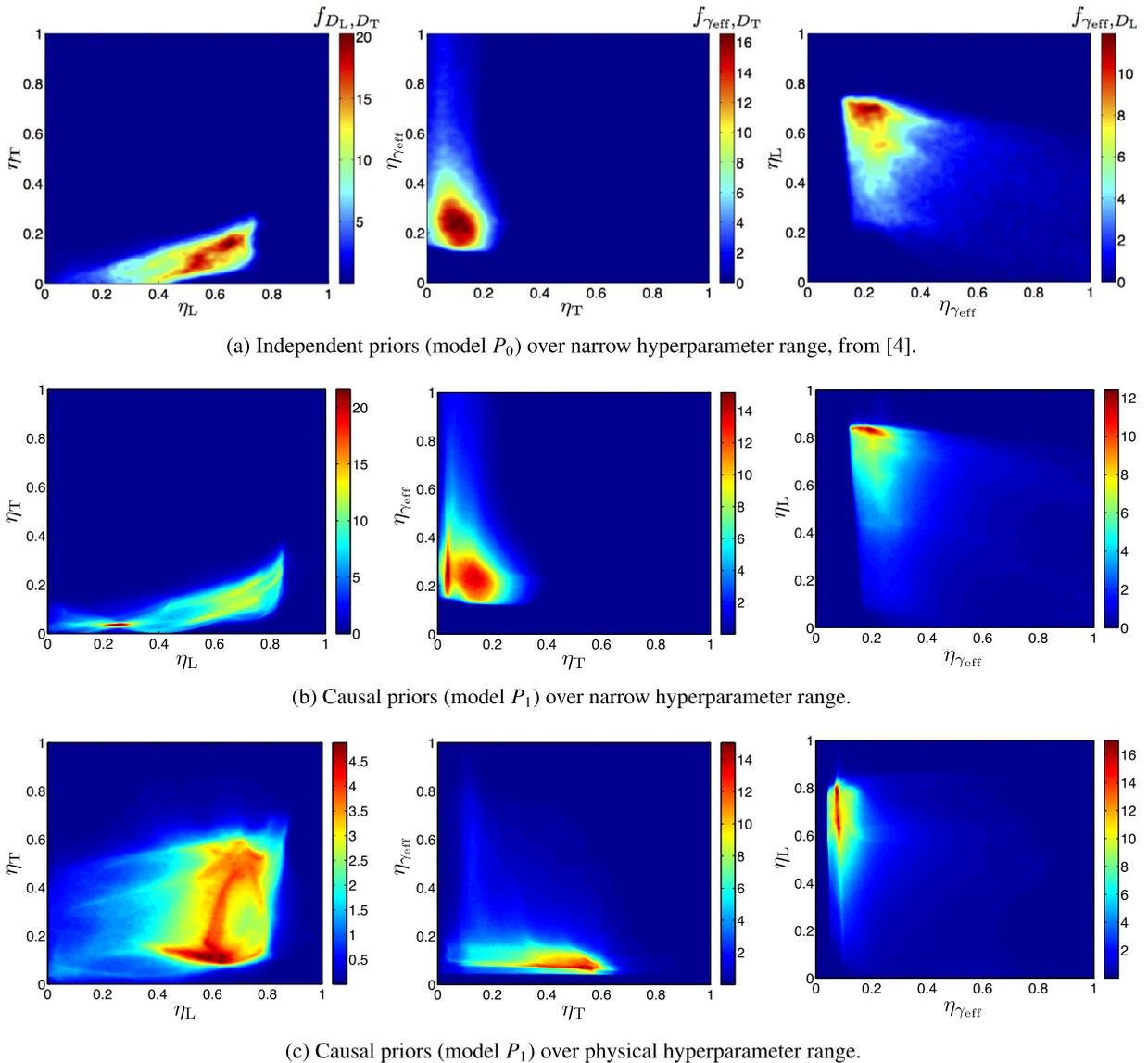


Fig. 8. A comparison of the joint densities for Darcy-scale flow variables above further underscores the importance of including causal relationships in the modeling process due to the impact for decision support. The causal relationships included in model P_1 in (20) guarantee that the pore-scale geometries sampled under P_1 are consistent with the hierarchical nanoporous material in Fig. 1 over the physical hyperparameter range in Table 2. In Fig. 8c, we observe that the QoIs related to model P_1 with physical hyperparameters realize a richer range of transverse diffusions than the QoIs depicted in Figs. 8a and 8b, which correspond to the models P_0 and P_1 over the narrow hyperparameter range in Table 1, thereby differentially impacting decision tasks.

Table 3

Results for a two-way nonparametric Cramér test [19] on the hypothesis of equality of the empirical distributions for comparable variables displayed in Figs. 7a, 7b, 8a and 8b each based on 2000 values sampled using a gPCE.

Variable	Cramér-statistic	Critical value	Conf. interval	p -value	Result	Figures
D_L	10.16	0.3116	0.95	<0.001	reject	} 7a vs. 7b
D_T	1.154	0.1049	0.95	<0.001	reject	
γ_{eff}	0.2232	0.3835	0.95	0.154	accept	
(D_L, D_T)	10.2	0.3154	0.95	<0.001	reject	} 8a vs. 8b
$(D_T, \gamma_{\text{eff}})$	1.3	0.3854	0.95	<0.001	reject	
$(\gamma_{\text{eff}}, D_L)$	7.525	0.4533	0.95	<0.001	reject	

between the Darcy-scale flow QoIs for model P_1 over physical hyperparameter range, with non-trivial correlations between pore-scale features, and for model P_0 or P_1 over the narrow hyperparameter range suggest flows with different physical properties. In the present context, these differences in physics can have a profound impact on decision support tasks

such as experimental design thereby underscoring the importance of including domain knowledge through causal relationships and structured priors in mathematical and statistical modeling processes. In the section that follows, we investigate methodologies for GSA that further emphasize the impact of causality on uncertainty in Darcy flow variables and QoIs.

Remark 2. In any of the Bayesian Network representations (15), (20), (23) and (38), uncertainty and error in the homogenization can be included by putting a distribution on \mathbf{X} that is not trivial, i.e., by replacing (11) with a distribution that captures error in the homogenization map or that compares distributions resulting from different upscaling techniques. Uncertainty and error in other relevant processes, such as the choice of the KDE, might also be incorporated into the Bayesian Network for the full statistical model and analyzed similarly. Thus, the Bayesian Network PDE formulation provides a systematic framework for building complete predictive models including forward physical models, transitions between scales, and uncertainties in parameters, mechanisms and parameter or model constraints.

5. Mutual Information for GSA and effect ranking in Bayesian Network PDEs

We are interested in simulating nanoporous media to inform the design of metamaterials with targeted macroscopic properties, such as effective diffusions and sorption constants, through engineered microscopic pore-scale structures. The sensitivity of macroscopic QoIs with respect to uncertainty in pore-scale properties is important to analyze as simulations of macroscopic material performance that are highly sensitivity to distributional changes in microscopic pore-scale processes and structures would undermine the generality of such investigations. In this context, sensitivity analysis corresponds to understanding how changes in input parameters and distributions affect the output simulation of a forward physical model that will ultimately inform an engineering decision task.

In general, sensitivity analysis is a key component of UQ and the analysis of model robustness, identifiability, and reliability. Local sensitivity analysis is suited to situations where the (hyper)parameters are known with some confidence and small perturbations are relevant. In contrast, in our present application of interest the mapping from pore-scale input distributions to Darcy-scale flow variables is nonlinear, includes several computational steps, and we have no *a priori* information on the form of the model for the Darcy-scale variables (cf. Section 3.3). These features demand GSA methods that explore the whole space of uncertain input factors [20].

Variance-based GSA methods, such as Sobol' indices [21] and total sensitivity indices [22], rank input factors and higher order interactions of input factors in terms of their contributions to the variance of a QoI. In particular, Sobol' sensitivity indices are used in [4] to analyze the global sensitivity of Darcy-scale QoIs to first and second order interactions among independent input parameters for the multiscale porous media model considered here. However, variance-based methods are not easily interpreted for correlated or dependent spaces of input parameters [10,8,23,9]. Moreover, we observe that the distributions for macroscopic Darcy-scale flow variables exhibit non-Gaussian behavior (cf. marginal and joint densities in Figs. 7 and 8 in Section 4 are highly skewed and even multimodal) and methods that rely on a finite number of moments may be insufficient to capture the full complexity of these interactions.

Our application of interest thus fixes the scope of the sensitivity analysis (i.e., global versus local) and additionally presents two challenges that are not easily handled with standard variance-based methods: (i) correlated and dependent inputs and (ii) highly skewed and even multimodal QoIs.

5.1. Mutual Information leverages structure inherent in Bayesian Network PDEs for sensitivity analysis

The Bayesian Network PDE framework represents the probabilistic models for components of a forward physical model, e.g., inputs, outputs, QoIs, transitions between scales, as a Bayesian Network thereby attaching additional layers of significance to the nodes in the network. On the one hand, this enables us to systematically build Bayesian Networks that incorporate domain knowledge through causal relationships that respect physical properties and expert knowledge. On the other hand, we are able to leverage the structure inherent in the Bayesian Network that suggest information theoretic measures of component sensitivity. An important contribution here is that we provide an interpretation of information theoretic sensitivity measures in the context of simulation-based studies (i.e., that incorporate forward physical models) intended to inform decision tasks such as optimal experimental design. In our numerical experiments below, the inclusion of structured priors with non-trivial correlations using Bayesian Network PDE yields different information-based GSA effect rankings than independent priors and moreover these rankings are more consistent with the anticipated physics of a model problem (cf. Section 5.3 and in particular Fig. 10).

We propose and interpret global sensitivity indices based on information theoretic concepts [11,12] that rely on empirical distributions acquired through simulation as opposed to moments. There is a rich literature on moment-independent indices for local sensitivity analysis [24–28] as well as GSA [29–34,23]. In particular, GSA indices based on various information theoretic notions are well established [35,31,36,37]. In [36], discrete Mutual Information-based sensitivity indices are demonstrated as an effective tool for discrete probability distributions arising in biochemical reaction networks in systems biology. For our application of interest, sensitivity indices and rankings based on the *differential Mutual Information* provide a suitable measure of effect that overcomes the twin challenges of causally related inputs and non-Gaussian QoI. The differential Mutual Information has explicit connections to more general information theoretic concepts and we provide interpretations of these in the context of uncertainty quantification and global sensitivity analysis.

The differential Mutual Information between continuous random variables V and W ,

$$I(V; W) := \iint \log \left(\frac{f_{V,W}(v, w)}{f_V(v)f_W(w)} \right) f_{V,W}(v, w) dv dw, \quad (29)$$

$0 \leq I(V; W) < \infty$, quantifies the statistical dependence between V and W , that is, the amount of shared information, provided that all of the densities above exist and the marginals are non-zero. Importantly, (29) applies to dependent random variables V and W , such as random variables that share a causal relationship, and we observe that $I(V, W) = 0$ if V and W are independent, i.e., if $P(V, W) = P(V)P(W)$. Further, (29) applies to random variables with very general marginal and joint distributions including distributions that are skewed and multimodal. The differential Mutual Information is precisely the relative entropy \mathcal{R} (or Kullback–Leibler divergence) between the joint and marginal densities,

$$I(V; W) = \mathcal{R}(P(V, W) \parallel P(V)P(W)), \quad (30)$$

a well-known pseudo-distance used, e.g., in variational inference [38,39], machine learning [40], and model selection [41]. Lastly, the differential Mutual Information $I(V, W)$ is the limiting value of the discrete Mutual Information (the supremum over all partitions of V and W) and therefore shares all of the same properties as its discrete counterpart [11]; in particular, we will compute the differential Mutual Information using empirical distributions.

5.2. Estimators for Mutual Information global sensitivity indices

For each QoI g , we consider a global sensitivity index,

$$S_{\Theta}(g) := I(g; \Theta), \quad (31)$$

based on the Mutual Information between a Darcy-scale QoI g and each uncertain pore-scale input parameter $\Theta \in \Theta = (\Theta_1, \dots, \Theta_{N_p})$. Intuitively, the index (31) measures the predictability of g given knowledge of Θ through the discrepancy between the joint density and the product of the marginal densities appearing in (29). We observe that in (31), $S_{\Theta}(g) \geq 0$ for all Θ ; in contrast, Sobol' indices that represent the correlated share of the total influence can become negative (see, e.g., [10]) which leads to a problem with interpretation.

We estimate the sensitivity index (31) using empirically estimated distributions, i.e., KDEs, for the density functions of Darcy-scale QoIs. Indeed, a benefit to using the Mutual Information is the availability of methods relying on plug-in estimators for (29) with corresponding numerical analyses (including [42–44]). Recall that we obtain approximate densities \tilde{f} using univariate and multivariate KDEs, in (26) and (27), respectively, that are in turn obtained using the gPCE surrogates \hat{g} defined in (24). Due to the availability of these surrogates, we do not sample input-output pairs as suggested by the form of (29) but instead consider the equivalent representation (change of measure),

$$I(V; W) = \iint \log \left(\frac{f_{V,W}(v, w)}{f_V(v)f_W(w)} \right) \frac{f_{V,W}(v, w)}{f_V(v)f_W(w)} f_V(v)f_W(w) dv dw. \quad (32)$$

Thus we compute the statistical estimator $\hat{S}_{\Theta}(g) \approx S_{\Theta}(g)$,

$$\hat{S}_{\Theta}(g) := \frac{1}{N} \sum_{k=1}^N \log \left(\frac{\tilde{f}_{\hat{g}, \Theta}(\hat{g}^k, \Theta^k)}{\tilde{f}_{\hat{g}}(\hat{g}^k) f_{\Theta}(\Theta^k)} \right) \frac{\tilde{f}_{\hat{g}, \Theta}(\hat{g}^k, \Theta^k)}{\tilde{f}_{\hat{g}}(\hat{g}^k) f_{\Theta}(\Theta^k)}, \quad (33)$$

based on a MC approximation of (32) using KDEs as plug-in estimates where appropriate.

In the left-hand side of Table 4, we report the value of the statistical estimator $\hat{S}_{\Theta}(g)$ related to the numerical experiments presented in Section 4.2 concerning the probabilistic model P_1 in (20) with causal inputs (see Fig. 4) over both the narrow and physical hyperparameter ranges in Tables 1 and 2. We are interested in the global sensitivity of the Darcy-scale QoIs $g = D_L$, $g = D_T$, and $g = \gamma_{\text{eff}}$ with respect to each of input parameters $\Theta \in \Theta = (\Theta_R, \Theta_{\theta}, \Theta_d, \Theta_i)$ representing the pore-scale features identified in the hierarchical nanoporous material in Fig. 1. The computed estimators (33) are based on first generating random variables Θ^k and \hat{g}^k for $k = 1, \dots, N = 10^7$ using the respective gPCE surrogate obtained by the workflow outlined in Section 4. These samples are used to form the respective KDE on an η -grid of size 128 by 128 which in turn are then used to form the plug-in quantity

$$X^j = \log \left(\frac{\tilde{f}_{\Theta, \hat{g}}(\Theta^j, \hat{g}^j)}{\tilde{f}_{\Theta}(\Theta^j) \tilde{f}_{\hat{g}}(\hat{g}^j)} \right) \frac{\tilde{f}_{\Theta, \hat{g}}(\Theta^j, \hat{g}^j)}{\tilde{f}_{\Theta}(\Theta^j) \tilde{f}_{\hat{g}}(\hat{g}^j)},$$

where we define $\log\left(\frac{0}{0}\right) \cdot \frac{0}{0} = 0$ (re-using 10^5 of the samples (Θ^j, \hat{g}^j) generated during density estimation). The sensitivity index is then estimated using the MC estimator, $\hat{S}_{\Theta}(\hat{g}) = \frac{1}{M} \sum_{j=1}^M X^j$ with $M = 10^5$ samples. We observe that the $M = 10^5$ samples utilized for the MC estimator is suggested to be sufficient by the convergence demonstrated in Fig. 9 for the experiment with model P_1 over the range Table 1; similar convergence observations were made for the other experiments.

Table 4

The global sensitivity index S in (31) based on the Mutual Information quantifies the effect of pore-scale uncertainties in terms of how additional knowledge of the input $\Theta \in \Theta$ reduces uncertainty in our prediction of the macroscopic Darcy-scale variables D_L , D_T , and γ_{eff} . Below, the estimator \widehat{S} in (33) and the associated ranking \widehat{r} in (35) are given for the probabilistic model P_1 in (20) over both the hyperparameter ranges in Table 1 and Table 2 (cf. graphical representation of rankings in Fig. 10b that also include error bars indicating two standard deviations); bold figures indicate maximal column values.

Θ	Mutual Information (33)			Ranking effects (35)			Hyperparameters
	$\widehat{S}_{\Theta}(D_L)$	$\widehat{S}_{\Theta}(D_T)$	$\widehat{S}_{\Theta}(\gamma_{\text{eff}})$	$\widehat{r}_{\Theta}(D_L)$	$\widehat{r}_{\Theta}(D_T)$	$\widehat{r}_{\Theta}(\gamma_{\text{eff}})$	
Θ_R	0.0419	0.0424	0.8955	0.0724	0.1177	0.8509	Table 1
Θ_{θ}	0.5074	0.2049	0.0366	0.8770	0.5692	0.0348	
Θ_d	0.0150	0.0878	0.0271	0.0259	0.2438	0.0258	
Θ_l	0.0143	0.0249	0.0932	0.0247	0.0692	0.0885	
Θ_R	0.0655	0.0714	0.2878	0.1354	0.2470	0.5242	Table 2
Θ_{θ}	0.3539	0.0312	0.0251	0.7312	0.1079	0.0457	
Θ_d	0.0225	0.1646	0.0823	0.0465	0.5697	0.1499	
Θ_l	0.0420	0.0218	0.1539	0.0868	0.0754	0.2802	

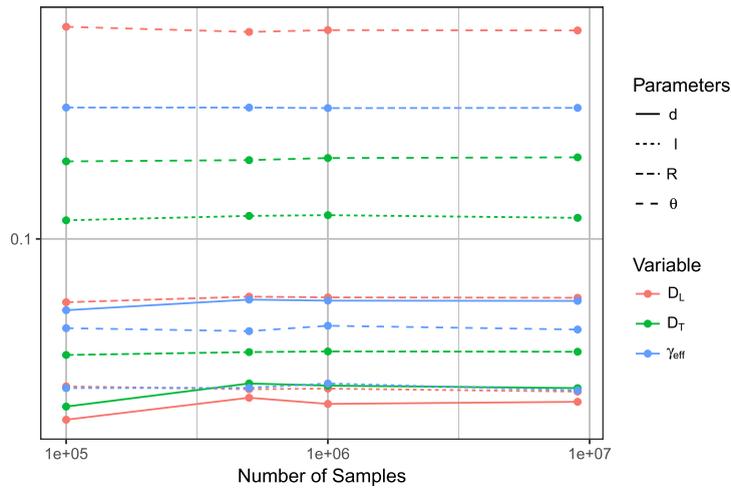


Fig. 9. The convergence of the MC estimator (33) for model P_1 in (20) over the narrow hyperparameter range in Table 1 for each Darcy-scale flow variable for each input parameter demonstrates that 10^5 samples is sufficient for the numerical experiments.

Remark 3. The differential Mutual Information (29) can be expressed as an expected value of the relative entropy between conditional and marginal distributions,

$$I(g; \Theta) = \mathbf{E}_{\Theta} [\mathcal{R}(P(g | \Theta) \| P(g))] , \tag{34}$$

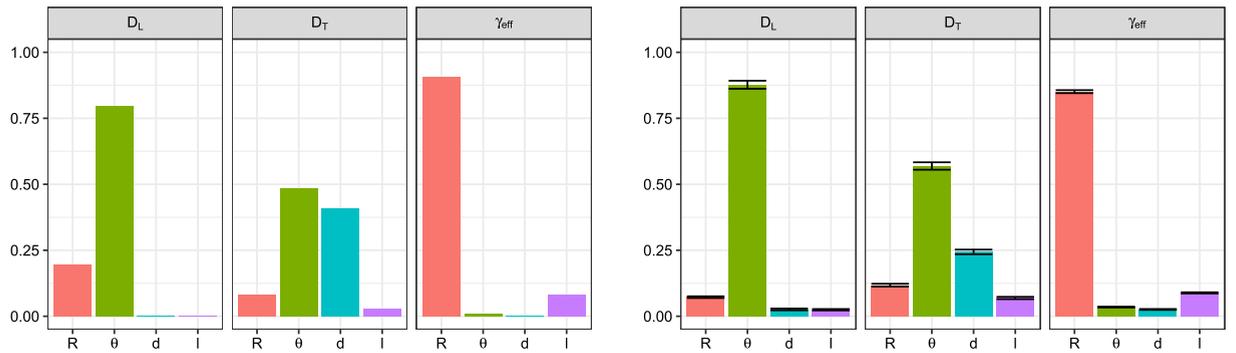
for the input-output pair $(g(Y(\Theta)), \Theta)$ provided $P(g, \Theta) = P(g, | \Theta)P(\Theta)$, i.e., all the densities exist. In [23], a family of sensitivity measures is given by replacing \mathcal{R} in (34) with a general class of Csiszár φ -divergences [45–47]; here we demonstrate an implementation using gPCE surrogates that requires the representation (32) and restrict our attention to the differential Mutual Information owing to the clear interpretation as a measure of effect in terms of statistical dependence and shared information.

Remark 4. One can also consider higher order effects that include interactions between a subset of parameters using the conditional differential Mutual Information that takes the form of conditional expectations of the relative entropy between joints and marginals, as in [36]. These higher order interactions can be interpreted similarly in terms of dependence and shared information.

5.3. Ranking impact of uncertainty in correlated pore-scale inputs on Darcy-scale QoIs

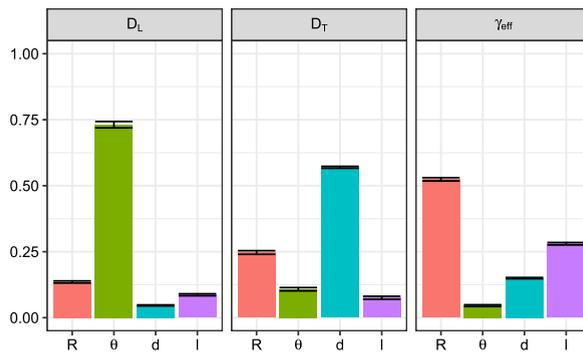
Using the global sensitivity index (31) we form the ranking,

$$r_{\Theta}(g) = \frac{S_{\Theta}(g)}{\sum_{V \in \Theta} S_V(g)} , \tag{35}$$



(a) First order Sobol' indices for independent priors model P_0 over narrow hyperparameter range (reproduced from data in [4] for the convenience of the reader).

(b) Ranking effects $\widehat{r}_{\Theta}(\hat{g})$ for causal model P_1 over narrow hyperparameter range.



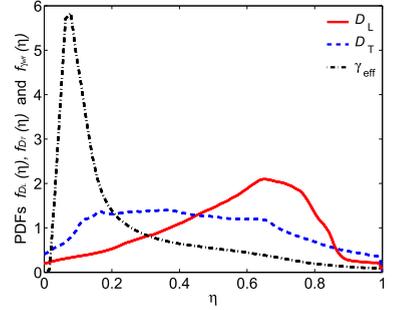
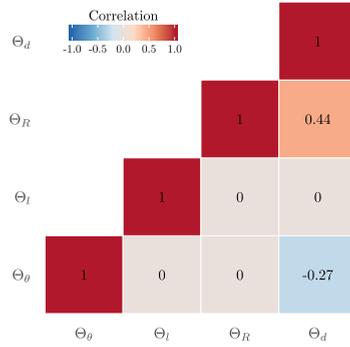
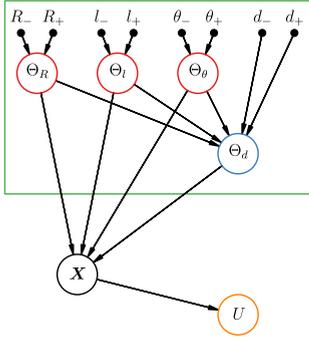
(c) Ranking effects $\widehat{r}_{\Theta}(\hat{g})$ for causal model P_1 over physical hyperparameter range.

Fig. 10. The rankings in Fig. 10c associated with the causal model P_1 in (20) over the physical hyperparameter range in Table 2 yield parameter rankings that are consistent with our understanding of the physics of the hierarchical nanoporous material in Fig. 1. In contrast, the Sobol' index rankings in Fig. 10a for the independent priors model P_0 in (15) over the narrow hyperparameter range in Table 1 suggest that the transverse diffusion D_T is most sensitive to the parameter θ , related to the angle of overlap between adjacent mesopores in a nanotunnel. The Mutual Information rankings in Fig. 10b for model P_0 are consistent with the Sobol' index rankings in Fig. 10a for model P_0 owing to the similarity of the correlation structures for pore-scale features over the narrow hyperparameter range (cf. Fig. 6). In Figs. 10b and 10c, error bars, that indicate the relative error associated with plus/minus two standard deviations of the computed sensitivity index, provide an indication of confidence in the ranking value.

$0 \leq r_{\Theta}(g) \leq 1$, of the relative contribution of each pore-scale parameter $\Theta \in \Theta$ to the global sensitivity of each Darcy-scale variable g . We then obtain the estimate $\widehat{r}_{\Theta}(g)$ reported in the right-hand side of Table 4, by using the estimator \widehat{S} in (33) in place of S . The values \widehat{r} are also displayed graphically in Fig. 10 with error bars that indicate the relative error associated with plus or minus two standard deviations of the computed sensitivity index to provide an indication of confidence in the ranking.

In Fig. 10b, we observe that the rankings suggested by the model P_1 in (20) with causal inputs in Fig. 4 over the narrow hyperparameter range in Table 1 are consistent with the Sobol' index rankings in Fig. 10a that were obtained in [4] for the model P_0 with independent priors. As observed in Section 4.2 this is to be expected due to the similarity in the correlation structures in Figs. 6a and 6b over the narrow hyperparameter range. The θ , which is related to the mesopore radius, is the most influential parameter for both the longitudinal diffusion D_L and the transverse diffusion D_T . The mesopore radius R , which is related to the size of the fluid-solid interface in Fig. 1, is the most influential parameter for the sorption rate constant γ_{eff} . In general, it is important to interpret these sensitivity rankings in the context of the hyperparameter range; the ranking \widehat{r}_{Θ_d} in Fig. 10b is likely to be small as the range for d in Table 1 is very narrow (i.e., the Darcy-scale QoIs are insensitive over the narrow range of admissible d values).

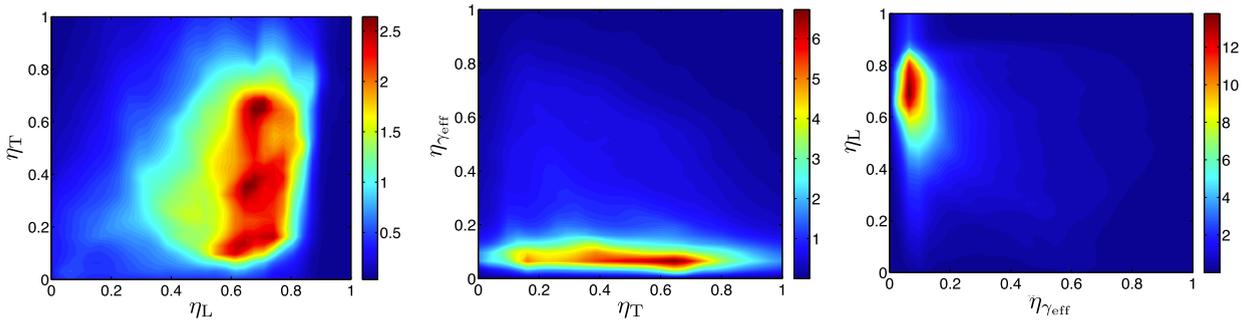
In contrast, the rankings in Fig. 10c for the model P_1 over the physical hyperparameter range in Table 2 demonstrates considerably different rankings to Figs. 10a and 10b thereby highlighting once again the impact of causal relationships and the Bayesian Network PDE modeling approach. The rankings in Fig. 10c indicate that while the θ and the R are still the most influential parameters for, respectively, the longitudinal diffusion, D_L , and sorption rate constant, γ_{eff} , it is d , which is related to the diameter of the nanotube, that is the most influential parameter for the transverse diffusion D_T through nanotubes. Thus, for the simple hierarchical nanoporous material in Fig. 1, the rankings in Fig. 10c using model P_1 over the physical



(a) Bayesian Network for model P_2 encoding different causal relationships from P_1 (cf. Figures 3 and 4).

(b) Correlation structure for P_2 over physical hyperparameter range in Table 2 (cf. Figure 6).

(c) Marginal KDEs for Darcy-scale flow variables for model P_2 over the physical hyperparameter range (cf. Figure 7).



(d) Causal priors (model P_2) over the physical hyperparameter range (cf. Figure 8).

Fig. 11. An alternative model P_2 in (38) encodes causal relationships that are different from model P_1 in (20) but also respects structural constraints realized by the hierarchical nanoporous material in Fig. 1. The model P_2 is viewed as mirroring engineering processes and design workflows that are distinct from those of P_1 .

hyperparameter range reflect our expectations of the physics better than the experimental observations with respect to models over the narrow range. Importantly, the rankings in Fig. 10 are moment-independent and therefore suitable for the non-Gaussian behavior of the Darcy-scale QoIs observed in Figs. 7 and 8. Moreover, as the rankings are based on the Mutual Information, they can be interpreted as ranking the impact of the pore-scale parameter (whether correlated or not) on the total uncertainty in the Darcy-scale quantity of interest. To summarize the practical outcomes of this experiment, we observe that the inclusion of structured priors with non-trivial correlations (i) yields Mutual Information GSA indices for Bayesian Network PDEs and corresponding rankings that are different from those obtained with uncorrelated and independent priors and (ii) moreover, these rankings are more consistent with the anticipated physics of a model problem.

6. Alternative probabilistic models reflect different engineering and design causal relationships

We view the construction of the full statistical model for the multiscale porous media system as mirroring engineering processes and design workflows. From this perspective, it may be desirable to build models containing different causal relationships among the pore-scale parameters than those previously considered. Recall that the model P_1 in (20) with causal priors given by the Bayesian Network in Fig. 4 is related to the design of nanotunnels/mesopores, i.e., the pore-scale features R and θ in the hierarchical nanoporous media in Fig. 1. In contrast, if it is desirable or possible to choose three aspects, such as the features R , θ , and l , independently while constraining the only remaining parameter, d , then one obtains a second model represented by the Bayesian Network in Fig. 11a.

The Bayesian Network in Fig. 11a corresponds to placing all of the constraints on Θ_d . That is, choosing a subset of independent parameters $\{\Theta_R, \Theta_\theta, \Theta_l\}$ and assuming uniform priors,

$$\Theta_R \sim \text{unif}(R_-, R_+) \quad \text{and} \quad \Theta_\theta \sim \text{unif}(\theta_-, \theta_+) \quad \text{and} \quad \Theta_l \sim \text{unif}(l_-, l_+), \tag{36}$$

constrains the distribution of Θ_d . Specifically, in order for the choice of d to be consistent with the features depicted in Fig. 1, then (i) the nanotube diameter $d < 2R \cos \theta$ and (ii) if the gap between vertical mesopores is zero than the nanotube

Table 5

Two-way nonparametric Cramér test [19] rejects the hypothesis of equality of the empirical marginal and joint distributions for comparable variables in Figs. 7c, 8c, 11c and 11d with high statistical significance (each test is based on 2000 values sampled using a gPCE).

Variable	Cramér-statistic	Critical value	Conf. interval	<i>p</i> -value	Result	Figures
D_L	4.688	0.3389	0.95	<0.001	reject	} 11c vs. 7c
D_T	79.08	0.7396	0.95	<0.001	reject	
γ_{eff}	0.4619	0.3958	0.95	0.029	reject	
(D_L, D_T)	72.31	0.7827	0.95	<0.001	reject	} 11d vs. 8c
$(D_T, \gamma_{\text{eff}})$	72.63	0.8284	0.95	<0.001	reject	
$(\gamma_{\text{eff}}, D_L)$	4.734	0.4882	0.95	<0.001	reject	

diameter must be less than $d < \sqrt{4lR - l^2}$ (i.e., to avoid nanotube “goiters”, see further Fig. 5). The corresponding conditional distribution is then given by

$$\Theta_d | \Theta_R, \Theta_\theta, \Theta_l \sim \text{unif}(d_-, \min\{\sqrt{4\Theta_l\Theta_R - \Theta_l^2}, 2\Theta_R \cos \Theta_\theta, d_+\}), \quad (37)$$

where again we assume a uniform model. The form of the full statistical model that includes the causal relationships in Fig. 11a is then given by

$$P_2 := \delta_U(U - u(\mathbf{x}, t; \mathbf{X})) \delta_{\mathbf{X}}(\mathbf{X} - \chi(\xi, t; \Theta)) P(\Theta_d | \Theta_R, \Theta_\theta, \Theta_l) P(\Theta_R) P(\Theta_\theta) P(\Theta_l), \quad (38)$$

where the distributions for conditionals and priors above are given in (36) and (37).

We observe that the Bayesian Networks in Fig. 4 and Fig. 11a are not equivalent. Over the physical range of hyperparameters in Table 2, we observe that correlation structure in Fig. 11d is distinct from Fig. 8c and thus the joint density $P(\Theta)$ for the Bayesian Networks in Fig. 11a and Fig. 4 are different. Although it may be possible to select priors to ensure equality between the distributions P_1 and P_2 , this is not in general the goal. Further, we can compare the Darcy-scale Qols for the model P_2 to the model P_1 in Fig. 11c over the physical hyperparameter ranges in Table 2, i.e., by comparing Fig. 11c to Fig. 7c and Fig. 11d to Fig. 8c. Importantly, incorporating constraints directly into the models P_1 and P_2 allows us to ensure that we sample from a joint distribution that ensures realistic geometries that are consistent with the hierarchical nanoporous material in Fig. 1 over the physical range of hyperparameters. On the one hand, the densities in Fig. 7c (and Fig. 8c) are qualitatively similar to those in Fig. 11c (respectively, Fig. 11d) owing to the closeness of the input densities $P(\Theta)$, see the empirical correlations in Figs. 6c and 11b. On the other hand, we observe that the models P_1 and P_2 lead to distinct distributions on Darcy-scale flow variables as the Cramér tests reported in Table 5 reject the hypothesis of equality of the empirical distributions with high statistical significance. We emphasize that we do not give here a method for selecting among various models but instead a collection of tools for breaking the stochastic modeling task into smaller, manageable components that enables a systematic way of building a full statistical model that incorporates engineering design constraints.

7. Conclusions

The sensitivity of Darcy-scale observables to changes in pore-scale properties and rigorous quantification of the uncertainty in predictions are some of the least studied aspects of multiscale models of flow and transport in porous materials. Our analysis leads to the following major conclusions.

- Causal relationships are natural and stem from physical or chemical constraints, engineering design, and expert knowledge. These relationships exist between model parameters, scales, and model components in multi-scale and multi-physics models. In the context of hierarchical nanoporous materials, we observe that causal relationships suggested by geometrical structural constraints among microscopic features yield non-trivial correlations among pore-scale model parameters over physical ranges.
- We incorporate domain knowledge-informed causal relationships, and thereby correlations, in a unified random, multiscale PDE model using structured probabilistic graphical models, in this case, Bayesian Networks. This perspective ultimately gives rise to Bayesian Network (random) PDEs.
- Due to causal relationships and resulting correlations between model parameters, GSA is not straightforward. Furthermore, predictive PDFs of Qols are not necessarily Gaussian (highly skewed, multimodal) or otherwise of known analytical form. For these two reasons, it is necessary to depart from moment-based sensitivity analysis, e.g., ANOVA methods, and deploy PDF-based methods relying on Mutual Information.
- The non-intrusive uncertainty quantification workflow relying on Rosenblatt transforms, gPCE surrogates for Qols (obtained using DAKOTA), and kernel density estimation techniques makes exhaustive sampling for prediction and uncertainty quantification feasible.
- The proposed Mutual Information global sensitivity indices for Bayesian Network PDE and corresponding Darcy-scale Qols yield parameter rankings that are consistent with our understanding of the physics of a hierarchical nanoporous

metamaterial. We demonstrate that correlations stemming from causal relationships turn out to be important in determining the most influential model parameters/mechanisms and impact predictions of QoIs.

The structure of Bayesian Networks or PGMs for the pore-scale parameter correlations can, in general, be learned from experimental or simulated data rather than being based on natural structural constraints as in this work and moreover they need not adopt causal relationships that mirror engineering design processes. One future direction includes learning the structure of the Bayesian Networks from available data and inferring the distributions of pore-scale features directly. The flexibility of the Bayesian Network PDE approach for incorporating domain knowledge and uncertainty into physical models will also enable explorations of model-form uncertainty that advance traditional uncertainty quantification techniques. Another future direction for research is to apply hybrid information divergences (e.g., [48]) to bound families of model predictions based on various pore-scale parameter representations and based on different upscaling techniques (i.e., different probabilistic models) to address questions of model-form uncertainty and tracking different levels of fidelity attached to system components.

Acknowledgements

This research was supported in part by the Department of Energy (DOE) Office of Science under grant 0000241868, by the Air Force Office of Scientific Research under grants FA9550-17-1-0417 and FA9550-18-1-0214, and by National Science Foundation under grants 1606192 and DMS-1515712.

References

- [1] X. Zhang, K. Urita, I. Moriguchi, D.M. Tartakovsky, Design of nanoporous materials with optimal sorption capacity, *J. Appl. Phys.* 117 (2015) 244304, <https://doi.org/10.1063/1.4923057>.
- [2] X. Zhang, D.M. Tartakovsky, Optimal design of nanoporous materials for electrochemical devices, *Appl. Phys. Lett.* 110 (2017) 143103, <https://doi.org/10.1063/1.4979466>.
- [3] B. Ling, A.M. Tartakovsky, I. Battiato, Dispersion controlled by permeable surfaces: surface properties and scaling, *J. Fluid Mech.* 801 (2016) 13–42, <https://doi.org/10.1017/jfm.2016.431>.
- [4] K. Um, X. Zhang, M. Katsoulakis, P. Plechac, D.M. Tartakovsky, Global sensitivity analysis of multiscale properties of porous materials, *J. Appl. Phys.* 123 (2018) 075103, <https://doi.org/10.1063/1.5009691>.
- [5] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Los Altos, 1988.
- [6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, revised second printing ed., Elsevier, 2014.
- [7] D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2009.
- [8] 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>.
- [9] B. Iooss, C. Prieur, Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications, <https://hal.inria.fr/hal-01556303v4>, 2018.
- [10] M. Navarro, J. Witteveen, J. Blom, Polynomial chaos expansion for general multivariate distributions with correlated variables, arXiv:1406.5483, 2014.
- [11] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, second ed., Wiley-Interscience, Hoboken, NJ, 2006.
- [12] E.S. Soofi, Capturing the intangible concept of information, *J. Am. Stat. Assoc.* 89 (1994) 1243–1254, <https://doi.org/10.2307/2290988>.
- [13] B.M. Adams, W.J. Bohnhoff, K.R. Dalbey, J.P. Eddy, M.S. Eldred, D.M. Gay, K. Haskell, P.D. Hough, L.P. Swiler, DAKOTA, a Multilevel Parallel Object-oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.0 User's Manual, Tech. Rep. SAND2010-2183, Sandia National Laboratories, Albuquerque, NM, 2009.
- [14] D. Foreman-Mackey, D.W. Hogg, DAFT: beautifully rendered probabilistic graphical models, <http://daft-pgm.org/>, 2012.
- [15] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton, NJ, 2010.
- [16] J.A. Paulson, E.A. Buehler, A. Mesbah, Arbitrary polynomial chaos for uncertainty propagation of correlated random variables in dynamic systems, *IFAC-PapersOnLine* 50 (2017) 3548–3553, <https://doi.org/10.1016/j.ifacol.2017.08.954>.
- [17] M. Rosenblatt, Remarks on a multivariate transformation, *Ann. Math. Stat.* 23 (1952) 470–472, <https://doi.org/10.1214/aoms/1177729394>.
- [18] E. Torre, S. Marelli, P. Embrechts, B. Sudret, A general framework for uncertainty quantification under non-Gaussian input dependencies, arXiv:1709.08626, 2017.
- [19] L. Baringhaus, C. Franz, On a new multivariate two-sample test, *J. Multivar. Anal.* 88 (2004) 190–206, [https://doi.org/10.1016/S0047-259X\(03\)00079-4](https://doi.org/10.1016/S0047-259X(03)00079-4).
- [20] 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.
- [21] I.M. Sobol, Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.* 1 (1993) 407–414.
- [22] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Saf.* 52 (1996) 1–17, [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6).
- [23] S. Rahman, The f -sensitivity index, *SIAM/ASA J. Uncertain. Quantificat.* 4 (2016) 130–162, <https://doi.org/10.1137/140997774>.
- [24] Y. Pantazis, M.A. Katsoulakis, D.G. Vlachos, Parametric sensitivity analysis for biochemical reaction networks based on pathway information theory, *BMC Bioinform.* 14 (2013) 1, <https://doi.org/10.1186/1471-2105-14-311>.
- [25] Y. Pantazis, M.A. Katsoulakis, A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics, *J. Chem. Phys.* 138 (2013) 054115, <https://doi.org/10.1063/1.4789612>.
- [26] M. Komorowski, M.J. Costa, D.A. Rand, M.P. Stumpf, Sensitivity, robustness, and identifiability in stochastic chemical kinetics models, *Proc. Natl. Acad. Sci. USA* 108 (2011) 8645–8650, <https://doi.org/10.1073/pnas.1015814108>.
- [27] A.J. Majda, B. Gershgorin, Improving model fidelity and sensitivity for complex systems through empirical information theory, *Proc. Natl. Acad. Sci. USA* 108 (2011) 10044–10049, <https://doi.org/10.1073/pnas.1105174108>.
- [28] A.J. Majda, B. Gershgorin, Quantifying uncertainty in climate change science through empirical information theory, *Proc. Natl. Acad. Sci. USA* 107 (2010) 14958–14963, <https://doi.org/10.1073/pnas.1007009107>.
- [29] M.-H. Chun, S.-J. Han, N.-I. Tak, An uncertainty importance measure using a distance metric for the change in a cumulative distribution function, *Reliab. Eng. Syst. Saf.* 70 (2000) 313–321, [https://doi.org/10.1016/S0951-8320\(00\)00068-5](https://doi.org/10.1016/S0951-8320(00)00068-5).

- [30] E. Borgonovo, Measuring uncertainty importance: investigation and comparison of alternative approaches, *Risk Anal.* 26 (2006) 1349–1361, <https://doi.org/10.1111/j.1539-6924.2006.00806.x>.
- [31] H. Liu, W. Chen, A. Sudjianto, Relative entropy based method for probabilistic sensitivity analysis in engineering design, *J. Mech. Des.* 128 (2006) 326–336, <https://doi.org/10.1115/1.2159025>.
- [32] E. Borgonovo, A new uncertainty importance measure, *Reliab. Eng. Syst. Saf.* 92 (2007) 771–784, <https://doi.org/10.1016/j.ress.2006.04.015>.
- [33] Q. Liu, T. Homma, A new computational method of a moment-independent uncertainty importance measure, *Reliab. Eng. Syst. Saf.* 94 (2009) 1205–1211, <https://doi.org/10.1016/j.ress.2008.10.005>.
- [34] W. Castainings, 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>.
- [35] C.K. Park, K.-I. Ahn, A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment, *Reliab. Eng. Syst. Saf.* 46 (1994) 253–261, [https://doi.org/10.1016/0951-8320\(94\)90119-8](https://doi.org/10.1016/0951-8320(94)90119-8).
- [36] 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 (2008) 223–235, <https://doi.org/10.1098/rsif.2007.1079>.
- [37] 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>.
- [38] M.J. Wainwright, M.I. Jordan, Graphical models, exponential families, and variational inference, *Found. Trends Mach. Learn.* 1 (2008), <https://doi.org/10.1561/22000000001>.
- [39] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational inference: a review for statisticians, *J. Am. Stat. Assoc.* 112 (2017) 859–877, <https://doi.org/10.1080/01621459.2017.1285773>.
- [40] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag, New York, 2006.
- [41] K.P. Burnham, D.R. Anderson, *Model Selection and Multimodel Inference, second ed.*, Springer-Verlag, New York, 2002.
- [42] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, *Phys. Rev. E* 69 (2004) 066138, <https://doi.org/10.1103/PhysRevE.69.066138>.
- [43] L. Paninski, Estimation of entropy and mutual information, *Neural Comput.* 15 (2003) 1191–1253, <https://doi.org/10.1162/089976603321780272>.
- [44] A. Antos, I. Kontoyiannis, Convergence properties of functional estimates for discrete distributions, *Random Struct. Algorithms* 19 (2001) 163–193, <https://doi.org/10.1002/rsa.10019>.
- [45] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, *J. R. Stat. Soc. B* (1966) 131–142, <https://www.jstor.org/stable/2984279>.
- [46] I. Csizsár, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hung.* 2 (1967) 299–318.
- [47] F. Liese, I. Vajda, On divergences and informations in statistics and information theory, *IEEE Trans. Inf. Theory* 52 (2006) 4394–4412, <https://doi.org/10.1109/TIT.2006.881731>.
- [48] 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 (2018) 1364–1394, <https://doi.org/10.1137/17M1143344>.