October 2019

MODEL-FORM UNCERTAINTY QUANTIFICATION FOR PREDICTIVE PROBABILISTIC GRAPHICAL MODELS

Jinchao Feng
University of Massachusetts Amherst

MODEL-FORM UNCERTAINTY QUANTIFICATION FOR PREDICTIVE PROBABILISTIC GRAPHICAL MODELS

A Dissertation Presented

by

JINCHAO FENG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2019

Department of Mathematics and Statistics

MODEL-FORM UNCERTAINTY QUANTIFICATION FOR PREDICTIVE PROBABILISTIC GRAPHICAL MODELS

A Dissertation Presented

by

JINCHAO FENG

Approved as to style and content by:

Markos A. Katsoulakis, Chair

Luc Rey-Bellet, Member

John Staudenmayer, Member

Peng Bai, Member

Nathaniel Whitaker, Department Head
Mathematics and Statistics

DEDICATION

To my family.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisors, Professor Markos A. Katsoulakis and Professor Luc Rey-Bellet, for their thoughtful, patient guidance and continuous support on both my research and life during my Ph.D. study. It was my honor and pleasure to be their student.

Besides, I would like to thank Professor John Staudenmayer and Professor Peng Bai for their service as my committee members, and insightful comments, helpful suggestions on my dissertation. I also want to express my appreciation to Professor Dionisios G. Vlachos and his research team in UDel for their collaborating in research and valuable support.

Finally, I would like to thank all the other faculties and staffs in the Department of Mathematics and Statistics at UMass Amherst. They create such a great environment for studying and working. I am also grateful to all my friends for their accompany and encouragement. I won't have such a colorful life in these years without them.

ABSTRACT

MODEL-FORM UNCERTAINTY QUANTIFICATION FOR PREDICTIVE PROBABILISTIC GRAPHICAL MODELS

SEPTEMBER 2019

JINCHAO FENG,

B.S., UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Markos Katsoulakis and Professor Luc Rey-Bellet

In this thesis, we focus on Uncertainty Quantification and Sensitivity Analysis, which can provide performance guarantees for predictive models built with both aleatoric and epistemic uncertainties, as well as data, and identify which components in a model have the most influence on predictions of our quantities of interest. In the first part (Chapter 2), we propose non-parametric methods for both local and global sensitivity analysis of chemical reaction models with correlated parameter dependencies. The developed mathematical and statistical tools are applied to a benchmark Langmuir competitive adsorption model on a close packed platinum surface, whose parameters, estimated from quantum-scale computations, are correlated and are limited in size (small data). The proposed mathematical methodology employs gradient-based methods to compute sensitivity indices. We observe

that ranking influential parameters depends critically on whether or not correlations between parameters are taken into account. The impact of uncertainty in the correlation and the necessity of the proposed non-parametric perspective are demonstrated.

In the second part (Chapter 3-4), we develop new information-based uncertainty quantification and sensitivity analysis methods for Probabilistic Graphical Models. Probabilistic graphical models are an important class of methods for probabilistic modeling and inference, probabilistic machine learning, and probabilistic artificial intelligence. Its hierarchical structure allows us to bring together in a systematic way statistical and multi-scale physical modeling, different types of data, incorporating expert knowledge, correlations, and causal relationships. However, due to multi-scale modeling, learning from sparse data, and mechanisms without full knowledge, many predictive models will necessarily have diverse sources of uncertainty at different scales. The new model-form uncertainty quantification indices we developed can handle both parametric and non-parametric probabilistic graphical models, as well as small and large model/parameter perturbations in a single, unified mathematical framework and provide an envelope of model predictions for our quantities of interest. Moreover, we propose a model-form Sensitivity Index, which allows us to rank the impact of each component of the probabilistic graphical model, and provide a systematic methodology to close the experiment - model - simulation - prediction loop and improve the computational model iteratively based on our new uncertainty quantification and sensitivity analysis methods. To illustrate our ideas, we explore a physicochemical application on the Oxygen Reduction Reaction (ORR) in Chapter 4, whose optimization was identified as a key to the performance of fuel cells.

In the last part (Chapter 5), we complete our discussion for the uncertainty quantification and sensitivity analysis methods on probabilistic graphical models by introducing a new sensitivity analysis method for the case where we know the real model sits in a certain parametric family. Note that the uncertainty indices above may be too pessimistic (as they are inherently non-parametric) when studying uncertainty/sensitivity questions for models confined within a given parametric family. Therefore, we develop a method using likelihood ratio and fisher information matrix, which can capture correlations and causal dependencies in the graphical models, and we show it can provide us more accurate results for the parametric probabilistic graphical models.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Uncertainty quantification (UQ) is a key mathematical and computational tool to assess the predictive ability of a model. This information can be used to map the envelope of model predictions, to improve a model via error reduction methods, and to inform control and optimization strategies in system tasks [45, 65, 79, 92, 95, 106, 110, 119, 128]. Sensitivity analysis (SA) is one of the most effective tools for identifying influential model parameters [2, 97, 109, 119]. The SA approaches are typically classified as local or global methods. Local sensitivity analysis (LSA) computes variability in model predictions due to infinitesimal perturbations in the model parameters [119]. The resulting local sensitivity indices (LSIs) include gradient [119] and information-based methods [28]. LSIs have also been applied to system optimization and model calibration problems [55, 137]. Global sensitivity analysis (GSA) determines variability in model predictions over a range of parameters due to uncertainty in those parameters. GSA techniques include analytical, regression, screening and variance-based methods [22, 74, 82, 86, 94, 119, 116, 125, 126, 139, 140, 141, 142].

While significant progress has been achieved in UQ and SA methods over the years, physical systems often exhibit correlated parameters. In recent work, we introduced such a mathematical framework and demonstrated the impact of correlations on a model predictive ability for a complex reaction network [125]. Our ability to understand and improve methods relies on developing simple but physically sound models that we can analyze mathematically.

In Chapter 2, we propose non-parametric methods for the GSA of chemical reaction models with correlated parameter dependencies. A Langmuir bimolecular hydrogen/oxygen competitive adsorption model is employed as a benchmark to motivate and concretely illustrate the derivation and algorithmic aspects of the proposed method. This system describes the competitive adsorption of hydrogen and oxygen on a Pt(111) surface. Such systems are encountered in catalytic oxidation, such as emissions abatement, small scale power generation, fuel cells and batteries. Here, parameter correlations stem from correlated quantum-scale computational data calculated using Density Functional Theory (DFT). These correlations are assimilated into the model as an informed prior distribution for the model's parameters. The use of non-parametric methods in modeling parameter uncertainty and understanding global sensitivity is necessitated by the limited availability of quantum-scale data. The proposed mathematical methodology employs gradient-based methods to compute correlative local/global sensitivity indices (LSI/GSI) to illustrate the relative effects of parameter perturbations, errors and uncertainties in model parameters. We show that the ranking of influential parameters depends critically on whether or not correlations between parameters are taken into account. The impact of uncertainty in the correlation on the LSI/GSI is also demonstrated. Finally, we show the necessity of the proposed non-parametric perspective by comparing with a parametric approach.

Moreover, in contrast to uncertainty due to the inherent randomness of probabilistic models and their parameters as shown in above, it is common that there is significant uncertainty regarding the probabilistic model itself. For instance, model uncertainty can stem from the fact that a model (or components of it) may have been learned from available data which could be sparse, incomplete or imperfect, as is typically the case in physical-chemical and engineering applications, so we could not determine its probability distribution or the probabilistic structure (conditional dependency/causality) of the components in the model. Similarly, for the inference/prediction tasks we typically will use approximate inference methods, which create additional model uncertainty. Lastly, some physical mechanisms may be too complex to be fully incorporated in a model and an approximation or surrogate model. In these classes of model error two challenges emerge: (a) the "real" probabilistic model is a model Q (partly unknown or computationally intractable) but instead we have to use a baseline, surrogate or approximate model P, and (b) in applications we are interested in predicting correctly Quantities of Interest (QoIs), given by expected values with respect to our models, and not necessarily the entire model Q.

On the other hand traditional UQ methods which mostly consider parametric approaches, e.g., by perturbing, tuning, or inferring the model parameters with a known probability distribution [134] which are not suitable for the aforementioned models. And most classical sensitivity indices like gradient-based (derivative-based) sensitivity indices, the Sobol index, and its variants, etc. [73], are restricted on parametric models with independent parameters. Although there are some other new sensitivity analysis methods for correlated parameters or (Gaussian) Bayesian networks, e.g., using divergence measures (especially KL divergence) to compare different model structure (especially for Gaussian Bayesian networks) [47, 48, 46], analyzing the sensitivity of components by conditioning with $f(X)$-divergence [108], extended gradient based sensitivity indices for correlated parameters [33] (introduced in Chapter 2), using the gradient-based or variance-based (ANOVA-based) sensitivity analysis for Bayesian networks with deterministic structures (known distribution) and non-deterministic structures (KDE) [138, 129, 16], and using mutual information and conditional mutual information [85, 49], they cannot handle model-form UQ, and it is not obvious how they will take advantage of the inherent graphical structure in PGMs, such as conditional independence, or restricted with Gaussian Bayesian networks. Therefore, in Chapter 3, we developed tight, information-theoretic and computable bounds for QoIs that provide such *predictive guarantees* [29, 50, 64, 56] .

To accomplish the goal, we use Probabilistic Graphical Models (PGM), an important class of methods for probabilistic modeling and inference, and constitutes the mathematical foundation of modeling uncertainty in Artificial Intelligence (AI). PGMs can bring together in a systematic way modeling, data and experiments at different scales, and expert knowledge from scientific groups. They are widely used in many real-world applications, like medical diagnostics, natural language processing, computer vision, robotics, computational biology, and cognitive science, e.g., [37], [60, 6], [81], [78], [38]. Their general mathematical formulation was developed in the seminal work (over 25K citations) of J. Pearl [100, 102], that revolutionized AI.

Many problems in machine learning involve classification, analysis and predictions, using data sets of points which are *independent* of each other. For instance, given images of handwritten characters to predict correctly the digit between 0-9. However, this is not the case in many applications involving physicochemical systems, where dependencies and correlations in space/time and between model elements (molecules, parameters, mechanisms), causal relationships between inputs and outputs, couplings between scales and physics (from quantum to meso/macro-scale) are the norm rather than the exception. Therefore, we consider using PGM to provide the proper mathematical and computational framework for physico-chemical problems, which allows us to represent expert knowledge, and learn the models from available data.

A PGM is defined as a probability model P with density

$$p(x) = \prod_{i=1}^{n} p(x_i | x_{\pi_i}) \tag{1.1}$$

where $x = (x_1, \dots, x_n)$ are the values of random variables $X = (X_1, \dots, X_n)$, $x_{\pi_i} = \{x_{i_1}, \dots, x_{i_m}\} \subset \{x_1, \dots, x_n\}$ is the values of parents of each random variable X_i, and

$$p(x_i | x_{\pi_i}) \,. \tag{1.2}$$

is the Conditional Probability Density (CPD) for the conditional distribution $P_{i|\pi_i}$ with given parents $X_{\pi_i} = x_{\pi_i}$. For example, for an inhomogeneous markov chains, we have $p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1})$, where $p(x_1|x_0) := p(x_1)$, and $\pi_i = \{i-1\}$. This concept proved to be the key to constructing complex probability models with many parameters and nodes, allowed to incorporate data and expert knowledge, and organize distributed probability computations by "divide and conquer" using graph-theoretic model representations. PGMs can also represent causal relationships between random variables through Directed Acyclical Graphs (DAG), [100], [66].

Then for a QoI $f(X)$ with a given nominal model or baseline model P, which is computationally tractable and believed to be a good approximation for the physical model of X, and an alternative or perturbed model $Q \in \mathcal{Q}$, which can be considered as the true unknown model for X and belongs to a family of distributions $\mathcal{Q}$, we want to quantify the influence of uncertainty about the model when we try to use the easier computed quantity $\mathbb{E}_P\left[f(X)\right]$ to approximate the exact value $\mathbb{E}_Q\left[f(X)\right]$

by looking at the bias

$$\mathbb{E}_Q\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right] \tag{1.3}$$

Thus, we can define the **predictive uncertainty** of baseline model P for the QoI $f(X)$ as the biases on the worst case scenario with given family of alternative models $\mathcal{Q}$,

$$I^{\pm}(f(X), P; \mathcal{Q}) := \sup_{Q \in \mathcal{Q}}/\inf \mathbb{E}_Q\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right] \tag{1.4}$$

Note that the predictive uncertainty represents the robustness of the model P w.r.t. $\mathcal{Q}$, i.e. all the biases between the predictions of $f(X)$ with $Q \in \mathcal{Q}$ and P are bounded by the predictive uncertainty and the bounds are tight.

In Chapter 3, we will investigate three ambiguity sets: one for model-form uncertainty quantification in Section 3.2.1, defined in (3.1), where we consider all the possible alternative model Q (graph structure/parents and CPDs might varied) around the PGM P with the condition $R(Q||P) \leq \eta$ for some model misspecification η; and two for model-form sensitivity analysis in Section 3.2.3, where we consider the sensitivity of node X_l in the PGM by perturbing the graph structure/parents and CPD of the node under the constraint $R(Q_{l|\pi_l^Q}||P_{l|\pi_l^P}) \leq \eta_l$ for some model misspecification η_l, defined in (3.52), or only perturbing the CPD of the node under the constraint $R(Q_{l|\pi_l}||P_{l|\pi_l}) \leq \eta_l$ and fixed the graph structure/parents π_l, defined in (3.59).

An application of the model-form UQ and SA for PGMs is shown in Chapter 4 with a chemical example on oxygen reduction reaction (ORR), which occurs at the cathode of the fuel cell and its kinetic losses comprise more than half of all voltage losses at peak power density [115]. Therefore, we use the PGMs (especially a Gaussian Bayesian network(GBN) in our case [66]) to adopt and mathematically formulate a *System of Systems* (SoS) perspective in our predictive modeling, i.e., bring together in a systematic way statistical and multi-scale physical modeling (both thermodynamics and kinetic models), different types of data (from DFT or experiments), incorporating in expert knowledge, correlations and causal relationships, then try to optimize ORR catalysts to improve fuel cell performance using the predictive model.

However, since our data is limited in size and we do not have full knowledge for all the nodes on the PGM, we must consider the model-form uncertainty in our model as we discuss above. Therefore we apply the methods we proposed in Chapter 3 and we show in Chapter 4 that the model-form sensitivity index we proposed on PGM can allow us to isolate errors in specific parts of the model, rank them and study their individual impact on predictions for our QoIs. Therefore it can give us a methodology on how to modify the model towards improving its predictive capability for specific QoIs.

In the end, we close our discussion by introducing another sensitivity analysis method for parametric PGMs in Chapter 5. The proposed UQ and SA tools above are non-parametric in nature since our challenges can involve uncertainty in the probabilistic model itself. And since the uncertainty and sensitivity indices are based on KL divergence, they are inherently non-parametric and thus the resulting family of distributions allows for densities that may not be attainable within a particular parametric family. However, if we already know the probabilistic models we need to consider lie exclusively within a fixed parametric family, our non-parametric bounds can be too wide since the family includes many other distributions outside the parametric family at hand. For instance, like many PGMs with discrete random variables, we know it must follow a Bernoulli or categorical distribution, therefore we do not need to consider the model-form uncertainty but only the uncertainty on parameters. Thus, in Chapter 5, we propose a UQ and SA method, which can work on the cases where the parametric families of the true models are known, using likelihood ratio (LR) estimator [44] and fisher information matrix (FIM). We show that our method can take advantages of the structure of PGM and reduce the computational complexity, and present its connection with the non-parametric methods we introduced in Chapter 3.

CHAPTER 2

NON-PARAMETRIC CORRELATIVE UNCERTAINTY QUANTIFICATION AND SENSITIVITY ANALYSIS

2.1 Background on Sensitivity Analysis

2.1.1 Predictive models

In this section an appropriate mathematical framework is discussed for sensitivity analysis. First, we consider an ensemble of models of the general form

$$\Pi(x|\lambda)\, p(\lambda), \tag{2.1}$$

where $\Pi(x|\lambda)$ denotes the predictive forward mathematical models, *i.e.* the probability distribution function (PDF) of state $X = x$ for fixed K dimensional model parameters $\lambda = [\lambda^1 \quad \lambda^2 \quad \cdots \quad \lambda^K]^T \in \Lambda$, and Λ presents the parameter space. The term $p(\lambda)$ denotes the PDF of λ which contains knowledge of uncertainty in the model, *i.e.* once we have $p(\lambda)$ we can generate ensembles of X's for each λ. Note that X may represent a static random variable, a snapshot of the system at some fixed time, or an entire time-series for dynamics and λ may denote the model parameters or indexing of different models.

In our specific model, λ corresponds to the binding energy of atomic oxygen and hydrogen on a given metal surface. We look at the uncertainty in coverage (Π) given the binding energy and its associated uncertainties and correlations. Coverage can also depend on other quantities that could be represented by λ, such as binding site, inert species, surface defects, surface impurities, and surface temperature [25, 32, 111]. The formalism is though general and beyond the binding energy and the isotherm. Other physical systems that follow the $\Pi(x|\lambda)p(\lambda)$ relationship include the dependency of molecular frequency (Π) on coverage (λ) [10] and forecasted temperature changes (Π), with CO_2, methane, and other greenhouse gases (λ) [84, 120, 127].

The system observable can be defined over all possible realizations of the state

$$f(\lambda) = \int h(x)\, \Pi(x|\lambda)\, dx, \tag{2.2}$$

where $h(x)$ denotes a desired quantity. The correlations in the parameter vector λ are also included in $p(\lambda)$ and are propagated into the state X through the predictive forward model of $\Pi(x|\lambda)$. Finally, the averaged observable for the model can be defined by

$$\bar{f} = \int f(\lambda)\, p(\lambda)\, d\lambda = \int\int h(x)\, \Pi(x|\lambda)\, p(\lambda)\, dx\, d\lambda. \tag{2.3}$$

2.1.2 Derivative-based sensitivity indices

Consider a general class of nonlinear models of the form

$$f = f(\lambda), \tag{2.4}$$

where f is an arbitrary scalar function. The (relative) LSI of f with respect to λ of (2.4) at the nominal value of λ_* is

$$S_\lambda^f(\lambda_*) = \frac{\nabla_\lambda f(\lambda)}{f(\lambda)}\bigg|_{\lambda=\lambda_*} = \nabla_\lambda \ln f(\lambda)\bigg|_{\lambda=\lambda_*}, \tag{2.5}$$

where

$$\nabla_\lambda f(\lambda) = \left[\frac{\partial f(\lambda)}{\partial \lambda^1}\ \ \frac{\partial f(\lambda)}{\partial \lambda^2}\ \ \cdots\ \ \frac{\partial f(\lambda)}{\partial \lambda^K}\right]^T. \tag{2.6}$$

The LSI of (2.5) supplies useful sensitivity information in the case of almost certain parameters, *i.e.* for a relatively tight range of parameter values. To incorporate the knowledge of uncertain parameter distributions and provide sensitivity information over the entire range, we determine the relevant GSI by employing the partial derivative of the LSI as a basic building block to integrate the local sensitivities over the total range of parameter changes

$$\xi_\lambda^f = \int_\lambda |\nabla_\lambda \ln f(\lambda)|^q p(\lambda)\, d\lambda, \tag{2.7}$$

where q denotes the type of required index ($q = 1$: improved Morris index, $q = 2$: asymptotic limit of the standard Morris index).

Note that the PDF, which incorporates knowledge of the λ distribution in the GSI of (2.7), must be identified subject to available experimental and/or simulation-based data. The possible correlations between the system parameters, which may be discovered during regression of the data in statistical models, must be encoded in $p(\lambda)$. Such correlations play a deciding role in sensitivity analysis and their effects are quantified in the following sections.

2.2 Parameter Correlation Effects

Previously, LSA and GSA were treated for the case of independent parameters. To extend sensitivity analysis to models with correlated parameters, we partition the vector of parameters into two,

$$\begin{aligned} \lambda_1 &= [\lambda^1 \;\; \lambda^2 \;\; \cdots \;\; \lambda^m]^T \in \Lambda_1, \\ \lambda_2 &= [\lambda^{m+1} \;\; \lambda^{m+2} \;\; \cdots \;\; \lambda^K]^T \in \Lambda_2, \end{aligned} \tag{2.8}$$

where $\lambda = [\lambda_1 \;\; \lambda_2]^T$, $\Lambda = \Lambda_1 \oplus \Lambda_2$, λ_1 contains all independent parameters, and λ_2 contains all dependent parameters. Parameters can be classified through their correlations, which are identified using experiments and/or computational tools for specific case studies, and by applying causality statistical methods. When λ_1 and λ_2 are correlated, perturbations in one parameter affect the other. Proper mathematical tools are needed to quantify parameter correlations and their impact on model reliability.

The correlation between λ_1 and λ_2 can be described by their joint probability distribution,

$$p(\lambda_1, \lambda_2) = p(\lambda_2|\lambda_1)p(\lambda_1). \tag{2.9}$$

For the marginal distribution of λ_1,

$$p(\lambda_1) = \int_{\Lambda_2} p(\lambda_1, \lambda_2)\, d\lambda_2. \tag{2.10}$$

Identifying $p(\lambda_1)$ and $p(\lambda_2|\lambda_1)$ in a systematic way is an essential step in our CGSA.

The joint probability distribution of $p(\lambda_1, \lambda_2)$ can either be built directly, with a sufficiently large ensemble of experimental and/or simulation-based data [125], or computed sequentially by marginalization according to (2.9). The latter approach requires building PDFs with data for both $p(\lambda_2|\lambda_1)$ and $p(\lambda_1)$, followed by Monte Carlo sampling to calculate correlative indices. There are various linear regression (LR) methods that can identify the conditional PDF of $p(\lambda_2|\lambda_1)$; including deterministic (DLR), stochastic (SLR), and Bayesian (BLR) [107, 130]. DLR yields a deterministic linear model whose LCSIs can be computed exactly, while its GSIs depend on the choice of $p(\lambda_1)$. SLR uses a least squares model for $p(\lambda_2|\lambda_1)$, typically in a Gaussian form [130]. Usually, there is not enough data for adequate fitting of a least squares model. BLR can can bypass this shortcoming by putting a prior on the parameters in the linear fit [107].

Hierarchical or empirical Bayesian methods can identify the marginal PDF of $p(\lambda_1)$ via deterministic linear or stochastic nonlinear fitting to the data. Bootstrapping does not require fitting but instead relies on simulation [130], which is particularly appropriate for problems with little data, by creating synthetic new samples using random re-sampling according to the actual distribution of the data [130]. In this way we can enrich the histogram with more data points and obtain a Gaussian distribution [130]. If histograms are too sparse, smoothed bootstrapping can be used. This method applies a kernel to data from a standard histogram. A Bayesian approach can be used to fit data to a well known distribution that "looks like" the histogram. "Looks like" means that we pick a family of well known distributions and match the first few moments with the corresponding moments of the data's histogram, *i.e.* mean, variance, skewness, etc. [40, 130].

We perform the correlative sensitivity analysis by focusing on λ_1, while still accounting for the correlations with λ_2

$$F(\lambda_1) = \int_{\Lambda_2} f(\lambda_1, \lambda_2)\, p(\lambda_2|\lambda_1)\, d\lambda_2. \tag{2.11}$$

The correlative local sensitivity index (CLSI) at the nominal point λ_{1*} is obtained similarly to (2.5) by direct differentiation,

$$S^f_{\lambda_1,\,corr}(\lambda_{1*}) = \frac{\nabla_{\lambda_1} F(\lambda_1)}{F(\lambda_1)}\Big|_{\lambda=\lambda_{1*}} = \nabla_{\lambda_1} \ln F(\lambda_1)\Big|_{\lambda=\lambda_{1*}}. \tag{2.12}$$

The CGSI can then be formulated

$$\xi^f_{\lambda_1,\,corr} = \int_{\Lambda_1} |\nabla_{\lambda_1} \ln F(\lambda_1)|^q p(\lambda_1)\, d\lambda_1, \tag{2.13}$$

by employing the CLSI of (2.12) as building blocks where $q = 1$ or $q = 2$.

For deterministic correlation where $\lambda_2 = g(\lambda_1)$, we can simplify the λ_1-marginal PDF of (2.10) by considering $p(\lambda_2|\lambda_1) = \delta(g(\lambda_1) - \lambda_2)$,

$$p(\lambda_1, \lambda_2) = p(\lambda_1)\, \delta(g(\lambda_1) - \lambda_2), \tag{2.14}$$

where $\delta(\cdot)$ denotes the standard Dirac function. Therefore, from (2.11), we have

$$\begin{aligned} F(\lambda_1) &= \int_{\Lambda_2} f(\lambda_1, \lambda_2)\, p(\lambda_2|\lambda_1)\, d\lambda_2 \\ &= \int_{\Lambda_2} f(\lambda_1, \lambda_2)\, \delta(g(\lambda_1) - \lambda_2)\, d\lambda_2 \\ &= f(\lambda_1, g(\lambda_1)), \end{aligned} \tag{2.15}$$

and the CLSI can be simplified to the following form

$$\begin{aligned} S^f_{\lambda_1,\, corr}(\lambda_{1*}) &= \nabla_{\lambda_1} \ln f(\lambda_1, g(\lambda_1)) \\ &= \Big(\frac{\nabla_{\lambda_1} f(\lambda_1, g(\lambda_1))}{f(\lambda_1, g(\lambda_1))}\Big)\Big|_{\lambda_1=\lambda_{1*}} \\ &= \left(\frac{1}{f(\lambda_1,\lambda_2)}\frac{\partial f(\lambda_1,\lambda_2)}{\partial \lambda_1}\right. \\ &\quad \left. + \frac{1}{f(\lambda_1,\lambda_2)}\frac{\partial f(\lambda_1,\lambda_2)}{\partial \lambda_2}\frac{\partial g(\lambda_1)}{\partial \lambda_1}\right)\Bigg|_{\lambda_1=\lambda_{1*},\, \lambda_2=g(\lambda_{1*})}. \end{aligned} \tag{2.16}$$

The additional second term in the CLSI differs from uncorrelative LSI in that the derivative with respect to the parameter λ_2 comes directly into play. The CGSI formulation for such a simplified case is

$$\begin{aligned} \xi^f_{\lambda_1,\, corr} &= \int_{\Lambda_1} \Big|\nabla_{\lambda_1} \ln \int_{\Lambda_2} f(\lambda_1,\lambda_2)\, p(\lambda_2|\lambda_1)\, d\lambda_2\Big|^q p(\lambda_1)\, d\lambda_1 \\ &= \int_{\Lambda_1} \Big|\nabla_{\lambda_1} \ln f(\lambda_1, g(\lambda_1))\Big|^q p(\lambda_1)\, d\lambda_1. \end{aligned} \tag{2.17}$$

The implementation of the sampling algorithm used to compute the correlative local/global sensitivity index (CLSI/CGSI) is described in the Appendix A.5.

2.3 A Langmuir Bimolecular Adsorption Model

We consider a Langmuir bimolecular adsorption model which describes competitive dissociative adsorption of hydrogen (H_2) and oxygen (O_2) on a catalyst surface,

$$\begin{aligned} H_2 + 2^* &\rightleftharpoons 2H^*, \\ O_2 + 2^* &\rightleftharpoons 2O^*, \end{aligned} \tag{2.18}$$

where H_2 and O_2 denote the hydrogen and oxygen molecules in the gas phase, 2^* are two active sites on the metal surface, and H^* and O^* represent the adsorbed hydrogen and oxygen atoms on the surface, respectively. A schematic of this adsorption process is illustrated in Figure 1. The physical system is related to hydrogen oxidation in fuel cells and batteries [7, 54, 80, 83, 93].

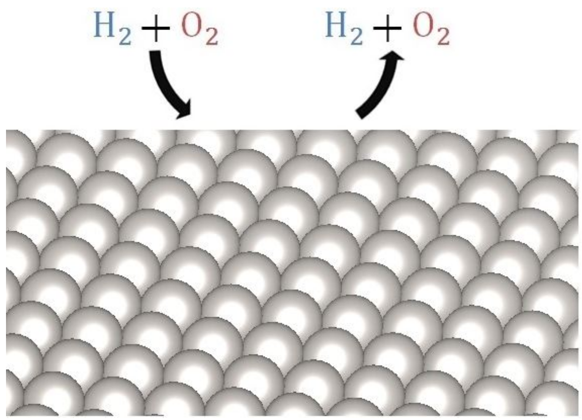

Figure 1. Competitive dissociative adsorption of hydrogen and oxygen on a catalyst surface.

The coverages dynamics can be formulated by the following set of ordinary differential equations

$$\begin{aligned} \frac{d\hat{\theta}_{H^*}}{dt} &= k_{H_2}^{ads} P_{H_2}(1-\hat{\theta}_{H^*}-\hat{\theta}_{O^*})^2 - k_{H_2}^{des}\hat{\theta}_{H^*}^2, \quad \theta_{H^*}^0 = \hat{\theta}_{H^*}(0), \\ \frac{d\hat{\theta}_{O^*}}{dt} &= k_{O_2}^{ads} P_{O_2}(1-\hat{\theta}_{H^*}-\hat{\theta}_{O^*})^2 - k_{O_2}^{des}\hat{\theta}_{O^*}^2, \quad \theta_{O^*}^0 = \hat{\theta}_{O^*}(0), \end{aligned} \tag{2.19}$$

where $\theta_{H^*}^0$ and $\theta_{O^*}^0$ represent the initial hydrogen and oxygen coverages, respectively. P_{H_2} and P_{O_2} are the partial pressures of the gas phase species [18], and we set $P_{H_2} = 1.01325 \times 10^{-10}\ N/m^2$, $P_{O_2} = 1.01325 \times 10^{-50}\ N/m^2$ in this chapter.

The hydrogen and oxygen coverages at equilibrium are

$$\begin{aligned} \hat{\theta}_{H^*} &= \frac{(K_{H_2}P_{H_2})^{\frac{1}{2}}}{1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}}, \\ \hat{\theta}_{O^*} &= \frac{(K_{O_2}P_{O_2})^{\frac{1}{2}}}{1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}}, \end{aligned} \tag{2.20}$$

where P is partial pressure and $K = \dfrac{k^{ads}}{k^{des}}$ is the equilibrium constant [24, 87]. K is determined from DFT calculations.

By focusing on variations of binding energies and fixing the other parameters, the coverages are

$$\hat{\theta}_{H^*} = \hat{\theta}_{H^*}(\Delta E_H, \Delta E_O), \qquad \hat{\theta}_{O^*} = \hat{\theta}_{O^*}(\Delta E_H, \Delta E_O). \tag{2.21}$$

where ΔE_H and ΔE_O are the binding energies of atomic hydrogen and oxygen to the surface. It is the effect that uncertainty and correlations in the binding energies have on the coverages that we explore below. The detailed formulas of $\hat{\theta}_{H^*}$ and $\hat{\theta}_{O^*}$ is derived in the Appendix A.1.

The Langmuir adsorption isotherm is strictly valid at low coverage with adsorption at a single site, which is our system of interest. For dissociative adsorption, which we study here, the Langmuir model requires adjacent empty sites on the catalyst surface. An ab-initio molecular dynamics (AIMD) study of hydrogen on Pd(100) showed that regardless of coverage, only two adjacent catalyst sites are necessary to dissociate hydrogen[53]. Although applications of AIMD to heterogeneous catalysis are rapidly advancing, the computational cost is still prohibitive for it to be used in generating adsorption isotherms[52]. Less computational intensive methods, such as Monte Carlo[118] and molecular dynamics with force fields[14], are used instead to generate an isotherm.

Seller et al. have shown that, when combined with the Bragg-Williams coverage model, the Langmuir adsorption isotherm accurately recreates experimental isotherms for several systems[118]. Furthermore, the same study found that the Langmuir adsorption isotherm with mean field treatment compares favorably with coverages predicted from lattice based grand canonical Monte Carlo (GCMC) simulations under certain conditions. A force field based molecular dynamics simulation of dimethyl methylphosphonate (DMMP) also supports the validity of the Langmuir model[14].

2.4 Data and Correlations

2.4.1 Methods

Electronic contributions to adsorption enthalpies are calculated with DFT, using the Vienna ab initio Simulation Package (VASP), version 5.3.2 [70, 71, 68, 69], with the plane wave basis set, PAW pseudopotentials[11, 72], and periodic boundary conditions. Simulation parameters are similar to those used in our previous work [90]. All VASP input files are created using the Atomistic Simulation Environment, an open-source python-based software program [8]. We employ the PBE exchange-correlation functional with D3 dispersion corrections [104, 51]. Spin-polarized calculations are performed for molecules in a vacuum and systems containing Ni and Co. The first Brillouin zone is sampled using the Monkhurst-Pack (3x3x1) mesh [91]. For the purposes of this work, the level of accuracy achieved using this mesh size was sufficient. Ionic force cut-off for all calculations is set to 0.05 eV/. In slab calculations, we use the (4x4) supercell containing four layers of atoms, with the positions of the bottom two layers fixed. We use an adsorbate coverage of 1/16 monolayers in all calculations.

The DFT dissociative adsorption energy for molecular oxygen on a metal surface is defined in Equation 2.22. A similar relation holds for dissociative adsorption of molecular hydrogen.

$$\Delta E_{O_2 \rightarrow 2O^*} = -\Big(2E_{O^*} - (E_{O_2} + 2E_*)\Big) \tag{2.22}$$

such that

$$\Delta E_O = \frac{1}{2}(\Delta E_{O_2 \rightarrow 2O^*} + D_0). \tag{2.23}$$

In Equation 2.22, E_{O_2} is the DFT energy of an O_2 molecule in a vacuum, E_* is the DFT energy of the pristine metal slab, E_{O^*} is the energy of the adsorbate-metal system, D_0 is the gas-phase bond dissociation energy at 0 K, and E_O is DFT adsorption energy of an oxygen atom. The calculated energies for a variety of metal surfaces and the resulting scaling relationship between electronic contributions of H and O dissociative adsorption energies are shown in Figure 2. We obtain a linear correlation between hydrogen and oxygen adsorption energies with a R^2 value of 0.87, i.e.

$$\Delta E_O = a\Delta E_H + b \tag{2.24}$$

where $a = 2.51$, $b = -2.46$ (eV).

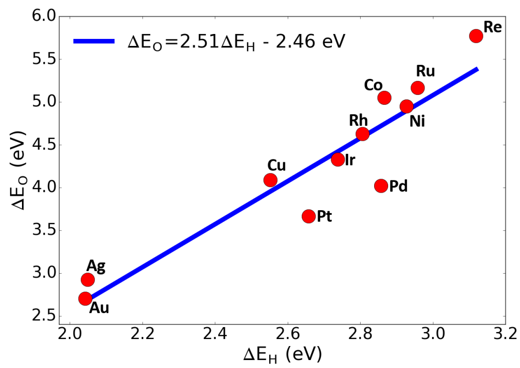

Figure 2. Correlation between oxygen and hydrogen adsorption energies on close-packed metal surfaces as defined in (2.24). Adsorbed atomic species are assumed to occupy fcc hollow sites.

Vibrational contributions and other temperature effects to adsorption enthalpy and entropy are accounted for in calculations of adsorption equilibrium constants (see Appendix A.1 for details). Zero point energy (ZPE) corrections for gas phase H_2 and O_2 are calculated using their experimental vibrational frequencies [34, 132].

Kinetic energy loss upon adsorption is accounted for by using the ideal gas value of $\frac{3}{2}RT$. Harmonic and rigid rotor approximations were utilized to account for vibrational and rotational degrees of freedom, respectively [88]. Hessian matrices are constructed using 0.015 displacements in x, y, and z directions from adsorbate equilibrium positions.

2.4.2 First principles adsorption data and errors

In order to validate our computational setup and provide an error estimate, we compare the calculated and experimental adsorption enthalpies of oxygen and hydrogen on platinum in Table 1 [63, 17, 23, 57]. The DFT calculations reproduce the experimental data well.

Table 1. Experimental and DFT-calculated enthalpies of adsorption for atomic oxygen and hydrogen on Pt(111).

Adsorbate	Experimental enthalpy	DFT computed enthalpy ‡
O	$3.71 \pm 0.07^{*} eV$	3.68 eV
H	$2.63^{\dagger}$ eV	2.69 eV

2.4.3 Correlations and prediction

Figure 3 highlights the differences in our model, defined in (2.1) for $(\hat{\theta}_{H^*}, \hat{\theta}_{O^*})$, resulting from correlations between ΔE_H and ΔE_O. Consider

$$\begin{aligned} & p(\hat{\theta}_{H^*}, \hat{\theta}_{O^*}) \\ = & \int \Pi(\hat{\theta}_{H^*}, \hat{\theta}_{O^*} | \Delta E_H, \Delta E_O)\, p(\Delta E_H, \Delta E_O)\ d(\Delta E_H, \Delta E_O), \end{aligned} \tag{2.25}$$

where $\Pi(\hat{\theta}_{H^*}, \hat{\theta}_{O^*} | \Delta E_H, \Delta E_O) = \delta(\hat{\theta}_{H^*}(\Delta E_H, \Delta E_O), \hat{\theta}_{O^*}(\Delta E_H, \Delta E_O))$, δ is the standard Dirac function and both $\hat{\theta}_{H^*}(\Delta E_H, \Delta E_O)$ and $\hat{\theta}_{O^*}(\Delta E_H, \Delta E_O)$ are given by (2.21).

Then, for the uncorrelated case (subscript uc), we can assume that

$$p(\Delta E_H, \Delta E_O) = p_{uc}(\Delta E_H, \Delta E_O) = p(\Delta E_H)p(\Delta E_O), \tag{2.26}$$

where $p(\Delta E_H)$ and $p(\Delta E_O)$ are defined as density functions of Gamma distributions, given by (2.30) and (2.31) in next section.

In the correlated case (subscript c), we assume that

$$p(\Delta E_H, \Delta E_O) = p_c(\Delta E_H, \Delta E_O) = p(\Delta E_H)p(\Delta E_O|\Delta E_H), \tag{2.27}$$

where $p(\Delta E_H)$ is still given by a Gamma distribution but $p(\Delta E_O|\Delta E_H)$ comes from a normal distribution with mean $a\Delta E_H + b$ and variance determined by the data in Section 2.6.2, which gives $p_c(\Delta E_H, \Delta E_O)$ a lower variance than $p_{uc}(\Delta E_H, \Delta E_O)$.

We can use changing of variables such that

$$p(\hat{\theta}_{H^*}, \hat{\theta}_{O^*}) = p(\Delta E_H, \Delta E_O)|\det(J)|, \tag{2.28}$$

where J is the Jacobian of the inverse of coverage function $\hat{\theta}(\Delta E_H, \Delta E_O)$ from (2.8), evaluated at $(\hat{\theta}_{H^*}, \hat{\theta}_{O^*})$. Figure 3 shows the density function contours $p(\hat{\theta}_{H^*}, \hat{\theta}_{O^*})$, in the uncorrelated case using (2.26), and correlated case using (2.27). Note that the correlation of ΔE_H and ΔE_O reduces the variance of our model.

2.5 Uncorrelated Sensitivity Index

In this section, we compute the uncorrelated local and global sensitivity index defined in Section 2.1.2 for the coverages $\hat{\theta}_{H^*}$ and $\hat{\theta}_{O^*}$ with respect to ΔE_H, and will turn to the correlated cases in the next three sections. In section 2.5.1, we compute the LSIs according to (2.21); and in section 2.5.2, we construct an uncorrelated prior distribution for ΔE_H and ΔE_O, and then use this distribution and LSIs to compute GSIs.

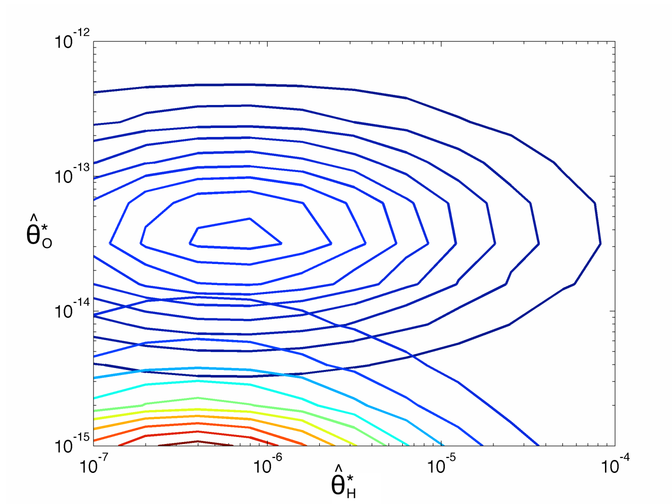

Figure 3. **Contour plot of $p(\hat{\theta}_{H^*}, \hat{\theta}_{O^*})$ in log-scale where warmer colors represents higher densities. The upper contour plot with cooler colors corresponds to the uncorrelated case, which suggests that the density function in this case is flatter; the lower contour plot with warmer colors corresponds to the correlated case which suggests the density function has a higher mode located close to the bottom and the left of the figure. The model with correlations has significantly lower variance than the uncorrelated one, yielding an overall more predictive model.**

2.5.1 Uncorrelated local sensitivity index (LSI)

Using the binding energies of adsorbed hydrogen and oxygen from Figure 2, we can analyze the relative LSIs for $\hat{\theta}_{H^*}$ and $\hat{\theta}_{O^*}$ with respect to ΔE_H. S_H^H and S_H^O are identified using (2.5) and the model given in (2.21) (detailed calculations are presented in Appendix A.2). For Pt, $\Delta E_H = 2.6581(eV)$, $\Delta E_O = 3.6604(eV)$,and the sensitivity of H and O coverages with respect to the H binding energy are $S_H^H = 38.9080$ and $S_H^O = -0.0138$. As expected, the H binding energy has a large effect on its coverage and a slight effect on the O coverage (some coupling is expected due to the competitive nature of adsorption).

2.5.2 Uncorrelated global sensitivity index (GSI)

To compute the corresponding GSIs on Pt by (2.7), we need to construct the distribution of our parameters, $p(\Delta E_H, \Delta E_O)$. In the uncorrelated case, we have

$$p(\Delta E_H, \Delta E_O) = p(\Delta E_H)p(\Delta E_O|\Delta E_H), \tag{2.29}$$

where $p(\Delta E_H)$ is the prior information for ΔE_H on Pt, and $p(\Delta E_O|\Delta E_H) = p(\Delta E_O)$ since ΔE_H and ΔE_O are independent, assuming no correlation.
Using the experimental and DFT data shown in Table 1, we construct an informative prior for $\Delta E_H \in \mathbb{E}_H$. Let $x_H = 2.63$ (eV), $x_O = 3.71$ (eV), $y_H = 2.69$ (eV) and $y_O = 3.68$ (eV), where x_i are the values given by experiment and y_i are given by DFT, $i = H, O$. To quantify uncertainty from DFT error, we assume that ΔE_H on Pt follows a gamma distribution with mean x_H and the standard deviation given by the difference between experiment and DFT, $(x_i - y_i)$. We can construct the distribution for $\Delta E_O \in \mathbb{E}_O$ in the same way under the uncorrelated assumption. The explicit density functions are shown below,

$$\begin{aligned} &\mathbb{E}_H \sim Gamma(a_H, b_H), \\ &p(\Delta E_H) = \frac{1}{b_H^{a_H}\Gamma(a_H)} \Delta E_H^{a_H - 1} \exp(-\frac{\Delta E_H}{b_H}) \quad \text{for } x > 0, \end{aligned} \tag{2.30}$$

$$p(\Delta E_O) = \frac{1}{b_O^{a_O}\Gamma(a_O)} \Delta E_O^{a_O - 1} \exp(-\frac{\Delta E_O}{b_O}) \quad \text{for } x > 0, \tag{2.31}$$

where $a_i = x_i^2/(x_i - y_i)^2$ and $b_i = (x_i - y_i)^2/x_i$, $i = H, O$.
The GSIs, ξ_H^H and ξ_H^O, are formulated by (2.7) with $q = 2$ and

$$p(\lambda) = p(\Delta E_H)p(\Delta E_O).$$

Computing the integral in (2.7) numerically gives $\xi_H^H = 1509.8865$ and $\xi_H^O = 0.2194$. Again, the H binding energy has a major effect only on the H coverage and a slight effect on the O coverage.

2.6 Correlated Local Sensitivity Index (CLSI)

The following sections cover the CLSI. In section 2.6.1, we consider the simplest correlation model, both deterministic and linear, to compute the CLSIs defined in Section 2.2. In section 2.6.2, we construct parametric models for $p(\Delta E_O|\Delta E_H)$ using the data shown in Figure 2, and compute the corresponding CLSIs. Section 2.7 covers non-parametric models.

2.6.1 CLSI with linear, deterministic correlations

To calculate the CLSI for $\hat{\theta}_{H^*}$ and $\hat{\theta}_{O^*}$ with respect to ΔE_H, with the formula defined in (2.12), we use the conditional probability $p(\Delta E_O|\Delta E_H)$. In the deterministic case, we can assume the conditional distribution of $\mathbb{E}_O$ has a mean of $g(\Delta E_H)$ and zero variance whose PDF can be described by the standard Dirac function

$$p(\Delta E_O|\Delta E_H) = \delta(g(\Delta E_H) - \Delta E_O), \tag{2.32}$$

as presented in Section 2.2. Then, using the data from Figure 2, one can determine the function $g(\Delta E_H)$ with different fitting models, like polynomials and smoothing splines. In this chapter, we use the linear function and set $g(\Delta E_H) = a\,\Delta E_H + b$, as shown in Figure 2. Then, the conditional distribution can be written as

$$p(\Delta E_O|\Delta E_H) = \delta(a\Delta E_H + b - \Delta E_O). \tag{2.33}$$

For brevity, we only consider the CLSI formulation, as defined in (2.12) for the hydrogen coverage $\hat{\theta}_{H^*}$, with respect to adsorbed hydrogen binding energy on the surface. The rest of CLSIs can be formulated by following the same procedure.

The CLSI for $\hat{\theta}_{H^*}$ with respect to ΔE_H at the nominal hydrogen binding energy of $\widehat{\Delta E_H}$ according to (2.12) takes the following form

$$\begin{aligned} S^H_{H,\,corr} &= \left[\frac{\partial(\ln\hat{\theta}_{H^*})}{\partial(\Delta E_H)}\right]_{corr} \\ = \quad & \frac{2}{\hat{\theta}_{H^*}}\left[\frac{\partial\hat{\theta}_{H^*}}{\partial K_{H_2}}\frac{\partial K_{H_2}}{\partial(\Delta G_{H_2\to 2H^*})} + \right. \\ & \left.\left. a\frac{\partial\hat{\theta}_{H^*}}{\partial K_{O_2}}\frac{\partial K_{O_2}}{\partial(\Delta G_{O_2\to 2O^*})}\right]\right|_{\Delta E_H=\widehat{\Delta E_H},\,\Delta E_O=a\,\widehat{\Delta E_H}+b}, \end{aligned} \tag{2.34}$$

where

$$\begin{aligned} \frac{\partial\hat{\theta}_{H^*}}{\partial K_{H_2}} &= \frac{P_{H_2}\Big(1+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)}{2(K_{H_2}P_{H_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}, \\ \frac{\partial\hat{\theta}_{H^*}}{\partial K_{O_2}} &= -\frac{P_{O_2}(K_{H_2}P_{H_2})^{\frac{1}{2}}}{2(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}, \\ \frac{\partial K_{H_2}}{\partial(\Delta G_{H_2\to 2H^*})} &= -\frac{1}{RT}\exp\Big(-\frac{\Delta G_{H_2\to 2H^*}}{RT}\Big), \\ \frac{\partial K_{O_2}}{\partial(\Delta G_{O_2\to 2O^*})} &= -\frac{1}{RT}\exp\Big(-\frac{\Delta G_{O_2\to 2O^*}}{RT}\Big). \end{aligned} \tag{2.35}$$

The CLSI derivations in the presence of deterministic linear correlation are briefly described in Appendix A.2 and the results of $S^H_{H,\,corr}$ and $S^O_{H,\,corr}$ are shown in FIG. 1. The numerical results for Pt are $S^H_{H,\,corr} = 38.9021$ and $S^O_{H,\,corr} = 97.7442$. The corresponding uncorrelated LSI indices from (2.5) are $S^H_H = 38.9080$, and $S^O_H = -0.0138$; hence the correlation between ΔE_H and ΔE_O does not affect the sensitivity of $\hat{\theta}_{H^*}$ with respect to ΔE_H, but does impact the sensitivity of $\hat{\theta}_{O^*}$ with respect to ΔE_H. The LSI changes from slightly negative to highly positive. This is rationalized from the slope of the correlation depicted in Figure 2. Specifically, an increase in the binding energy of H leads a much higher increase in the binding energy of O and thus to an increase of the O coverage.

2.6.2 CLSI with stochastic correlations: parametric probabilistic models

The above is a perfect linear model (deterministic) and ignores the variation around the linear fit of the binding energies. To capture correlations from the linear model, we set up a linear probabilistic model for ΔE_H and ΔE_O by introducing a random variable, ω, in the correlation [62],

$$\Delta E_O = a\,\Delta E_H + b + \omega, \qquad \omega \in \Omega. \tag{2.36}$$

To determine the distribution of ω, we can fit the data or adjusted data (to match the required domain of some distribution) using parametric models, like normal or gamma. Here we choose the normal distribution and fit the parameters using MATLAB by the Maximum Likelihood Estimation (MLE) method [62]. The result is shown in Figure 4.

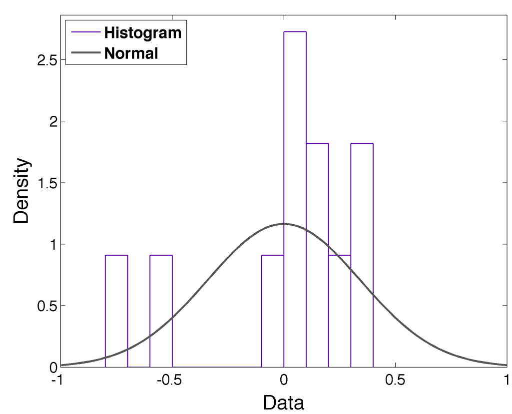

Figure 4. Model fitting for the random variable ω in (2.36) using a normal distribution; here we compare the best fit to the data's histogram. The normal distribution is not a good approximation for the data since it does not properly capture the outlier values between -1 and -0.5, depicted in the histogram. Other parametric models give similar results.

Using (2.12) to compute the CLSI of $\hat{\theta}_{H^*}$ with respect to ΔE_H, we consider

$$\begin{aligned} F_H^H(\Delta E_H) &= \int_{\Delta E_O} \hat{\theta}_{H^*}(\Delta E_H, \Delta E_O)\, p(\Delta E_O|\Delta E_H)\, d\,\Delta E_O \\ &= \int_{\omega} \hat{\theta}_{H^*}\big(\Delta E_H, a\,\Delta E_H + b + \omega\big)\, p(\omega)\, d\,\omega, \end{aligned} \tag{2.37}$$

where $p(\omega)$ is the PDF for ω in (2.36). Instead of using the Monte Carlo method (discussed in Appendix A.5), we can also use numerical integration to approximate the integral in (2.37).

The CLSI at the nominal point ΔE_H* can then be obtained by direct differentiation

$$\begin{aligned} S_{H,\,corr}^{H}(\Delta E_H*) &= \frac{\big(F_H^H(\Delta E_H)\big)'}{F_H^H(\Delta E_H)}\Big|_{\Delta E_H=\Delta E_H*} \\ &= \big(\ln F_H^H(\Delta E_H)\big)'\Big|_{\Delta E_H=\Delta E_H*} \end{aligned} \tag{2.38}$$

The gradient of $\ln F_H^H(\Delta E_H)$ is commonly estimated, such that

$$\big(\ln F_H^H(\Delta E_H)\big)' \approx \frac{\ln F_H^H(\Delta E_H+\epsilon) - \ln F_H^H(\Delta E_H-\epsilon)}{2\epsilon}. \tag{2.39}$$

The CLSI for $\hat{\theta}_{O^*}$, $S_{H,\,corr}^{O}(\Delta E_H*)$, is computed similarly. The numerical results for Pt are $S_{H,\,corr}^{H} = 35.9874$ and $S_{H,\,corr}^{O} = 9.4965$. Compared to the deterministic model results in the previous subsection, we find that uncertainty significantly impacts $S_{H,\,corr}^{O}$. This is a rather interesting result because the correlation in the data (linear) results in the H binding energy having a significant effect on the O coverage but uncertainty significantly diminishes this effect. We give results from other parametric models in the Appendix A.3.

2.7 Correlative Local Sensitivity Index (CLSI) with Stochastic Correlations: Non-parametric Models

For small data sets, such as ours, parametric models are not usually adequate. Instead, we consider non-parametric methods [131]. A possible non-parametric method is the empirical distribution function,

$$\hat{P}(\omega) = \frac{1}{11}\sum_{i=1}^{11} I(X_i \leq \omega), \tag{2.40}$$

where I is the identity function. With this method, $\mathbb{E}_{\hat{P}}[f]$ for some function f can be approximated via bootstrapping [131].

For categorical distributions, the bootstrap distribution is close to the posterior distribution with a non-informative symmetric Dirichlet prior according to Bayes method. It also has the same support, mean, and nearly the same covariance matrix as the data in the histogram. The bootstrap distribution is obtained without specifying either the prior or sampling from the posterior distribution [36].

We can also use curve estimation for our model [131]. A simple density estimator is a histogram, which is a piece-wise constant function where the height of the function is proportional to number of observations in each bin

$$\hat{p}_n(\omega) = \sum_{i=1}^{n} \frac{\nu_i}{nh} I(\omega \in B_i), \tag{2.41}$$

where $B_1, \ldots, B_n$ are the histogram bins, $h = 1/n$ is the bin-width, and ν_i is the number of observations in B_i, as shown in Figure 4.

Smoother estimators, called kernel density estimators [131], converge faster to the true density than fitting from histograms because histograms are discontinuous

$$\hat{p}_n(\omega) = \frac{1}{11}\sum_{i=1}^{11} \frac{1}{h} K(\frac{\omega - X_i}{h}), \tag{2.42}$$

where $h > 0$ is the bandwidth and K is the kernel, defined to be any smooth function satisfying $K(x) \geq 0$, $\int K(x)dx = 1$, $\int xK(x)dx = 0$ and $\sigma_K^2 = \int x^2 K(x)dx > 0$.

In the main text of this work we use the histogram to approximate the distribution of ω, and use

$$F_H^H(\Delta E_H) = \int_\omega \hat{\theta}_{H^*}(\Delta E_H, a\,\Delta E_H + b + \omega)\,\hat{p}_n(\omega)\,d\,\omega, \tag{2.43}$$

to compute $S^H_{H,\,corr}(\Delta E_H*)$ and $S^O_{H,\,corr}(\Delta E_H*)$ using Equation 2.38. The numerical results for adsorption on Pt are $S^H_{H,\,corr} = 35.2196$ and $S^O_{H,\,corr} = 12.4873$. Results from the kernel density estimators with uniform and standard normal kernel, $\mathcal{N}(0,1)$, are given in the Appendix A.4.

Figure 5 summarizes all local sensitivity analysis results (the mangitude is plotted so a semi-log scale can be used). Correlations play a significant role as demontsrated in our earlier work [125]. Clearly, the uncertainty in the correlations must properly by accounted for and, given the limited number of data we have for physical models, non-parametric models of the uncertainty are essential. For a large sample size, both (parametric and non-parametric) models should converge to the real distribution [130, 131]. Because we only have a few samples here, the non-parametric models approximate the noise term better, as shown in Figure 4.

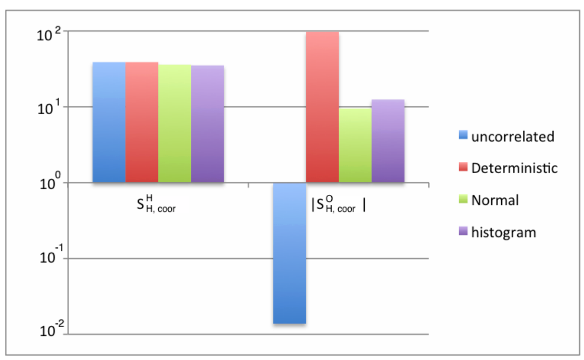

Figure 5. Results of S^H_H, S^O_H and $S^H_{H,\,corr}$, $S^O_{H,\,corr}$ for Pt. The bandwidth of histogram is 0.1. The sensitivities of $\hat{\theta}_{H^*}$ with respect to ΔE_H are almost identical for uncorrelated and correlated models. However, the correlation between ΔE_H and ΔE_O significantly impacts the sensitivity of $\hat{\theta}_{O^*}$ on ΔE_H, and changes the correlation from being slightly negative to highly positive. The overall shift in correlation is three orders of magnitude. Furthermore, the uncertainty ω in (2.36) also has a significant effect on $S^O_{H,\,corr}$: using a stochastic (parametric or non-parametric) model yields a sensitivity index smaller than the value from the deterministic model by an order of magnitude.

2.8 Correlated Global Sensitivity Index (CGSI)

In this section we compute the CGSIs for the correlation model previously used to determine the CLSIs. According to (2.13), the CGSIs are

$$\xi^{H}_{H,\,corr} = \int \ |S^{H}_{H,\,corr}(\Delta E_H)|^2 p(\Delta E_H)\ d(\Delta E_H), \tag{2.44}$$

$$\xi^{O}_{H,\,corr} = \int \ |S^{O}_{H,\,corr}(\Delta E_H)|^2 p(\Delta E_H)\ d(\Delta E_H). \tag{2.45}$$

As discussed in Section 7, we assume the prior distribution of ΔE_H on Pt satisfies

$$\Delta E_H \in \mathbb{E}_H, \qquad \mathbb{E}_H \sim Gamma(a_H, b_H) \tag{2.46}$$

with a PDF of

$$p(\Delta E_H) = \frac{1}{b_H^{a_H}\Gamma(a_H)} \Delta E_H^{\ a_H - 1} \exp(-\frac{\Delta E_H}{b_H}) \qquad for\ x > 0, \tag{2.47}$$

using the data in Table 1 according to (2.30). Then, from $S^{H}_{H,\,corr}$ and $S^{O}_{H,\,corr}$, we numerically calculate the CGSIs according to (2.44) and (2.45). The results are shown in Figure 6. Correlations have only a slight effect on the H coverage as we consider the H binding energy as an independent variable and the O binding energy as the dependent parameter.

2.9 Remarks on Non-parametric Correlated GSIs using Generalized Polynomial Chaos

In Section 2.2 and the Appendix A.5 we discuss the computation of the proposed correlated sensitivity indices using either direct numerical integration or Monte Carlo methods. Here, we briefly discuss the use of the Polynomial Chaos Expansion (PCE) method as an alternative to numerical integration (which is limited by the

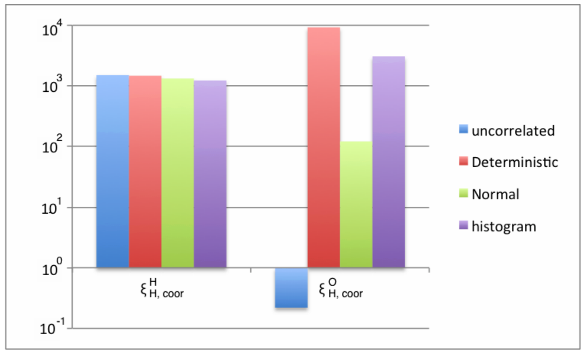

Figure 6. Uncorrelated and correlated GSI results, ξ_H^H, ξ_H^O and $\xi_{H,\,coor}^H$, $\xi_{H,\,coor}^O$ of Pt, computed by (2.7), (2.44) and (2.45). The correlation between ΔE_H and ΔE_O does not influence the sensitivity of $\hat{\theta}_{H^*}$ with respect to ΔE_H. Correlations do, however, impact the sensitivity of $\hat{\theta}_{O^*}$ with respect to ΔE_H. For $\xi_{H,\,coor}^O$, we find that the CGSI from the purely data-driven non-parametric model are significantly higher than the that from the parametric (normal distribution) model.

dimensions of the parameter space)[42, 135]. Polynomial Chaos methods rely on expanding the model $f(\lambda)$, defined in (2.2) in a series expansion, resulting in an approximation of the type

$$f(\lambda) \approx \sum_{i=1}^{d} c_i P^{(i)}(\lambda), \tag{2.48}$$

where d is the order of expansion approximation, c_i are the expansion coefficients and $P^{(i)}(\lambda)$ are the polynomials forming the basis $\{P^{(0)}, \ldots, P^{(d)}\}$. The chosen polynomials are orthogonal with respect to the probability measure of λ, i.e.

$$\int_{\Lambda} P^{(k)}(\lambda) P^{(l)}(\lambda) p(\lambda)\; d\lambda = \delta_{kl}, \qquad \forall k, l = 0, \ldots, d \tag{2.49}$$

where δ_{kl} is the Kronecker delta function, $p(\lambda)$ is the PDF of model parameters λ and Λ denotes the parameter space. Usually, $\lambda = (\lambda_1, \ldots, \lambda_n)$ is assumed to be independent. Carefully selecting the distribution (Gaussian, Gamma, etc) allows the corresponding basis to be given through the Askey scheme [136] and can be implemented using the software DAKOTA [1]. Using a previous approximation (2.48) allows calculation of the variance-based global sensitivity index (Sobol's indices) directly and without extra cost [21]; see also the implementations in [1]. For instance, in the context of the applications discussed here, and in [26], the authors use PCE to analyze the problem of global sensitivity analysis for chemical processes, assuming uniformly distributed, uncorrelated parameters.

PCE can also be generalized to arbitrary distributions with the non-parametric models considered here. Such models include the use of histograms or kernel-based distributions. Indeed, in [98], the authors introduce a PCE with arbitrary probability measures, which can be either discrete, continuous, or discretized continuous. This form of PCE can also be specified either analytically (as probability density/cumulative distribution functions) or numerically (as various histograms or as raw data sets, like the ones arising in non-parametric methods). Only a few moments of the underlying distribution, and not on the specific functional form of the probability distribution functions, are required for these methods. Therefore, these methods do not apply to distributions which are not characterized by their moments, such as the lognormal.

We also carry out PCE for parameters λ which have correlated components. Indeed, in [96], Navarro et al. give us a way to instruct PCE for general multivariate distributions with correlated variables. In our case, the Sobols indices are not necessarily positive, and the contribution due to correlation can completely cancel the contribution from the variable itself, resulting in a small Sobol's value even though such a variable can have a large impact on the outcome [96]. It should be possible to apply the derivative-based sensitivity, as defined in section 2.1.2, by replacing $f(\lambda)$ with the approximate PCE of the model. And it is also possible to combine the methods of [98] for the non-parametric aspects of the problem, and use [96] to address the correlations in the parameters. We expect to return to this implementation of PCE for non-parametric correlative sensitivity analysis in future work.

2.10 Conclusions

In this chapter we proposed a non-parametric method for the local and global sensitivity analysis of models with correlated parameter dependencies. The resulting mathematical tools are applied on a benchmark Langmuir competitive adsorption model. Such systems are encountered in catalytic oxidation, such as emissions abatement, small scale power generation, fuel cells and batteries. In the system considered here, parameter correlations stem from correlated quantum-scale computational data. The necessity of using non-parametric methods arose from the limited amount of available quantum-scale data. In our methodology, we employed gradient-based methods to compute correlative local and global sensitivity indices to illustrate the relative effects of parameter perturbations (or errors and uncertainties) in the hydrogen and oxygen binding energies on the coverages. We observed that identification of influential parameters depends critically on whether or not correlations between parameters are taken into account. Furthermore, the impact of uncertainty in the correlation and the necessity of non-parametric approaches on the sensitivity indices are demonstrated. Finally, we briefly discussed the applicability of Polynomial Chaos expansion methods for the efficient simulation of sensitivity indices.

CHAPTER 3

MODEL-FORM UNCERTAINTY QUANTIFICATION FOR PROBABILISTIC GRAPHICAL MODELS

In this Chapter, we develop UQ and SA methods for PGMs, along with rigorous, robust and computable prediction guarantees. Key UQ challenges in the PGMs include: (a) model-form and parametric uncertainties due to sparse, heterogeneous data used to learn the PGM; (b) multiple sources of uncertainty from the learning of each one of PGM nodes; (c) uncertainty in the learned graph structures. Therefore, our goal is to build an Uncertainty Quantification (UQ) framework for PGMs which, takes advantage of the graphical structure of the PGM, is able to quantify and distinguish the multiple sources of uncertainties in the model as well as assess and/or discover correlations and causal relationships between components of the model. Our mathematical tools to address such issues are based in part on information theory, precisely due to the scalability of the Kullback-Leibler (KL) divergence on graphs.

3.1 Background

3.1.1 Model-form UQ for general probabilistic models

Uncertainties arising from the fluctuations of the QoI's associated to a given baseline model P are of referred to as *aleatoric* and occur when sampling the model. They are handled by standard tools (e.g. central limit theorems, concentrations inequalities, bayesian posteriors). By contrast *model-form* uncertainties are associated to an incomplete knowledge of the model itself (i.e. *model misspecification*) and the main goal is to understand the resulting biases for QoI's. This type of uncertainty (also known as epistemic) arises, for example, from lack of data and/or limited knowledge as well as when the real model is too complex to be handled computationally (model approximation or model reduction).

In general, to apply the "model-form UQ" around a baseline (approximate, surrogate, etc.) model P, we consider all possible models Q of X which is "close to" P in KL-divergence, i.e., consider the ambiguity set $\mathcal{Q}$ defined by

$$\mathcal{Q} := \mathcal{D}^\eta := \{\text{all PGM } Q : R(Q||P) \leq \eta\} \tag{3.1}$$

with model misspecification η. Then in this case, the **predictive uncertainty** for the QoI $f(X)$, as defined in (1.4), would be

$$I^{\pm}(f(X), P; \mathcal{D}^\eta) = \sup_{Q \in \mathcal{D}^\eta}/\inf \, \mathbb{E}_Q\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right] \tag{3.2}$$

A key point is that the parameter η is *not* necessarily small! Furthermore, η can be either calculated as the KL distance of the baseline model P *from the available data*–see Fig. 7(R), or η can take arbitrary fixed values that correspond to model perturbations associated with *local* or *global* sensitivity analysis, see Section 3.3 for a more complete discussion.

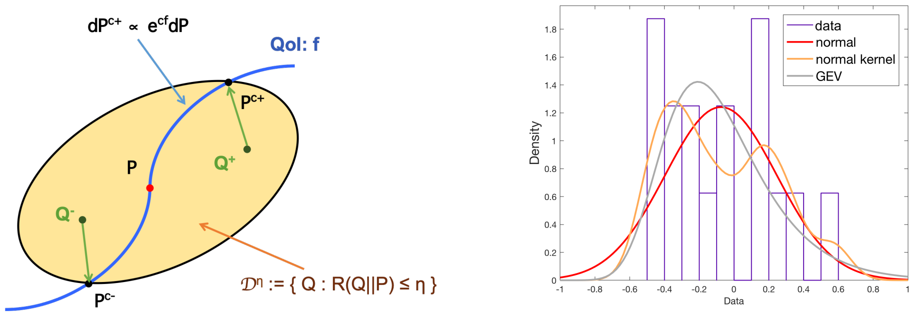

Figure 7. Left: [∞-dimensional, non-parametric] neighborhood of model P in KL divergence; the blue line represents a parametric family; $P^{\pm}$ is where we achieve the UQ indices/bounds $I^{\pm}$ on the space w.r.t QoI $f(X)$ (i.e., tightness of the bounds, see Lemma B.4). Right: Example of a source for model-form uncertainty: different probabilistic models/CPDs for sparse data of a PGM node. The red curve is used to build a baseline Gaussian model, P, the gray curve is another parametric model (Generalized Extreme Value (GEV) distribution) which fits the data better, and the yellow curve is a non-parametric model (Kernel Density Estimation (KDE) with normal kernel).

We remark that the domain $\mathcal{D}^\eta$ is an infinite dimensional space with respect to model parameters, as it includes not only parametric models but also non-parametric models. However, the predictive uncertainty shown in (3.2) is computable by a one dimensional optimization problem, and it is tight with only the baseline model P due to the properties of KL divergence. More specifically:

Theorem 3.1 *Let P be a probability measure with X, and $f(X)$ be a QoI depends on X. If $f(X)$ has finite moment generating function (MGF), $\mathbb{E}_P\left[e^{\pm c\bar{f}(X)}\right]$, in a neighborhood of the origin, then for the predictive uncertainty defined in* (3.2)*, there exist $0 < \eta_{\pm} \leq \infty$, such that for any $\eta \leq \eta_{\pm}$,*

$$\begin{aligned} I^{\pm}(f(X), P; \mathcal{D}^\eta) &= \sup_{Q\in\mathcal{D}^\eta}/\inf \mathbb{E}_Q\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right] \\ &= \pm\inf_{c>0}\left[\frac{1}{c}\log\mathbb{E}_P\left[\pm e^{c\bar{f}(X)}\right] + \frac{\eta}{c}\right] \\ &= \mathbb{E}_{Q^{\pm}}\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right] \end{aligned} \tag{3.3}$$

where $\bar{f}(X)$ is the centered QoI, $\bar{f}(X) := f(X) - \mathbb{E}_P[f(X)]$, and $Q^{\pm} = Q^{\pm}(\eta)$ are probability measures given by the elements $Q^{\pm} = P^{\pm c_{\pm}}$ where

$$dP^{\pm c_{\pm}} = \frac{e^{\pm c_{\pm} f(x)}}{\mathbb{E}_P\left[e^{\pm c_{\pm} f(X)}\right]}\, dP \tag{3.4}$$

and $c_{\pm}$ are the unique solutions of

$$R(P^{\pm c_{\pm}}||P) = \eta\,. \tag{3.5}$$

To prove the theorem, we first show Lemma B.3 and Lemma B.4 which are presented in [29, 50], and we include the proof for the lemmas and the theorem in Appendix B for completeness.

Example: Consider a random variable X, for which we have samples shown in Figure 7 (Right) as a histogram. Using the data, we build a baseline Gaussian model P with density $p(x) \sim \mathcal{N}(\mu_P, \sigma_P^2)$ (for instance, using MLE). Then for the QoI $f(X) = X$, and any other alternative model $\tilde{Q}$ satisfying $\tilde{Q} \in \mathcal{D}^\eta$ in (3.1) (which may include other possible models like generalized extreme value (GEV) distribution or kernel density estimation (KDE) shown in Figure 7 (Right), or the unknown real model). By Theorem 3.1 (a), we have

$$\begin{aligned} \mathbb{E}_{\tilde{Q}}[f] - \mathbb{E}_P[f] &\leq \sup_{\mathcal{D}^\eta} \mathbb{E}_Q[f] - \mathbb{E}_P[f] = I^+(f(X), P; \mathcal{D}^\eta) \\ &= \inf_{c>0}\left[\frac{1}{c}\log\int e^{c(x-\mu_P)}P(dx) + \frac{\eta}{c}\right] \\ &= \inf_{c>0}\left[\frac{1}{2}\sigma_P^2 c + \frac{\eta}{c}\right] = \sigma_P\sqrt{2\eta} \end{aligned} \tag{3.6}$$

where we use the Gaussian property that the MGF of P, $\mathbb{E}_P\left[e^{cX}\right] = e^{\mu_P c + \sigma_P^2 c/2}$. Similarly we obtain the lower bound,

$$\mathbb{E}_{\tilde{Q}}[f] - \mathbb{E}_P[f] \geq \inf_{\mathcal{D}^\eta} \mathbb{E}_Q[f] - \mathbb{E}_P[f] = I^-(f(X), P; \mathcal{D}^\eta) = -\sigma_P\sqrt{2\eta} \tag{3.7}$$

therefore, we can quantify the model-form uncertainty of P for the prediction of f by the indices $I^{\pm}(f(X), P; \mathcal{D}^\eta)$ in the set $\mathcal{D}^\eta$.

Furthermore, by Theorem 3.1 (b), we can find the optimizer $Q^{\pm} \in \mathcal{D}^{\eta}$ which achieve the equality, i.e.

$$q^{\pm}(x) \propto e^{\pm c_{\pm} x} p(x) \qquad \Rightarrow \qquad q^{\pm}(x) \sim \mathcal{N}(\mu_P \pm c_{\pm}\sigma_P^2, \sigma_P^2) \tag{3.8}$$

and

$$R(P^{\pm c_{\pm}} || P) = \eta \qquad \Rightarrow \qquad c_{\pm} = \sqrt{\frac{2\eta}{\sigma_P^2}} \tag{3.9}$$

thus, $q^{\pm}(x) \sim \mathcal{N}(\mu_P \pm \sqrt{2\sigma_P^2\eta}, \sigma_P^2)$, and it satisfies

$$\mathbb{E}_{Q^{\pm}}[f] - \mathbb{E}_P[f] = I^{\pm}(f(X), P; \mathcal{D}^{\eta}) = \pm\sigma_P\sqrt{2\eta} \tag{3.10}$$

Note that $Q^{\pm}$ still follow the Gaussian distribution in this case.

3.2 Main Results

3.2.1 Model-form UQ indices for PGMs

Here we want to extend the model-form UQ methods for the PGMs, along with rigorous, robust and computable prediction *guarantees*: (a) model-form and parametric uncertainties due to sparse, heterogeneous data used to learn the PGM; (b) multiple sources of uncertainty from the learning of each one of PGM nodes; (c) uncertainty in the learned graph structures. Therefore, for a PGM $p(x) = \prod_{i=1}^{n} p(x_i|x_{\pi_i})$, we want to look at the predictive uncertainty (3.2), for a QoI which is a function of one node in the PGM, i.e.,

$$\text{for } f(X_k),\ 1 \le k \le n \tag{3.11}$$

with the model misspecification η,

$$\sup_{Q \in \mathcal{D}^{\eta}}/\inf\ \mathbb{E}_Q[f(X_k)] - \mathbb{E}_P[f(X_k)] \tag{3.12}$$

where $\mathcal{D}^\eta$ is the ambiguity set defined in (3.1), i.e., when we perturb the baseline model P to an alternative model Q, altering both the structure and the CPDs, under model misspecification η. Then we obtain the following theorem which is a PGM analogue of Theorem 3.1:

Theorem 3.2 *Let P be a PGM defined as* (1.1)*, and $f(X_k)$ be a QoI only depends on X_k. If $f(X_k)$ has finite moment generating function (MGF), $\mathbb{E}_P\left[e^{c\bar{f}(X_k)}\right]$, in a neighborhood of the origin, then for the predictive uncertainty defined in* (3.12)*, there exist $0 < \eta_\pm \le \infty$, such that for any $\eta \le \eta_\pm$,*

$$\begin{aligned} I^\pm(f(X_k), P; \mathcal{D}^\eta) &= \sup_{Q\in\mathcal{D}^\eta}/\inf \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\ &= \pm \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{\{k\}}}\left[\pm e^{c\bar{f}(X_k)}\right] + \frac{\eta}{c}\right] \\ &= \mathbb{E}_{Q^\pm}\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \end{aligned} \tag{3.13}$$

where $\bar{f}(X_k)$ is the centered QoI, $\bar{f}(X_k) := f(X_k) - \mathbb{E}_P\left[f(X_k)\right]$, $P_{\{k\}}$ is the marginal distribution of X_k with respect to P, and $Q^\pm = Q^\pm(\eta) \in \mathcal{D}^\eta$ are probability measures given by the elements $Q^\pm = P^{\pm c_\pm}$ where

$$\frac{dP^{\pm c_\pm}}{dP} = \frac{e^{\pm c_\pm f(x_k)}}{\mathbb{E}_P\left[e^{\pm c_\pm f(X_k)}\right]} \tag{3.14}$$

and $c_\pm$ are the unique solutions of

$$R(P^{\pm c_\pm}||P) = \eta \tag{3.15}$$

More specifically, without loss of generality, if we assume $j < i$ for all $j \in \pi_i^P$, then $Q^\pm$ is given by

$$q^\pm(x_i|x_{\pi_i^{Q^\pm}}) \equiv p(x_i|x_{\pi_i^P}) \qquad \text{for all } i > k \text{ and } \pi_i^{Q^\pm} \equiv \pi_i^P \tag{3.16}$$

$$q^\pm(x_k|x_{\pi_k^{Q^\pm}}) = \frac{e^{\pm c_\pm f(x_k)}}{\mathbb{E}_{P_{k|\pi_k^P}}\left[e^{\pm c_\pm f(X_k)}\right]} \cdot p(x_k|x_{\pi_k^P}) \qquad \text{for all } x_{\pi_k^{Q^\pm}} \text{ and } \pi_k^{Q^\pm} = \pi_k^P \tag{3.17}$$

and

$$q^{\pm}(x_i|x_{\pi_i^{Q^{\pm}}}) = \frac{\mathbb{E}_{P_{i+1|\pi_{i+1}^P}}\left[\cdots \mathbb{E}_{P_{k|\pi_k^P}}\left[e^{\pm c_{\pm} f(X_k)}\right]\right]}{\mathbb{E}_{P_{i|\pi_i^P}}\left[\mathbb{E}_{P_{i+1|\pi_{i+1}^P}}\left[\cdots \mathbb{E}_{P_{k|\pi_k^P}}\left[e^{\pm c_{\pm} f(X_k)}\right]\right]\right]} p(x_i|x_{\pi_i^P}) \tag{3.18}$$

for all $i = 1, 2, \ldots, k-1$ *and* $\pi_i^P \subset \pi_i^{Q^{\pm}} \subset \{1, \ldots, i-1\}$.

Proof: The proof of Theorem 3.2 relies in part on Theorem 3.1, however a new important element is the role of the structure of the graph of the PGM, as is described precisely in (3.16)-(3.18). For part (a), consider $f(X) = f(X_k)$ and $p(x) = \prod_{i=1}^n p(x_i|x_{\pi_i})$, by (3.3), we have

$$\begin{aligned} & \sup_{Q \in \mathcal{D}^\eta}/\inf \ \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\ = \ & \pm \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_P\left[e^{\pm c\bar{f}(X_k)}\right] + \frac{\eta}{c}\right] \\ = \ & \pm \inf_{c>0}\left[\frac{1}{c}\log \underbrace{\int \cdots \int}_{x_1,\ldots,x_n} e^{\pm c\bar{f}(x_k)} \prod_{i=1}^n P(dx_i|x_{\pi_i}^P) + \frac{\eta}{c}\right] \\ = \ & \pm \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{\{k\}}}\left[e^{\pm c\bar{f}(X_k)}\right] + \frac{\eta}{c}\right] \end{aligned} \tag{3.19}$$

Then for part (b), if P satisfies $j < i$ for all $j \in \pi_i^P$, by (3.4) we have

$$\begin{aligned} & \prod_{i=1}^n q^{\pm}(x_i|x_{\pi_i^{Q^{\pm}}}) \\ = \ & \frac{e^{\pm c_{\pm} f(x_k)}}{\mathbb{E}_P\left[e^{\pm c_{\pm} f(X_k)}\right]} \prod_{i=1}^n p(x_i|x_{\pi_i^P}) \\ = \ & \frac{1}{\mathbb{E}_P\left[e^{\pm c_{\pm} f(X_k)}\right]} \prod_{i=k+1}^n p(x_i|x_{\pi_i^P}) \cdot e^{\pm c_{\pm} f(x_k)} p(x_k|x_{\pi_k^P}) \cdot \prod_{i=1}^{k-1} p(x_i|x_{\pi_i^P}) \end{aligned} \tag{3.20}$$

where $\pm c_{\pm}$ are the unique solutions of $R(P^{\pm c_{\pm}}||P) = \eta$. Therefore, we can define $Q^{\pm}$ as

$$q^{\pm}(x_i|x_{\pi_i^{Q^{\pm}}}) \equiv p^{\pm}(x_i|x_{\pi_i^P}) \qquad \text{for all } i > k \text{ and } \pi_i^{Q^{\pm}} \equiv \pi_i^P \tag{3.21}$$

$$q^{\pm}(x_k|x_{\pi_k^{Q^\pm}}) = \frac{e^{\pm c_\pm f(x_k)}}{\mathbb{E}_{P_{k|\pi_k^P}}\left[e^{\pm c_\pm f(X_k)}\right]} \cdot p(x_k|x_{\pi_k^P}) \qquad \text{for all } x_{\pi_k^{Q^\pm}} \text{ and } \pi_k^{Q^\pm} = \pi_k^P \tag{3.22}$$

and

$$q^{\pm}(x_i|x_{\pi_i^{Q^\pm}}) = \frac{\mathbb{E}_{P_{i+1|\pi_{i+1}^P}}\left[\cdots \mathbb{E}_{P_{k|\pi_k^P}}\left[e^{\pm c_\pm f(X_k)}\right]\right]}{\mathbb{E}_{P_{i|\pi_i^P}}\left[\mathbb{E}_{P_{i+1|\pi_{i+1}^P}}\left[\cdots \mathbb{E}_{P_{k|\pi_k^P}}\left[e^{\pm c_\pm f(X_k)}\right]\right]\right]} p(x_i|x_{\pi_i^P}) \tag{3.23}$$

for all $i = k-1, \ldots, 1$, where the denominators are the normalization factors for CPDs when $i \leq k$, and since the factors may depend on some values of the ancestors of X_k, x_{ρ_k}, $\pi_i^{Q^\pm}$ may differ from π_i^P as shown in Figure 10, and we have $\pi_i^P \subset \pi_i^{Q^\pm} \subset \{1, \ldots, i-1\}$. ■

Example (Inhomogeneous Markov chains): Consider the Markov chain models as a special case for the PGMs (1.1), i.e., let P with $p(x) = \prod_{i=1}^n p(x_i|x_{i-1})$ (where $p(x_1|x_0) := p(x_1)$, $\pi_i = \{i-1\}$) to be a probability measure defined on a Markov chain as shown in the following Figure:

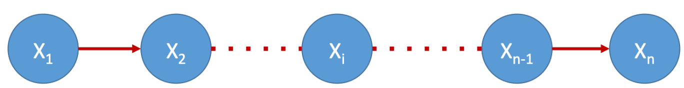

Figure 8. An inhomogeneous Markov chain consists of $X = \{X_1, X_2, \ldots, X_n\}$ with $p(x) = \prod_{i=1}^n p(x_i|x_{i-1})$.

then consider the QoI $f(X_k)$, if we perturb P with the constraint $R(Q||P) \leq \eta$, i.e. consider $Q \in \mathcal{D}^\eta$, then by Theorem 3.2, we have

$$I^{\pm}(f(X_k), P; \mathcal{D}^\eta) = \pm \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{\{k\}}}\left[e^{\pm c\tilde{f}(X_k)}\right] + \frac{\eta}{c}\right] \tag{3.24}$$

where $p_{\{k\}}(x_k) = \int \prod_{i=1}^k p(x_k|x_{k-1}) dx_{\{1,\ldots,k-1\}}$ and using (3.16)-(3.18), the optimizer $Q^\pm$ in Theorem 3.2 is obtained when

$$q^{\pm}(x_j|x_{j-1}) \equiv p(x_j|x_{j-1}) \qquad \text{for } j = k+1, \ldots, n \tag{3.25}$$

$$q^{\pm}(x_k|x_{k-1}) = \frac{e^{\pm c_\pm f(x_k)}}{\mathbb{E}_P\left[e^{\pm c_\pm f(X_k)}|x_{k-1}\right]} p(x_k|x_{k-1}) \tag{3.26}$$

$$q^{\pm}(x_j|x_{j-1}) = \frac{\mathbb{E}_P\left[e^{\pm c_{\pm} f(x_k)}|x_j\right]}{\mathbb{E}_P\left[e^{\pm c_{\pm} f(X_k)}|x_{j-1}\right]} p(x_j|x_{j-1}) \qquad \text{for } j = 1, \dots, k-1 \tag{3.27}$$

where $c_{\pm}$ are the unique solutions of

$$R(P^{\pm c_{\pm}}||P) = \eta \tag{3.28}$$

for $P^{\pm c_{\pm}}$ defined in (3.4) and $\mathbb{E}_P\left[e^{\pm c_{\pm} f(X_k)}|x_0\right] := \mathbb{E}_P\left[e^{\pm c_{\pm} f(X_k)}\right]$. Note that $Q^{\pm}$ is still a inhomogeneous Markov chain in this case.

Example (Gaussian Bayesian Networks): Gaussian Bayesian Networks (GBN), [66], is a special class of Probabilistic Graphical Models commonly used in natural and social sciences and where the CPDs (1.2) are linear and Gaussian. More specifically, for a GBN consisting of variables X, every node X_i is a linear Gaussian of its parents, i.e.,

$$p(x_i|x_{\pi_i}) = \mathcal{N}(\beta_{i0} + \beta_i^T x_{\pi_i}, \sigma_i^2) \tag{3.29}$$

with some β_0, β, and σ_i, or

$$X_i = \beta_{i0} + \beta_i^T X_{\pi_i} + \epsilon_i \tag{3.30}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. By the conjugacy properties of Gaussians, the joint distribution P becomes $p(x) = \mathcal{N}(\mu, \mathcal{C})$, i.e. it is also a Gaussian with parameters μ, $\mathcal{C}$, which can be calculated from β_{i0}, β_i, and σ_i [9].

For concreteness, we consider the GBN $p(x) = \mathcal{N}(\mu, \mathcal{C})$ in Figure 9:

Then for the QoI $f(X_4) = X_4$, if we perturb P with the constraint $R(Q||P) \le \eta$, i.e. consider $Q \in \mathcal{D}^{\eta}$, by Theorem 3.2 (3.13), we conclude that

$$I^{\pm}(f(X_4), P; \mathcal{D}^{\eta}) = \pm\sqrt{2\mathcal{C}_{44}\eta} = \pm\sqrt{2(\sigma_4^2 + \beta_{43}^2\sigma_3^2 + \beta_{42}^2\sigma_2^2)\eta} \tag{3.31}$$

and by (3.16) - (3.18), the optimizer in Theorem 3.2 is obtained when

$$q^{\pm}(x_5|x_{\pi_5^{Q^{\pm}}}) \equiv p(x_5|x_4, x_1) = \mathcal{N}(\beta_{50} + \beta_{54}x_4 + \beta_{51}x_1, \sigma_5^2) \tag{3.32}$$

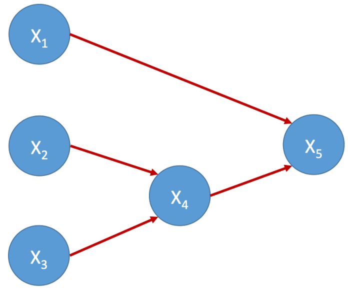

Figure 9. A GBN consists of $X = \{X_1, X_2, \ldots, X_5\}$ with $p(x) = p(x_5|x_4, x_1)p(x_4|x_3, x_2)\ p(x_3)p(x_2)p(x_1)$ where $p(x_5|x_4, x_1) = \mathcal{N}(\beta_{50} + \beta_{54}x_4 + \beta_{51}x_1, \sigma_5^2)$, $p(x_4|x_3, x_2) = \mathcal{N}(\beta_{40} + \beta_{43}x_3 + \beta_{42}x_2, \sigma_4^2)$, $p(x_3) = \mathcal{N}(\beta_{30}, \sigma_3^2)$, $p(x_2) = \mathcal{N}(\beta_{20}, \sigma_2^2)$, and $p(x_1) = \mathcal{N}(\beta_{10}, \sigma_1^2)$.

where $\pi_5^{Q^\pm} \equiv \pi_5^P = \{4, 1\}$,

$$
\begin{aligned}
q^\pm(x_4|x_{\pi_4^{Q^\pm}}) &= \frac{e^{\pm c_\pm x_4}}{\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]} \cdot p(x_4|x_{\pi_4^P}) \\
&= \frac{e^{\pm c_\pm x_4} e^{-\frac{(x_4-\beta_{40}-\beta_{43}x_3-\beta_{42}x_2)^2}{2\sigma_4^2}}}{\int_{x_4} e^{\pm c_\pm x_4} e^{-\frac{(x_4-\beta_{40}-\beta_{43}x_3-\beta_{42}x_2)^2}{2\sigma_4^2}}\ dx_4} \\
&= \frac{e^{-\frac{(x_4-\beta_{40}-\beta_{43}x_3-\beta_{42}x_2\mp c_\pm\sigma_4^2)^2}{2\sigma_4^2}}\ e^{\pm c_\pm(\beta_{43}x_3+\beta_{42}x_2)}}{\int_{x_4} e^{-\frac{(x_4-\beta_{40}-\beta_{43}x_3-\beta_{42}x_2\mp c_\pm\sigma_4^2)^2}{2\sigma_4^2}}\ dx_4\ e^{\pm c_\pm(\beta_{43}x_3+\beta_{42}x_2)}} \\
&= \mathcal{N}\left(\beta_{40} + \beta_{43}x_3 + \beta_{42}x_2 \pm c_\pm\sigma_4^2, \sigma_4^2\right)
\end{aligned}
\tag{3.33}
$$

where $\pi_4^{Q^\pm} \equiv \pi_4^P = \{3, 2\}$, and

$$
\begin{aligned}
q^\pm(x_3|x_{\pi_3^{Q^\pm}}) &= \frac{\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]}{\mathbb{E}_{P_3}\left[\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]\right]} p(x_3) \\
&= \frac{e^{\pm c_\pm(\beta_{43}x_3+\beta_{42}x_2)} e^{-\frac{(x_3-\beta_{30})^2}{2\sigma_3^2}}}{\int_{x_3} e^{\pm c_\pm(\beta_{43}x_3+\beta_{42}x_2)} e^{-\frac{(x_3-\beta_{30})^2}{2\sigma_3^2}}\ dx_3} \\
&= \frac{e^{-\frac{(x_3-\beta_{30}\mp c_\pm\beta_{43}\sigma_3^2)^2}{2\sigma_3^2}}\ e^{\pm c_\pm(\beta_{42}x_2)}}{\int_{x_3} e^{-\frac{(x_3-\beta_{30}\mp c_\pm\beta_{43}\sigma_3^2)^2}{2\sigma_3^2}}\ dx_3\ e^{\pm c_\pm(\beta_{42}x_2)}} \\
&= \mathcal{N}\left(x_3 - \beta_{30} \mp c_\pm\beta_{43}\sigma_3^2, \sigma_3^2\right)
\end{aligned}
\tag{3.34}
$$

$$
\begin{aligned}
q^{\pm}(x_2|x_{\pi_2^{Q^\pm}}) &= \frac{\mathbb{E}_{P_3}\left[\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]\right]}{\mathbb{E}_{P_2}\left[\mathbb{E}_{P_3}\left[\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]\right]\right]} p(x_2) \\
&= \frac{e^{\pm c_\pm(\beta_{42}x_2)} e^{-\frac{(x_2-\beta_{20})^2}{2\sigma_2^2}}}{\int_{x_2} e^{\pm c_\pm(\beta_{42}x_2)} e^{-\frac{(x_2-\beta_{20})^2}{2\sigma_2^2}} dx_2} \\
&= \frac{e^{-\frac{(x_3-\beta_{20}\mp c_\pm\beta_{42}\sigma_2^2)^2}{2\sigma_2^2}}}{\int_{x_2} e^{-\frac{(x_2-\beta_{20}\mp c_\pm\beta_{42}\sigma_2^2)^2}{2\sigma_2^2}} dx_2} \\
&= \mathcal{N}\left(x_2-\beta_{20}\mp c_\pm\beta_{42}\sigma_2^2, \sigma_2^2\right) \qquad (3.35)
\end{aligned}
$$

$$
\begin{aligned}
q^{\pm}(x_1|x_{\pi_1^{Q^\pm}}) &= \frac{\mathbb{E}_{P_2}\left[\mathbb{E}_{P_3}\left[\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]\right]\right]}{\mathbb{E}_{P_1}\left[\mathbb{E}_{P_2}\left[\mathbb{E}_{P_3}\left[\mathbb{E}_{P_{4|\pi_4^P}}\left[e^{\pm c_\pm X_4}\right]\right]\right]\right]} p(x_1) \\
&= p(x_1) = \mathcal{N}(\beta_{10}, \sigma_1^2) \qquad (3.36)
\end{aligned}
$$

where $\pi_3^{Q^\pm} = \pi_2^{Q^\pm} = \pi_1^{Q^\pm} = \emptyset$. Then by (3.15), we have

$$
\begin{aligned}
&\pm c_\pm \mathbb{E}_{Q^\pm}[x_4] - \log \mathbb{E}_P\left[e^{\pm c_\pm x_4}\right] = \eta \\
\Rightarrow \qquad &\pm c_\pm = \pm\sqrt{\frac{2\eta}{\mathcal{C}_{44}}} = \pm\sqrt{\frac{2\eta}{\sigma_4^2+\beta_{43}^2\sigma_3^2+\beta_{42}^2\sigma_2^2}} \qquad (3.37)
\end{aligned}
$$

thus,

$$
q^{\pm}(x_4|x_3,x_2) = \mathcal{N}\left(\beta_{40}+\beta_{43}x_3+\beta_{42}x_2 \pm \sigma_4^2\sqrt{\frac{2\eta}{\sigma_4^2+\beta_{43}^2\sigma_3^2+\beta_{42}^2\sigma_2^2}}, \sigma_4^2\right) \qquad (3.38)
$$

$$
q^{\pm}(x_3) = \mathcal{N}\left(\beta_{30} \pm \sigma_3^2\sqrt{\frac{2\eta}{\sigma_4^2+\beta_{43}^2\sigma_3^2+\beta_{42}^2\sigma_2^2}}, \sigma_3^2\right) \qquad (3.39)
$$

$$
q^{\pm}(x_2) = \mathcal{N}\left(\beta_{20} \pm \sigma_2^2\sqrt{\frac{2\eta}{\sigma_4^2+\beta_{43}^2\sigma_3^2+\beta_{42}^2\sigma_2^2}}, \sigma_2^2\right) \qquad (3.40)
$$

Note that for $q^{\pm}(x_3|x_{\pi_3^{Q^\pm}})$, x_2 show up in the normalization factors based on (3.18), however, since $f(X_4) = X_4$ is linear and all the random variables are linearly depend on their parents in GBN as shown in (3.30), the terms with x_2 are canceled out from numerator and denominator, i.e. $2 \notin \pi_3^{Q^\pm}$ and $\pi_3^{Q^\pm} = \pi_3^P$. In general, we can conclude the result by the following Corollary for this special case in GBN:

Corollary 3.3 *Let P be a GBN satisfies* (3.29)*, and $f(X_k) = aX_k + b$ be a QoI only depends on X_k linearly. Then for the predictive uncertainty defined in* (3.12)*, we have*

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta}) = \pm\sqrt{2a^2\mathcal{C}_{kk}\eta} \tag{3.41}$$

where $\mathcal{C}_{kk}$ is the variance for the marginal distribution of X_k. Furthermore, the optimizer $Q^{\pm} = Q^{\pm}(\eta) \in \mathcal{D}^{\eta}$ given by Theorem 3.2 (3.16)-(3.18) *are also GBNs with same graph structure as P.*

Example (General PGM): Consider a general PGM as shown in the left of Figure 10, and given by

$$p(x) = p(x_1)p(x_2)p(x_3|x_2, x_1)p(x_4)p(x_5|x_3)p(x_6|x_4, x_3)p(x_7|x_6, x_5) \tag{3.42}$$

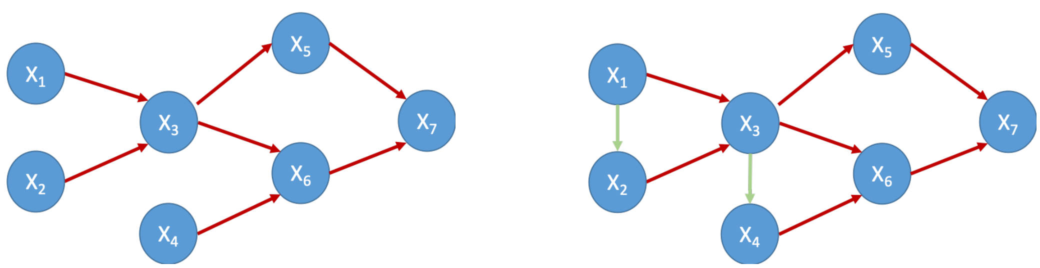

Figure 10. Left: An example of the structure of baseline PGM P; Right: The structure of optimizer $Q^{\pm}$ in Theorem 3.2 with QoI $f(X_6)$ based on (3.16) - (3.18). Note that since the normalization factor for $q^{\pm}(x_6|x_{\pi_6^{Q^{\pm}}})$ depends on X_3 and X_4, i.e. $\pi_6^{Q^{\pm}} = \{3, 4\}$, it propagates to $q^{\pm}(x_4|x_{\pi_4^{Q^{\pm}}})$ by (3.18), so $\pi_4^{Q^{\pm}} = \{3\} \cup \pi_4^P = \{3\}$, which create a new connection from X_3 to X_4 in $Q^{\pm}$. Same for the new connection from X_1 to X_2.

Then for a QoI $f(X_6)$, by Theorem 3.2, we have

$$\begin{aligned}
& I^{\pm}(f(X_6), P; \mathcal{D}^{\eta}) \\
= \;& \pm \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{\{k\}}}\left[\pm e^{c\bar{f}(X_k)}\right] + \frac{\eta}{c}\right] \\
= \;& \pm \inf_{c>0}\left[\frac{1}{c}\log \int \pm e^{c\bar{f}(x_6)} P(dx_6|x_4, x_3)P(dx_4)P(dx_3|x_2, x_1)P(dx_2)P(dx_1) + \frac{\eta}{c}\right]
\end{aligned} \tag{3.43}$$

and by (3.16) - (3.18), the optimizer in Theorem 3.2 is obtained when

$$q^{\pm}(x_7|x_{\pi_7^{Q^\pm}}) \equiv p(x_7|x_6, x_5) \tag{3.44}$$

where $\pi_7^{Q^\pm} \equiv \pi_7^P = \{6, 5\}$,

$$q^{\pm}(x_6|x_{\pi_6^{Q^\pm}}) = \frac{e^{\pm c_\pm x_6}}{\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]} \cdot p(x_6|x_4, x_3) \tag{3.45}$$

where $\pi_6^{Q^\pm} \equiv \pi_6^P = \{4, 3\}$, and

$$\begin{aligned} q^{\pm}(x_5|x_{\pi_5^{Q^\pm}}) &= \frac{\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]}{\mathbb{E}_{P_{5|\{3\}}}\left[\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]\right]} p(x_5|x_3) \\ &= p(x_5|x_3) \end{aligned} \tag{3.46}$$

since X_5 and X_6 are conditional independent given X_3, $\mathbb{E}_{P_{5|\{3\}}}\left[\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]\right] = \mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]$ given $X_3 = x_3$, so $\pi_5^{Q^\pm} \equiv \pi_5^P = \{3\}$. Note that, in general, we can conclude that only X_{ρ_k} may have different parents set in $Q^\pm$ with QoI $f(X_k)$, and

$$q^{\pm}(x_4|x_{\pi_4^{Q^\pm}}) = \frac{\mathbb{E}_{P_{5|\{3\}}}\left[\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]\right]}{\mathbb{E}_{P_4}\left[\mathbb{E}_{P_{5|\{3\}}}\left[\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]\right]\right]} p(x_4) \tag{3.47}$$

since both normalization factors on the numerator and denominator depend on $X_{\pi_6} \cup X_{\pi_5} = \{X_4, X_3\}$, so in general, we have $\pi_4^{Q^\pm} = \pi_4^P \cup \{3\} = \{3\}$, i.e., there is a new connection $X_3 \to X_4$ in $Q^\pm$, and

$$q^{\pm}(x_3|x_{\pi_3^{Q^\pm}}) = \frac{\mathbb{E}_{P_4}\left[\mathbb{E}_{P_{5|\{3\}}}\left[\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]\right]\right]}{\mathbb{E}_{P_{3|\{2,1\}}}\left[\mathbb{E}_{P_4}\left[\mathbb{E}_{P_{5|\{3\}}}\left[\mathbb{E}_{P_{6|\{4,3\}}}\left[e^{\pm c_\pm X_6}\right]\right]\right]\right]} p(x_3|x_2, x_1) \tag{3.48}$$

where $\pi_3^{Q^\pm} \equiv \pi_3^P = \{2, 1\}$ since the normalization factors do not contain other variables. And we can do the same for X_2 and X_1 to get the entire structure of $Q^\pm$ which has another new connection $X_1 \to X_2$, and the results are shown in Figure 10 (Right).

3.2.2 Chain rule and interpreting the model misspecification parameter in PGMs

For the *model misspecification* parameter η in the uncertainty domain $\mathcal{D}^\eta = \{Q : R(Q||P) \leq \eta\}$, it describes our confidence to the baseline model P and thus we refer to η as "model misspecification". For instance if η is small, $\mathcal{D}^\eta$ includes only small perturbations of the baseline P. however, a key point in our formulation is that the parameter η is not necessarily small in general. As we discuss in detail the Section 3.3, η can be calculated as the KL distance of the baseline model P from the available data–see Fig 7 (R); this η value would be *a surrogate* for the distance of the baseline model from the "real" model. Alternatively, η can take arbitrary fixed values that correspond to model perturbations associated with *local* (small η) or *global* sensitivity analysis (larger η) in the same mathematical framework.
Moreover, based on the PGM structure, we can apply the *chain rule of KL divergence* [20], which gives us

Lemma 3.1 ***[Chain Rule of Relative Entropy for PGMs]*** *For any two PGMs P and Q with densities $p(x) = \prod_{i=1}^n p(x_i|x_{\pi_i^P})$ and $q(x) = \prod_{i=1}^n q(x_i|x_{\pi_i^Q})$, we have*

$$R(Q||P) = \sum_{i=1}^n \mathbb{E}_{Q_{\pi_i^Q \cup \pi_i^P}} \left[R(Q_{i|\pi_i^Q}||P_{i|\pi_i^P}) \right] = \sum_{i=1}^n \mathbb{E}_{Q_{\pi_i^Q \cup \pi_i^P}} \left[\eta_i^{\pi_i^Q \cup \pi_i^P} \right] \tag{3.49}$$

where $\eta_i^{\pi_i^Q \cup \pi_i^P} := R(Q_{i|\pi_i^Q}||P_{i|\pi_i^P})$ are the conditional relative entropy between $Q_{i|\pi_i^Q}$ and $P_{i|\pi_i^P}$ with given $X_{\pi_i^Q \cup \pi_i^P} = x_{\pi_i^Q \cup \pi_i^P}$, i.e.

$$R(Q_{i|\pi_i^Q}||P_{i|\pi_i^P}) = \int \log \frac{Q(dx_i|x_{\pi_i^Q})}{P(dx_i|x_{\pi_i^P})} Q(dx_i|x_{\pi_i^Q}) \tag{3.50}$$

Proof.

$$\begin{aligned}
R(Q||P) &= \int \log \frac{\prod_{i=1}^n Q(dx_i|x_{\pi_i^Q})}{\prod_{i=1}^n P(dx_i|x_{\pi_i^P})} \prod_{j=1}^n Q(dx_j|x_{\pi_j^Q}) \\
&= \int \sum_{i=1}^n \log \frac{Q(dx_i|x_{\pi_i^Q})}{P(dx_i|x_{\pi_i^P})} \prod_{j=1}^n Q(dx_j|x_{\pi_j^Q}) \\
&= \sum_{i=1}^n \int \log \frac{Q(dx_i|x_{\pi_i^Q})}{P(dx_i|x_{\pi_i^P})} Q(dx_i|x_{\pi_i^Q}) \cdot \prod_{j\in\{\rho_i^Q \cup \rho_i^P\}} Q(dx_j|x_{\pi_j^Q}) \\
&= \sum_{i=1}^n \mathbb{E}_{Q_{\pi_i^Q \cup \pi_i^P}} \left[R(Q_{i|\pi_i^Q}||P_{i|\pi_i^P}) \right] \\
&= \sum_{i=1}^n \mathbb{E}_{Q_{\pi_i^Q \cup \pi_i^P}} \left[\eta_i^{\pi_i^Q \cup \pi_i^P} \right] \qquad (3.51)
\end{aligned}$$

where $\eta_i^{\pi_i^Q \cup \pi_i^P}$ is the KL divergence between CPDs $Q_{i|\pi_i^Q}$ and $P_{i|\pi_i^P}$ with given parents $x_{\pi_i^Q} \cup x_{\pi_i^P}$. ■

Therefore, we can break down the calculation of the aforementioned model misspecification R($Q||P$) in Theorem 3.2 into separate PGM components, which reduces the calculation of model misspecification η to individual node and CPD calculations. Furthermore, this decomposition localizes the uncertainty from multiple sources corresponding to different PGM components, and we will use this property to defines specific ambiguity sets which allow us to do model-form sensitivity analysis for each component on the PGM as shown in next subsection.

3.2.3 Model-form sensitivity indices for PGMs

Since the existing sensitivity analysis methods, e.g., gradient and ANOVA-based methods, (a) cannot handle UQ tasks with model uncertainty (not just parametric), e.g., Fig. 7(R), and (b) it is not obvious how they will take advantage of the inherent graphical structure in PGMs, such as conditional independence, here we use concept of predictive uncertainty in (1.4) with suitable ambiguity sets to discuss different kinds of model-form sensitivity analysis methods for PGM, where all make sense in different contexts/perturbations, and could be useful for different application/under different constraints. In all cases, we isolate a single node l on the PGM for a "stress test", and we keep all the other PGM nodes fixed; then we can vary a combination of parents and CPDs for the node l; the CPDs vary in a non-parametric neighborhood of a baseline CPD $p(x_l|x_{\pi_l})$ of the baseline PGM P with model misspecification η_l. The results give us a rank of sensitivities for each node which can provide a strategy to "close the data-model-predictions loop" and design better models by targeting the most under-performing components of our PGMs and address *trade-offs* between model complexity, data & predictive guarantees. Here we distinguish two cases, although various combinations can be considered with the same mathematical tools:

1. In Part 1 we keep all the nodes on PGM fixed except l, for which the parents and the CPDs can vary in a non-parametric ambiguity set $\mathcal{D}_l^{\eta_l}$, see the definition of (3.52).

2. In Part 2 we keep all the nodes on the PGM and their parents fixed, i.e., we keep the graph structure of the PGM, and allow non-parametric variability in the CPD of node l, see the definition of (3.59).

3.2.4 Model-form Sensitivity Indices, Part 1 – vary graph structure and CPD

To isolate and rank the impact uncertainties of each node, based on the results we find in previous subsection for the model misspecification η, we consider the specific domain that only has a perturbation on $P_{l|\pi_l}$ from P, and define the ambiguity set $\mathcal{Q}$ by

$$\mathcal{Q} := \mathcal{D}_l^{\eta_l} \left\{ \begin{array}{c} \text{all PGM } Q : R(Q_{l|\pi_l^Q} || P_{l|\pi_l^P}) \leq \eta_l \text{ for all } x_{\pi_i^P} \cup x_{\pi_i^Q}, \\ Q_{j|\pi_j} \equiv P_{j|\pi_j} \text{ for all } j \neq l \end{array} \right\} \tag{3.52}$$

where π_l^Q is indices of the parents set of X_l in Q which may be different from π_l^P, i.e., we can change the graph structure that directed to X_l.

By (3.49), we have $R(Q||P) \leq \eta_l$ for all $Q \in \mathcal{D}_l^{\eta_l}$, then we can consider the predictive uncertainty on $\mathcal{D}_l^{\eta_l}$ which measure and rank the impact of each part of the model in the PGM, $P_{l|\pi_l}$, as

$$I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) = \sup_{Q \in \mathcal{D}_l^{\eta_l}}/\inf \ \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \tag{3.53}$$

Moreover, similarly to Theorem 1, we can show that the predictive uncertainty in this case is also computable with only the baseline model P by the following Theorem:

Theorem 3.4 *Let P be a PGM defined as* (1.1)*, and $f(X_k)$ be a QoI that only depends on X_k. If $f(X_k)$ has finite moment generating function (MGF), $\mathbb{E}_P\left[e^{c\bar{f}(X_k)}\right]$, in a neighborhood of the origin, then for the predictive uncertainty mentioned in* (3.53)*, there exist $0 < \eta_{\pm} \leq \infty$, such that for any $\eta \leq \eta_{\pm}$,*

$$\begin{aligned}
& I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) \\
= \ & \sup_{Q \in \mathcal{D}_l^{\eta_l}}/\inf \ \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\
= \ & \begin{cases} \pm\mathbb{E}_{P_{\rho_l^P}}\left[\inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{l|\pi_l^P}}\left[e^{\pm c\bar{F}(X_l, X_{\rho_l^P})}\right] + \frac{\eta_l}{c}\right]\right] & l \in \rho_k^P \cup \{k\} \\ 0 & l \notin \rho_k^P \cup \{k\} \end{cases} \\
= \ & \mathbb{E}_{Q^{\pm}}\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right]
\end{aligned} \tag{3.54}$$

where ρ_i^P *is the index set of ancestors for* X_i *in* P*,*

$$F(X_l, X_{\rho_l^P}) = \mathbb{E}_{P_{\{k\}|\rho_l^P \cup \{l\}}} [f(X_k)] \tag{3.55}$$

$\bar{F}(X_l, X_{\rho_l^P}) = F(X_l, X_{\rho_l^P}) - \mathbb{E}_{P_{\rho_l^P \cup \{l\}}} \left[F(X_l, X_{\rho_l^P})\right] = F(X_l, X_{\rho_l^P}) - \mathbb{E}_P [f(X_k)]$*, and for* $l \in \rho_k^P \cup \{k\}$*, the probability measures* $Q^\pm$ *are given by*

$$q^\pm(x_i|x_{\pi_i^{Q^\pm}}) \equiv p(x_i|x_{\pi_i^P}) \qquad \textit{for all } i \neq l \textit{ and } \pi_i^{Q^\pm} \equiv \pi_i^P \tag{3.56}$$

and

$$q^\pm(x_l|x_{\pi_l^{Q^\pm}}) = \frac{e^{\pm c_\pm(x_{\rho_l^P}) F(x_l, x_{\rho_l^P})}}{\mathbb{E}_{P_{l|\pi_l^P}} \left[e^{\pm c_\pm(x_{\rho_l^P}) F(X_l, x_{\rho_l^P})}\right]} p(x_l|x_{\pi_l^P}) \quad \textit{for all } x_{\pi_l^{Q^\pm}}\,. \tag{3.57}$$

where $\pi_l^P \subset \pi_l^{Q^\pm} \subset \rho_l^P$ *and* $c_\pm(x_{\rho_l^P})$ *are the unique solutions of*

$$R(P^{c_\pm}_{l|\pi_l^P} || P_{l|\pi_l^P}) = \eta_l \tag{3.58}$$

for all $x_{\rho_l^P}$*.*

Proof of the theorem is shown in Appendix B.

3.2.5 Model-form Sensitivity Analysis, Part 2 – only vary CPD

Furthermore, if we are confident about the causality/connection between all the nodes on the PGM P, we could also consider the domain where the graph structure of alternative models are fixed to be the same as P, i.e., $\pi_l^Q \equiv \pi_l^P = \pi_l$, and investigate the ambiguity set $\mathcal{Q}$ defined by

$$\mathcal{Q} := \mathcal{D}_{l,P}^{\eta_l} = \left\{ \begin{array}{c} \text{all PGM } Q : R(Q_{l|\pi_l} || P_{l|\pi_l}) \leq \eta_l \text{ for all } x_{\pi_l}, \\ Q_{j|\pi_j} \equiv P_{j|\pi_j} \text{ for all } j \neq l \end{array} \right\} \tag{3.59}$$

Then the predictive uncertainty on $\mathcal{D}^{\eta_l}_{l,P}$, i.e.

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta_l}_{l,P}) = \sup/\inf_{Q \in \mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \tag{3.60}$$

can indicate the multiple model/data uncertainties that enter during the learning of the baseline model with given graph structure at each component CPD, and similar to the previous case, it satisfies the following Theorem:

Theorem 3.5 *(a)* ***[Uncertainty Bounds]*** *Let P be a PGM defined as* (1.1)*, and $f(X_k)$ be a QoI that only depends on X_k. If $f(X_k)$ has finite moment generating function (MGF), $\mathbb{E}_P\left[e^{c\bar{f}(X_k)}\right]$, in a neighborhood of the origin, then for the predictive uncertainty defined in* (3.60)*, there exist $0 < \eta_{\pm} \leq \infty$, such that for any $\eta \leq \eta_{\pm}$ and any $Q \in \mathcal{D}^{\eta_l}_{l,P}$, we have*

$$\mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \equiv 0 \qquad \textit{for any } l \notin \rho_k \cup \{k\} \tag{3.61}$$

and

$$\begin{aligned}
I^{+}(f(X_k), P; \mathcal{D}^{\eta_l}_{l,P}) &= \sup_{Q \in \mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\
&\leq \mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{l|\pi_l}}\left[e^{c\bar{F}(X_l, X_{\rho_l})}\right] + \frac{\eta_l}{c}\right]\right] \\
I^{-}(f(X_k), P; \mathcal{D}^{\eta_l}_{l,P}) &= \inf_{Q \in \mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\
&\geq -\mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{l|\pi_l}}\left[e^{-c\bar{F}(X_l, X_{\rho_l})}\right] + \frac{\eta_l}{c}\right]\right]
\end{aligned} \tag{3.62}$$

for any $l \in \rho_k \cup \{k\}$, where

$$F(X_l, X_{\rho_l}) = \mathbb{E}_{P_{\{k\}|\rho_l \cup \{l\}}}\left[f(X_k)\right] \tag{3.63}$$

and $\bar{F}(X_l, X_{\rho_l}) = F(X_l, X_{\rho_l}) - \mathbb{E}_{P_{\rho_l \cup \{l\}}}\left[F((X_l, X_{\rho_l}))\right] = F(X_l, X_{\rho_l}) - \mathbb{E}_P\left[f(X_k)\right]$.
(b) ***[Tightness]*** *If the assumption*

$$F(x_l, x_{\rho_l}) = F(x_l, x_{\pi_l}) \tag{3.64}$$

holds, then there exist probability measures $Q^{\pm} = Q^{\pm}(\eta) \in \mathcal{D}^{\eta_l}_{l,P}$ *such that*

$$\mathbb{E}_{Q^{\pm}}\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \;=\; \sup_{Q\in\mathcal{D}^{\eta_l}_{l,P}}/\inf\ \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \tag{3.65}$$

Furthermore, for $l \in \rho_k^P \cup \{k\}$*, the probability measures* $Q^{\pm}$ *are given by* (3.56) *-* (3.58)*.*

Proof of the theorem is shown in Appendix B.

Remark: The assumption in Step 2 ($F(x_l, x_{\rho_l}) = F(x_l, \pi_l)$) can be satisfied when $p(x_k|x_l, x_{\rho_l}) \equiv p(x_k|x_l, x_{\pi_l})$, or when $\rho_l \cap \rho_i \subset \pi_l$ for all $i \in \rho_k \cup \{k\} \setminus \rho_l \cup \{l\}$. Especially, for all Markov chains, tree structure model, etc... all the nodes in X_{ρ_l} are connected with X_k only through X_l, therefore, given X_l, X_{ρ_l} are independent of X_k, i.e. $p(x_k|x_l, x_{\rho_l}) \equiv p(x_k|x_l)$, so $F(x_l, x_{\rho_l}) = F(x_l, x_{\pi_l})$. Two simple examples where the assumption is satisfied or violated are shown below.

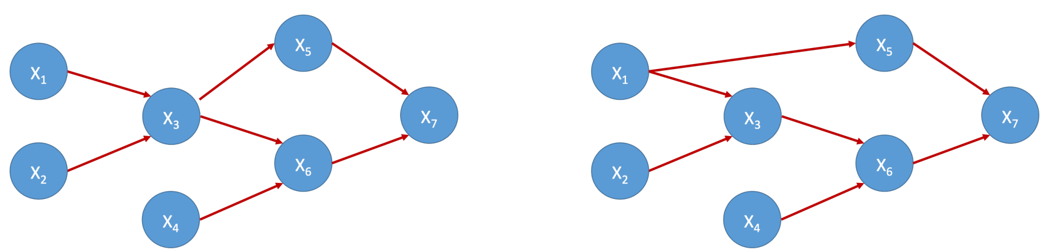

Figure 11. Two examples of the structure of PGM where one (left) could achieve the equality in (3.62) **for** $I^{\pm}(X_7, P; \mathcal{D}^{\eta_6}_{6,P})$**, while the other one (right) could not. For the left PGM, we have** $F = F(x_6, x_3)$**, while** $F = F(x_6, x_1)$ **for the right PGM, therefore, for the optimizer** Q_l^+**,** $\pi_6^Q = \{3, 4\} = \pi_6$ **for the left one, while** $\pi_6^Q = \{3, 4, 1\} \neq \pi_6$ **for the right one.**

Example (Inhomogeneous Markov chains): Again we consider the Markov chain models shown in Figure 8, and the QoI $f(X_k)$, then if we only perturb $P_{l|l-1}$, $l \leq k$, with the constraint $R(Q_{l|\pi_l^Q}||P_{l|l-1}) \leq \eta_l$, i.e. for $Q \in \mathcal{D}^{\eta_l}_l$, where $l \in \rho_k \cup \{k\}$, by Theorem 3.4, we have

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta_l}_l) = \pm\mathbb{E}_{P_{\{l-1\}}}\left[\inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_{P_{l|l-1}}\left[e^{\pm c\bar{F}(X_l, X_{\rho_l})}\right] + \frac{\eta_l}{c}\right]\right]. \tag{3.66}$$

where $F(x_l, x_{\rho_l}) = F(x_l) = \int f(x_k) \prod_{i=l+1}^{k} P(dx_i|x_{i-1})$. Note that $F(x_l, x_{\rho_l}) = F(x_l)$ satisfies the assumption on Theorem 3.5, so we have $I^{\pm}(f(X_k), P; \mathcal{D}_{l,P}^{\eta_l}) = I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l})$, and using (3.56)-(3.58), the optimizer in both Theorem 3.4 and 3.5 is obtained when

$$q^{\pm}(x_i|x_{i-1}) \equiv p(x_i|x_{i-1}) \qquad \text{for all } i \neq l \tag{3.67}$$

and

$$q^{\pm}(x_l|x_{l-1}) = \frac{e^{\pm c_{\pm}(x_{l-1})F(x_l)}}{\mathbb{E}_P\left[e^{\pm c_{\pm}(x_{l-1})F(X_l)}|x_{l-1}\right]} p(x_l|x_{l-1}) \tag{3.68}$$

where $c_{\pm}(x_{l-1})$ are the unique solutions of

$$R(P_{l|l-1}^{c_{\pm}} || P_{l|l-1}) = \eta_l \tag{3.69}$$

for all x_{l-1}. Moreover, if we only perturb $P_{l|l-1}$, $l > k$, with the constraint $R(Q_{l|\pi_l^Q}||P_{l|l-1}) \leq \eta_l$ or $R(Q_{l|l-1}||P_{l|l-1}) \leq \eta_l$, then by Theorem 3.4 and 3.5, we have $I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) = I^{\pm}(f(X_k), P; \mathcal{D}_{l,P}^{\eta_l}) = 0$.

Example (Gaussian Bayesian Networks): Here we consider GBN shown in Figure 9, for the QoI $f(X_4) = X_4$, using Theorem 3.4, 3.5, we conclude that

1. If we only perturb P_3 with the constraint $R(Q_{3|\pi_3^Q}||P_3) \leq \eta_3$ or $R(Q_3||P_3) \leq \eta_3$, i.e. consider $Q \in \mathcal{D}_3^{\eta_3}$ or $\mathcal{D}_{3,P}^{\eta_3}$, then by Theorem 3.4 and 3.5, since the function F in (3.55) satisfies

$$\begin{aligned} F(x_3, x_{\rho_3}) &= \int f(x_4) P(dx_4|x_3, x_2) P(dx_2) \\ &= \beta_{43}x_3 + \beta_{40} + \beta_{42}\beta_{20} \\ &= F(x_3) \end{aligned} \tag{3.70}$$

apply (3.54), we have

$$I^{\pm}(f(X_4), P; \mathcal{D}_3^{\eta_3}) = I^{\pm}(f(X_4), P; \mathcal{D}_{3,P}^{\eta_3}) = \pm|\beta_{43}|\sqrt{2\sigma_3^2\eta_3} \tag{3.71}$$

And by (3.56)- (3.58), the optimizer in both Theorem 3.4 and 3.5 is obtained when

$$\begin{aligned}
q^{\pm}(x_3|x_{\pi_l^{Q^\pm}}) &= \frac{e^{\pm c_\pm F(x_3)}}{\mathbb{E}_{P_3}\left[e^{\pm c_\pm F(X_3)}\right]}p(x_3) \\
&= \frac{e^{\pm c_\pm(\beta_{43}x_3+\beta_{40}++\beta_{42}\beta_{20})}e^{-\frac{(x_3-\beta_{30})^2}{2\sigma_3^2}}}{\int_{x_3} e^{\pm c_\pm(\beta_{43}x_3+\beta_{40}++\beta_{42}\beta_{20})}e^{-\frac{(x_3-\beta_{30})^2}{2\sigma_3^2}}\,dx_3} \\
&= \frac{e^{-\frac{(x_3-\beta_{30}\mp c_\pm\beta_{43}\sigma_3^2)^2}{2\sigma_3^2}}}{\int_{x_3} e^{-\frac{(x_3-\beta_{30}\mp c_\pm\beta_{43}\sigma_3^2)^2}{2\sigma_3^2}}\,dx_3} \\
&= \mathcal{N}\left(\beta_{30}\pm c_\pm\beta_{43}\sigma_3^2,\sigma_3^2\right) \qquad (3.72)
\end{aligned}$$

and

$$R(P^{c_\pm}_{l|\pi_l^P}||P_{l|\pi_l^P}) = \eta_l \qquad \Rightarrow \qquad \pm c_\pm = \pm\sqrt{\frac{2\eta_l}{\beta_{43}^2\sigma_3^2}} \qquad (3.73)$$

so $q^{\pm}(x_3) = \mathcal{N}\left(\beta_{30}\pm\frac{\beta_{43}}{|\beta_{43}|}\sqrt{2\eta_3\sigma_3^2},\sigma_3^2\right)$, and all other components are kept the same, i.e., $q^{\pm}(x_i|x_{\pi_i}) \equiv p(x_i|x_{\pi_i})$ for all $i \neq 2$.

2. if we only perturb P_1 with the constraint $R(Q_{1|\pi_1^Q}||P_1) \leq \eta_1$ or $R(Q_1||P_1) \leq \eta_1$, i.e. consider $Q \in \mathcal{D}_1^{\eta_1}$ or $\mathcal{D}_{1,P}^{\eta_1}$, then by Theorem 3.4 and 3.5, we have

$$I^{\pm}(f(X_4),P;\mathcal{D}_1^{\eta_1}) = I^{\pm}(f(X_4),P;\mathcal{D}_{1,P}^{\eta_1}) = 0 \qquad (3.74)$$

since $1 \notin \rho_4$

Now let us add some connections to the GBN in Figure 9, and consider a more complicated GBN as shown in the left of the Figure 12.

Then for the QoI $f(X_4) = X_4$, if we consider same ambiguity sets as above, i.e., only perturb P_3 with the constraint $R(Q_{3|\pi_3^Q}||P_3) \leq \eta_3$ or $R(Q_3||P_3) \leq \eta_3$, i.e. consider $Q \in \mathcal{D}_3^{\eta_3}$ or $\mathcal{D}_{3,P}^{\eta_3}$, then by Theorem 3.4 and 3.5, since the function F in

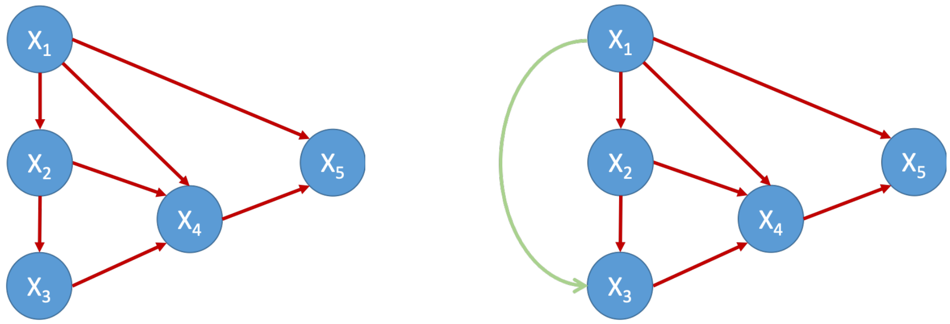

Figure 12. Left: A GBN consists of $X = \{X_1, X_2, \dots, X_5\}$ with $p(x) = p(x_5|x_4, x_1)\ p(x_4|x_3, x_2, x_1)\ p(x_3|x_2)p(x_2|x_1)p(x_1)$ where $p(x_5|x_4, x_1) = \mathcal{N}(\beta_{50} + \beta_{54}x_4 + \beta_{51}x_1, \sigma_5^2)$, $p(x_4|x_3, x_2, x_1) = \mathcal{N}(\beta_{40} + \beta_{43}x_3 + \beta_{42}x_2 + \beta_{41}x_1, \sigma_4^2)$, $p(x_3|x_2) = \mathcal{N}(\beta_{30} + \beta_{32}x_2, \sigma_3^2)$, $p(x_2|x_1) = \mathcal{N}(\beta_{20} + \beta_{21}x_1, \sigma_2^2)$, and $p(x_1) = \mathcal{N}(\beta_{10}, \sigma_1^2)$; Right: The structure of optimizer $Q^{\pm}$ in Theorem 3.4 with QoI $f(X_4)$ and perturbing X_3 based on (3.56) - (3.58). Note that since the function $F(X_3, X_{\rho_3^P})$ in (3.55) may depend on $X_{\rho_3^P}$, so the factor for $q^{\pm}(x_3|x_{\pi_3^{Q^{\pm}}})$ depends on X_2 and X_1, i.e. $\pi_3^{Q^{\pm}} = \{1, 2\}$ by (3.57), which creates a new connection from X_1 to X_3 in $Q^{\pm}$. However, for some special cases like GBN with linear QoI, the graph structure will keep the same, see Corollary 3.6.

(3.55) now is

$$
\begin{aligned}
F(x_3, x_{\rho_3}^P) &= \int f(x_4)P(dx_4|x_3, x_2, x_1) \\
&= \beta_{43}x_3 + \beta_{42}x_2 + \beta_{41}x_1 + \beta_{40} \\
&= F(x_3, x_2, x_1)\ . \qquad (3.75)
\end{aligned}
$$

Thus, for the ambiguity set $\mathcal{D}_3^{\eta_3}$, the optimizer $Q^{\pm}$ would have an extra connection $X_1 \to X_3$ in general by (3.57) as shown in Figure 12 (Right). However, in this case, we have

$$
\begin{aligned}
q^{\pm}(x_3|x_{\pi_l^{Q^\pm}}) &= \frac{e^{\pm c_\pm(x_{\rho_3^P})F(x_3,x_{\rho_3^P})}}{\mathbb{E}_{P_{3|\pi_3^P}}\left[e^{\pm c_\pm(x_{\rho_3^P})F(X_3,x_{\rho_3^P})}\right]}p(x_3|x_{\pi_3^P}) \\
&= \frac{e^{\pm c_\pm(x_2,x_1)F(x_3,x_2,x_1)}}{\mathbb{E}_{P_{3|\{2\}}}\left[e^{\pm c_\pm(x_2,x_1)F(X_3,x_2,x_1)}\right]}p(x_3|x_2) \\
&= \frac{e^{\pm c_\pm(x_2,x_1)(\beta_{43}x_3+\beta_{42}x_2+\beta_{41}x_1+\beta_{40})}e^{-\frac{(x_3-\beta_{30}-\beta_{32}x_2)^2}{2\sigma_3^2}}}{\int_{x_3} e^{\pm c_\pm(x_2,x_1)(\beta_{43}x_3+\beta_{42}x_2+\beta_{41}x_1+\beta_{40})}e^{-\frac{(x_3-\beta_{30}-\beta_{32}x_2)^2}{2\sigma_3^2}}\,dx_3} \\
&= \frac{e^{-\frac{(x_3-\beta_{30}-\beta_{32}x_2\mp c_\pm(x_2,x_1)\beta_{43}\sigma_3^2)^2}{2\sigma_3^2}}}{\int_{x_3} e^{-\frac{(x_3-\beta_{30}-\beta_{32}x_2\mp c_\pm(x_2,x_1)\beta_{43}\sigma_3^2)^2}{2\sigma_3^2}}\,dx_3} \\
&= \mathcal{N}\left(\beta_{30}+\beta_{32}x_2 \pm c_\pm(x_2,x_1)\beta_{43}\sigma_3^2, \sigma_3^2\right) \qquad (3.76)
\end{aligned}
$$

then by (3.58),

$$
R(P^{c_\pm}_{l|\pi_l^P}||P_{l|\pi_l^P}) = \eta_l \qquad \Rightarrow \qquad \pm c_\pm(x_2,x_1) = \pm\sqrt{\frac{2\eta_l}{\beta_{43}^2\sigma_3^2}} \qquad (3.77)
$$

so $c_\pm(x_2,x_1)$ does not depend on X_1, X_2, and we have $x_{\pi_l^{Q^\pm}} \equiv x_{\pi_l^P} = \{2\}$, $Q^\pm$ have same graph structure as P. And apply (3.54), we still have

$$
I^{\pm}(f(X_4),P;\mathcal{D}_3^{\eta_3}) = I^{\pm}(f(X_4),P;\mathcal{D}_{3,P}^{\eta_3}) \;=\; \pm|\beta_{43}|\sqrt{2\sigma_3^2\eta_3} \qquad (3.78)
$$

In general, we can conclude the result by the following Corollary for this special case in GBN:

Corollary 3.6 *Let P be a GBN satisfies* (3.29)*, and $f(X_k) = aX_k + b$ be a QoI only depends on X_k linearly. Then for the predictive uncertainties defined in* (3.53) *and* (3.60)*, we have*

$$
I^{\pm}(f(X_k),P;\mathcal{D}_l^{\eta_l}) \equiv I^{\pm}(f(X_k),P;\mathcal{D}_{l,P}^{\eta_l}) \qquad (3.79)
$$

and the optimizer $Q^\pm = Q^\pm(\eta) \in \mathcal{D}_{l,P}^{\eta_l} \subset \mathcal{D}_l^{\eta_l}$ given by (3.56) - (3.58) *are also GBNs with same graph structure as P. Furthermore, for $l \in \pi_k^P$ and $l \notin \rho_{\pi_j}^P$ for all $j \in \pi_k$,*

$j \neq l$, we have

$$I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) = \pm|\beta_{kl}|\sqrt{2\sigma_l^2\eta_l} \tag{3.80}$$

Moreover, for any $l \in \rho_k^P$, we also have

$$I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) = \pm|\tilde{\beta}_{kl}|\sqrt{2\sigma_l^2\eta_l} \tag{3.81}$$

for some constant $\tilde{\beta}_{kl}$. For example, if we perturb P_1 in Figure 12 with the constraint $R(Q_{1|\pi_1^Q}||P_1) \leq \eta_1$ or $R(Q_1||P_1) \leq \eta_1$, i.e. consider $Q \in \mathcal{D}_1^{\eta_1}$ or $\mathcal{D}_{1,P}^{\eta_1}$, since the function F in (3.55) now is

$$\begin{aligned} F(x_1) &= \int f(x_4)P(dx_4|x_3, x_2, x_1)P(dx_3|x_2)P(dx_2|x_1) \\ &= (\beta_{43}\beta_{32}\beta_{21} + \beta_{42}\beta_{21} + \beta_{41})x_1 + \beta_{40} + \beta_{43}\beta_{30} + \beta_{43}\beta_{32}\beta_{20} + \beta_{42}\beta_{20}. \end{aligned} \tag{3.82}$$

then by Theorem 3.4, 3.5 and (3.54), we can conclude that

$$I^{\pm}(f(X_4), P; \mathcal{D}_1^{\eta_1}) = I^{\pm}(f(X_4), P; \mathcal{D}_{1,P}^{\eta_1}) = \pm|\beta_{43}\beta_{32}\beta_{21} + \beta_{42}\beta_{21} + \beta_{41}|\sqrt{2\sigma_1^2\eta_1} \tag{3.83}$$

3.2.6 Model-form UQ and SA indices

Here we summarize all the results above and define the corresponding indices for model-form UQ and SA as following:

- **model-form UQ indices**

  we define the model-form UQ indices of the PGM P for the QoI $f(X_k)$, $1 \leq k \leq n$, by $I^{\pm}(f(X_k), P; \mathcal{D}^{\eta})$, i.e., we consider the the worst case scenarios in the ambiguity set $\mathcal{D}^{\eta}$ which contains all possible models Q with the aforementioned model misspecification η, then based on Theorem 3.2,

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta}) = \pm \inf_{c>0} \left[\frac{1}{c}\log \mathbb{E}_{P_{\{k\}}}\left[e^{\pm c\bar{f}(X_k)}\right] + \frac{\eta}{c}\right] \tag{3.84}$$

where $\bar{f}(X_k)$ is the centered QoI, $\bar{f}(X_k) := f(X_k) - \mathbb{E}_P[f(X_k)]$.

Remark: we can also use the UQ indices $I^{\pm}$ to define the **relative predictive uncertainty**, i.e., the relative error

$$\frac{I^{+}(f(X_k), P; \mathcal{D}^{\eta})}{|\mathbb{E}_P[f(X_k)]|} \tag{3.85}$$

which captures the uncertainty of the nominal model P within the family of models $Q \in \mathcal{D}^{\eta}$ for QoI $f(X_k)$.

- **model-form sensitivity indices 1**

we define the model-form sensitivity indices, which measure and rank the impact of each part of the model in the PGM, $P_{l|\pi_l}$, by $I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l})$ as discussed in Section 3.2.3, i.e., we consider the worst case scenarios in the ambiguity set $\mathcal{D}_l^{\eta_l}$ where we perturb the CPD and parents of node X_l with model misspecification η_l, then based on the results shown on Theorem 3.4,

$$\begin{aligned} & I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) \\ = & \begin{cases} \pm\mathbb{E}_{P_{\rho_l^P}}\left[\inf_{c>0}\left[\frac{1}{c}\log\mathbb{E}_{P_{l|\pi_l^P}}\left[e^{\pm c\bar{F}(X_l, X_{\rho_l^P})}\right] + \frac{\eta_l}{c}\right]\right] & l \in \rho_k^P \cup \{k\} \\ 0 & l \notin \rho_k^P \cup \{k\} \end{cases} \end{aligned} \tag{3.86}$$

where

$$F(X_l, X_{\rho_l^P}) = \mathbb{E}_{P_{\{k\}|\rho_l^P \cup \{l\}}}[f(X_k)] \tag{3.87}$$

and $\bar{F}(X_l, X_{\rho_l^P}) = F(X_l, X_{\rho_l^P}) - \mathbb{E}_{P_{\rho_l^P \cup \{l\}}}\left[F((X_l, X_{\rho_l^P}))\right] = F(X_l, X_{\rho_l^P}) - \mathbb{E}_P[f(X_k)]$.

- **model-form sensitivity indices 2**

we can also define the an alternative model-form sensitivity indices by $I^{\pm}(f(X_k), P; \mathcal{D}_{l,P}^{\eta_l})$ for which we consider the worst case scenarios in the ambiguity set $\mathcal{D}_{l,P}^{\eta_l}$, i.e., we still perturb the CPD of node X_l with model misspecification η_l but with the constraint that the parent set π_l is fixed, so is the graph structure of the PGM, then by Theorem 3.5, when P satisfies the assumption $F(x_l, x_{\rho_l}) = F(x_l, x_{\pi_l})$ where

$$F(X_l, X_{\rho_l}) = \mathbb{E}_{P_{\{k\}|\rho_l \cup \{l\}}} [f(X_k)] \tag{3.88}$$

we have

$$\begin{aligned} & I^{\pm}(f(X_k), P; \mathcal{D}^{\eta_l}_{l,P}) \\ = & \begin{cases} \pm\mathbb{E}_{P_{\rho_l}} \left[\inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{l|\pi_l}} \left[e^{\pm c\bar{F}(X_l, X_{\rho_l})}\right] + \frac{\eta_l}{c}\right]\right] & l \in \rho_k \cup \{k\} \\ 0 & l \notin \rho_k \cup \{k\} \end{cases} \end{aligned} \tag{3.89}$$

where

$$\bar{F}(X_l, X_{\rho_l}) = F(X_l, X_{\rho_l}) - \mathbb{E}_{P_{\rho_l \cup \{l\}}} [F((X_l, X_{\rho_l}))] = F(X_l, X_{\rho_l}) - \mathbb{E}_P [f(X_k)] \ .$$

Furthermore, note that $I^{\pm}(f(X_k), P; \mathcal{D}^{\eta_l}_{l,P}) = I^{\pm}(f(X_k), P; \mathcal{D}^{\eta_l}_{l})$ when P satisfies the assumption in Theorem 3.5 (b).

Note that all the indices we defined are bounds for the PGMs in infinite dimensional spaces, but they are computable (with some conditions) by a one dimensional optimization problem based on the Theorems we list above.

Remark [On the choice of KL divergence]: Given the abundance of different distances and pseudo-distances for probability models besides the Kullback-Leibler divergence, it is reasonable to wonder if any other such metrics or divergences (e.g. Wasserstein, χ^2, total variation, Hellinger, etc) can be used in place of Kullback-Leibler (KL) in the definition of the non-parametric family (3.52), (3.59), and the sensitivity index (3.86). It turns out that the choice of the KL divergence in the present work is crucial in obtaining computable sensitivity index (3.86). Indeed, in Section 3.2.2, we demonstrate that the derivation of (3.86) relies on taking advantage of the chain rule for the KL divergence, [20]. More specifically, we break down the calculation of any KL distance between different PGM models, in terms of conditional KL divergences between separate PGM nodes, i.e. CPDs $p(x_i|x_{\pi_i})$ in (1.1), see (B3-38). It is also this property of the Kullback-Leibler divergence that allows us to isolate the uncertainty impact on QoIs from multiple PGM components and data sources. The lack of such a decomposition property in other probabilistic metrics and divergences and its significance for UQ calculations is demonstrated in special cases of PGMs such as Markov Chains and Markov Random Fields (e.g. Boltzmann/Gibbs distributions), in [64].

3.3 How To Pick The Misspecification Parameters in PGMs?

Here we consider two perspectives in setting up the model misspecification parameters η/ η_j in the indices $I^{\pm}(f(X_k), P; \mathcal{D}^{\eta})$ or $I^{\pm}(f(X_k), P; \mathcal{D}_j^{\eta_j})$: (a) a fixed constant $\eta > 0$: for the UQ indices (as in Theorem 3.2) or the sensitivity indices (as in Theorem 3.4 and 3.5), we can consider perturbing the whole model P or each part of model, $P_{i|\pi_i}$, with the same amount of "distance" η, as "stress test", then comparing the indices $I(f(X_k), P; \mathcal{D}_j^{\eta_j})$ will give us a ranking of the impact of each component on the model. (b) Computed from data: we can also consider the η by the "distance" between data and the PGM P, where data is represented by a histogram or a KDE approximation of the histogram, or any given particular model Q from data or expert knowledge. In this case, we can estimate η_j values constitute *surrogates* for the distance of the baseline model from the unknown "real" model. And it may be different for different components or different given conditions.

By the chain rule of KL divergence [20], $\eta := R(Q||P)$ can be computed by

$$\begin{aligned}\eta &= \int \log \frac{dQ}{dP} dQ = \sum_{i=1}^{n} \mathbb{E}_Q \left[R(Q_{i|\pi_i}||P_{i|\pi_i}\right] dx_i \\ &= \sum_{i=1}^{n} \mathbb{E}_Q \left[\eta_i^{\pi_i}\right] \end{aligned} \tag{3.90}$$

where

$$\eta_i^{\pi_i} = \int \log \frac{Q_{i|\pi_i}}{P_{i|\pi_i}} Q_{i|\pi_i} dx_i \tag{3.91}$$

with given x_{π_i}.

Examples: For a Gaussian Bayesian network where $p(x_i|x_{\pi_i})$ satisfies $p(x_i|x_{\pi_i}) = \mathcal{N}(\beta_{i0} + \beta_i^T x_{\pi_i}, \sigma_i^2)$ for some β_{i0}, β_i, and σ_i^2, i.e.,

$$P_{i|\pi_i}: \quad X_i = \beta_{i0} + \beta_i^T X_{\pi_i} + \epsilon_i \tag{3.92}$$

where ϵ_i is a random variable with density $p_{\epsilon_i}(x) = \mathcal{N}(0, \sigma_i^2)$ which comes from fitting data with Maximum-Likelihood-Estimation. Then we consider alternative models to P such as

$$Q_{i|\pi_i}: \quad X_i = \beta_{i0} + \beta_i^T X_{\pi_i} + \tilde{\epsilon}_i \tag{3.93}$$

where $\tilde{\epsilon}_i$ follows another approximate distribution of the data with density $q_{\tilde{\epsilon}_i}(x)$, for instance any histogram or KDE. Therefore, for given x_{π_i}, we have

$$\begin{aligned}\eta_i^{\pi_i} &= \int \log \frac{q(x_i|x_{\pi_i})}{p(x_i|x_{\pi_i})} q(x_i|x_{\pi_i}) dx_i \\ &= \int \log \frac{q(x_i - \beta_{i0} - \beta_i^T x_{\pi_i}|x_{\pi_i})}{p(x_i - \beta_{i0} - \beta_i^T x_{\pi_i}|x_{\pi_i})} q(x_i - \beta_{i0} - \beta_i^T x_{\pi_i}|x_{\pi_i}) dx_i \\ &= \int \log \frac{q_{\tilde{\epsilon}_i}(x)}{p_{\epsilon_i}(x)} q_{\tilde{\epsilon}_i}(x) dx\,, \end{aligned} \tag{3.94}$$

thus, we have that $\eta_i^{\pi_i}$ is independent of π_i; in fact, we have

$$\eta_i^{\pi_i} \equiv \eta_i = \int \log \frac{q_{\tilde{\epsilon}_i}(x)}{p_{\epsilon_i}(x)} q_{\tilde{\epsilon}_i}(x) dx\,, \tag{3.95}$$

Therefore we can consider the estimation of model misspecification based on (3.95) with $Q_{\tilde{\epsilon}_i}$ as the histogram, i.e.,

$$q^{hist}_{\tilde{\epsilon}_i}(x) = \sum_{k=1}^{m} \frac{\nu_k}{nh} I(x \in B_k)\,, \tag{3.96}$$

where $B_1, \ldots, B_m$ are the histogram bins, h is the bin width, n is the number of observations and ν_k is the number of observations in B_k. Alternatively, we can consider the model $Q_{\tilde{\epsilon}_i}$ given by a kernel density estimator (KDE) viewed here as a high resolution but smooth approximation of the histogram, namely

$$q^{KDE}_{\tilde{\epsilon}_i}(x) = \sum_{k=1}^{n} \frac{1}{nh} K(\frac{x - x_i}{h})\,, \tag{3.97}$$

where $K(\cdot)$ is the normal kernel smoothing function with bin width h, $(x_1, \ldots, x_n)$ are the samples of ϵ_i. Similarly, we can consider other KDE kernels, [131], or any other probabilistic representations of the data in the histogram. It can be shown using the weak continuity properties of the KL divergence, [27], that $R(Q_{\tilde{\epsilon}_i}||P_{\epsilon_i})$ will converge to $R(Q_{\epsilon_i}||P_{\epsilon_i})$ in the large data limit, where Q_{ϵ_i} is the real distribution of ϵ_i, for more general results we also refer to [105].

3.4 Model Selection and Correctability

3.4.1 Model selection based on model-form UQ indices

Based on the predictive uncertainty indices (3.84), we intend to develop a new class of Information Criteria (IC) for model selection & evaluation that *include in the selection process specific QoIs of engineering interest*. In the existing AI literature, IC such as Akaike IC and Bayesian IC, are deployed for model selection tasks, [66], but do not take into consideration QoIs. Therefore, we propose to: (a) use the predictive uncertainty indices (3.84) to evaluate the predictive ability of different models; (b) compare and optimize model selection by minimizing the predictive uncertainty indices (3.84), where η is calculated as in typical AIC/BIC methods, [66, 9] as the distance between model and available data .

In order to explain the key idea and the main difference between existing IC methods that do not take into account QoIs $f(X_k)$, let us consider the linearization of the predictive uncertainty indices (3.84), [29],

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta}) = \pm\sqrt{2Var_P(f(X_k))}\eta^{1/2} + O(\eta) \tag{3.98}$$

where P is the baseline PGM model, and when P is a Gaussian network, the above expansion is exact, [50]. It is evident that (3.98) has both information-theoretic aspects as in standard IC via the KL term η, and also includes the engineering QoI $f(X)$ via the variance term.

3.4.2 Model improvement based on model-form sensitivity indices

We could also consider the uncertainty of each component on the PGM separately by the model-form sensitivity indices (3.86), then with a *desired tolerance* $TOL \in (0, 1)$ for predictive uncertainty, i.e. selecting a model P such that

$$\frac{I^{+}(f(X_k), P; \mathcal{D}_l^{\eta_l})}{\mathbb{E}_P\left[f(X_k)\right]} \leq TOL \qquad \text{for all PGM nodes } l\,. \tag{3.99}$$

we can improve the selection of a baseline model P as follows.

Step 1: Find data-based surrogates η_l's using for instance the approach in (3.91), or more generally:

$$\eta_l = \sup_{x_{\pi_l}} R(Q_{l|\pi_l}||P_{l|\pi_l})$$

where Q is the surrogate model given by KDE/histogram.

Step 2: Calculate the model-form sensitivity indices (3.86):

$$I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) \qquad \text{for all } l$$

with given QoI $f(X_k)$, and find the most uncertain component,

$$l^* = \underset{l}{\operatorname{argmax}}\ I^+(f(X_k), P; \mathcal{D}_l^{\eta_l})$$

Step 3: Determine whether the relative predictive uncertainty of P_{l^*} is within a given tolerance level $TOL \in (0, 1)$, i.e. satisfying (3.99) and thus

$$\frac{I^+(f(X_k), P; \mathcal{D}_{l^*}^{\eta_{l^*}})}{\mathbb{E}_P\left[f(X_k)\right]} \leq TOL\,. \tag{3.100}$$

Step 4: If (3.100) is not true, reduce $I^+(f(X_k), P; \mathcal{D}_{l^*}^{\eta_{l^*}})$ based on (3.86) and (3.98), i.e., we could consider decreasing $\mathrm{Var}_{P_{l^*|\pi_{l^*}}}(F(X_l, X_{\rho_l}))$ or η_{l^*} by acquiring more data for l^* or updating the CPD.

Note that, based on (3.86), the indices depend on all the CPDs on PGM (in general, for node l, the $\mathbb{E}_{P_{\rho_l}}[\cdot]$ part in the sensitivity indices may depend on all the CPDs of l's ancestors, and $F(X_l, X_{\rho_l})$ part may depend on the CPDs of all the other nodes), so if we decrease the uncertainty of l^* component by updating the CPD $P_{l^*|\pi_{l^*}}$, the indices for other components may increase. However, if the mean model of $f(X_k)$ does not change, i.e. $\bar{F}(X_l, X_{\rho_l})$ is fixed for all l when we improve the model, then updating $P_{l^*|\pi_{l^*}}$ would only affect the descendant components of l^*, therefore, we could make all the components satisfied (3.100) via the loop shown above.

Example: Consider a GBN defined as (3.29), where

$$X_i = \beta_{i0} + \beta_i^T X_{\pi_i} + \epsilon_i \tag{3.101}$$

with $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, if we only update the CPD $p(x_i|x_{\pi_i})$ by changing the distribution of ϵ_i from Gaussian to other mean zero distribution, then for $f(X_k) = X_k$, $1 \leq k \leq n$, we have

$$F(X_l, X_{\rho_l}) \equiv \tilde{\beta}_{kl}^T (X_l, X_{\rho_l})^T + \tilde{\beta}_{k0} \tag{3.102}$$

for some constants $\tilde{\beta}_{kl}$, $\tilde{\beta}_{k0}$, i.e., $\bar{F}(X_l, X_{\rho_l})$ is fixed for all l. In fact, since we know

$$I^+(f(X_k), P; \mathcal{D}_l^{\eta_l}) = |\beta_{kl}|\sqrt{2\sigma_l^2\eta_l} \tag{3.103}$$

by (B3-51), updating the CPD of any component l would only change the value of $I^+(f(X_k), P; \mathcal{D}_l^{\eta_l})$ if we fix the correlation between all the nodes (i.e., all the β_l's). Moreover, since we have

$$I^+(f(X_k), P; \mathcal{D}^{\eta}) = \sqrt{2\mathcal{C}_{kk}\eta} = \sqrt{2(\sum_{i\in\rho_k}\tilde{\beta}_{ki}^2\sigma_i^2)(\sum_{j=1}^{n}\eta_j)} \tag{3.104}$$

see (3.41) and (3.31) for example, so the model-form UQ index for the whole model will also be decreased when we decrease the model-form sensitivity index for component l either by decreasing σ_l^2 or η_l.

CHAPTER 4

PGMS IN CHEMISTRY: AN APPLICATION ON OXYGEN REDUCTION REACTION

4.1 Towards AI Chemistry: Causality, PGMs & Multi-scale Modeling

4.1.1 From the computational chemistry towards the AI chemistry

Computational Chemistry powered by groundbreaking developments in scientific computing and sophisticated multi-scale modeling from the quantum scale and up, has provided in the last years unprecedented new insights in areas ranging from chemical sciences, to materials and biology. However, in order to become truly predictive, reliable and robust enough to perform design and optimization tasks, these models still need to incorporate heterogeneous and multiscale data, e.g. electronic structure calculations, experimental data from the mesoscale or the device/engineering scale, highly correlated time series data, and so on. Furthermore, this statistical learning process needs to account for varying degrees of expert knowledge, e.g. some parts of a physico-chemical model may be less well-accepted or understood than others; data that are not easily collected manually and need to be retrieved from the literature; physical constraints, correlations and intrinsically causal relationships between model components such as parameters, mechanisms, input/output relationships and different quantities of interest. To this end, existing and potential new developments in Data Science methods such as approximate inference, probabilistic & causal networks, reinforcement learning, information retrieval, and UQ, need to be fused with Applied Mathematics and Computational Chemistry methods for multi-scale/-physics models, in order to advance the field towards full, predictive Artificial Intelligence (AI) for Chemistry, capable to first learn efficiently networks of multi-scale models based on imperfect and heterogeneous data and expert knowledge, and second, to close the experiment/data/model loop, i.e. continuously improve data and model selection towards enhancing predictive and robust design & optimization capabilities under uncertainty.

4.1.2 Probabilistic graphical modeling for chemistry

We started working in some of these directions in our recent work [125]; there we identified the importance of correlations in model parameters/reactions towards building more predictive chemical kinetics models; we also developed the necessary new UQ and non-parametric statistics methods to assess predictive capability in the presence of strong correlations, [33]. However here we want to move beyond correlations and build full causal models from available heterogeneous data and expert knowledge, and importantly, along with predictive guarantees. Finally we seek strategies to improve such models, i.e. their predictive guarantees as quantified here, by targeting with more data or improved modeling any under-performing components of our model. Our mathematical formulations rely on Probabilistic Graphical Models and new associated Uncertainty Quantification methods suitable for graphical models build on sparse and heterogeneous data.

In this chapter, we apply PGMs as models that can provide the *mathematical* *foundation* for *AI in Computational Chemistry*: here, correlations in space/time and between model elements (molecules, parameters, mechanisms), causal relationships between inputs and outputs/QoIs, couplings between scales and physics (from quantum to meso/macro-scale) are typically poresent and thus necessary in building complete, predictive models. In this direction, we intend to build PGM-based AI models for both modeling and design in physico-chemical systems. This class of proposed *Chemistry PGMs*, and in particular the proposed class of *Chemical Bayesian Networks*, allows us to combine expert knowledge (e.g. from multi-scale/multi-physics modeling), computational and experimental data, along with Uncertainty Quantification, Machine Learning and Information Theory to obtain mathematical and computational models with predictive guarantees. Finally the proposed Uncertainty Quantification (UQ) methods for Chemical PGMs allow for systematic strategies for model evaluation and adaptive model improvement.

4.1.3 Modeling and uncertainties in Oxygen Reduction Reaction

Due to the 100-fold higher energy density of fuels fuel cells are superior to batteries; they provide more power at lower weights, smaller volumes, and do not suffer from recharging challenges [99]. The hydrogen fuel cell is a mature technology that produces electricity via the Hydrogen Oxidation Reaction (HOR) at the anode and the Oxygen Reduction Reaction (ORR) at the cathode, see Figure 13(b). Polymer electrolyte membrane fuel cells are commercially available [39]. Due to the high cost of platinum (Pt) catalysts and stability problems of other materials in an acidic electrolyte, recent focus has been on developing alkaline electrolytes. This technology (see Figure 13(b)), while extremely promising, results in slower reaction rates (by ~ 2 orders of magnitude compared to Pt/acidic electrolyte) and thus for a need for bigger devices for the same performance, [117, 30]. Overcoming this slower-rate challenge requires discovery of new, multicomponent, e.g., core-shell alloy, catalysts.

Our objective is to demonstrate our new modeling paradigm via the use of probabilistic AI and PGMs on this important problem. The physical model we consider here is simple in order to enable mathematical analysis while obeying real constraints such as thermodynamics, real reactions, reaction stoichiometry, mass conservation, etc. The ORR reaction depends on the formation of surface hydroperoxyl (OOH^*) from molecular oxygen (O_2), and water (H_2O) from surface hydroxide (OH^*) [124]. The complete mechanism [13, 4, 61] involves four electron steps, see Figure 13(a). Among these, reactions R1 and R4 are slow [13]. Acceleration of ORR then translates into discovering materials that speed up the slower of R1 and R4. An approach to discovering new materials entails use of models to generate activity plots, see Figure 13(c), as a function of descriptor(s) whose properties can be generated quickly from quantum mechanical calculations, [113].

We compute the rate using a thermodynamic model based on the minimum free energy of reactions R1 and R4, i.e., $rate = \exp(-\max[\Delta G_1, \Delta G_4]/k_B T)$, where k_B is the Boltzmann constant and T is the temperature. The Gibbs free energy ΔG^f of a species is estimated from the electronic energy (EDFT) obtained using density functional theory (DFT), and corrected for both solvation (Esolv) in water and for temperature effects. Upon computing the formation free energies of $O*$, OOH^*, and OH^* on different monometallic catalysts, the free energies ΔG_1 and ΔG_4 are computed as linear combinations of free energies of species and are regressed vs. $\Delta G^f_{O^*}$ (the descriptor); see data in Figure 13(c). The intersection of the two lines (see Figure 13(c)) determines the max of the volcano curve and provides optimal material properties, i.e., the $\Delta G^f_{O^*}$, which can then be matched to those of multicomponent materials to maximize the rate. This approach was originally introduced

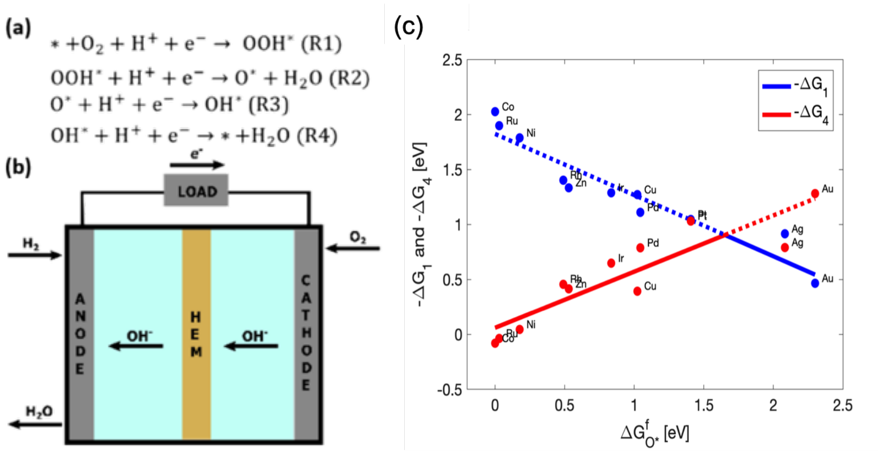

Figure 13. **(a) Key reaction steps (R1-R4) in alkaline fuel cells. R1: solvated O_2 forms adsorbed OOH^*; R2: OOH^* forms adsorbed surface oxygen O^* and solvated H_2O; R3: O^* forms adsorbed OH^*; R4: H_2O forms and regenerates the free catalyst site. * represents an unoccupied metal site and next to a species, e.g., OOH^*, an adsorbed species; $H+$ and $e-$ refer to proton and electron. (b) Schematic of an alkaline fuel cell. (c) Negative changes in Gibbs energies for reactions R1 and R4: OOH adsorption (blue) and OH desorption (red). The optimal $\Delta G^f_{O^*}$ is the intersection of the two lines. Shown are both DFT data on various metals (circles) and lines from linear regressions. The function given by $\min(-\Delta G_1, -\Delta G_4)$, corresponding to the rate, is indicated by the solid lines and is referred to in the literature as a "volcano curve"**

to discover a highly active Ni-Pt bimetallic for decomposition of ammonia, [58].

However, due to incomplete available data, expensive to compute quantities with quantum mechanical simulations, sparse data, lack of a full expert-knowledge library, and lack of quantified errors, the prediction of model accuracy and identifying under-performing components are impossible under a deterministic model. Therefore, we generate DFT data to estimate free energies (see Figure 13(c)), estimate error distributions, account for expert knowledge. Overall, we develop a workflow to account for errors, and build the first corresponding PGM (see Figure 16) that opens up the door for Probabilistic AI in Chemistry. More specifically, errors (see Figure 16 and Table 3) exist in experiments (ω_{ei}), DFT (ω_{di}), solvation energies (ω_{si}), and regressions (correlations) are used to determine the optimum ΔG^f_{O*} (ω_{ci}), a problem accentuated by the relatively sparse data available. Experimental errors (ω_{ei}) in ΔG^f_{O*} and ΔG^f_{OH*} arise from repeated measurements in (1) the same and (2) different labs. Repeated calorimetry and temperature-programmed desorption measurements for the dissociative adsorption enthalpy of O_2 will provide a distribution of errors for ΔG^f_{O*}. The distribution of DFT errors (ω_{di}) will be computed by comparing experimental and calculated (DFT) data across various metals. The ω_{si} distribution is estimated by simulating several hundred explicit water molecules using ab initio molecular dynamics. Multiple modeling choices are dictated by expert knowledge: for example, we choose O^* as a descriptor because it has the fewest local minima on a potential energy surface for faster quantum calculations. Because errors are independent, we will add their contributions in a linear manner. Furthermore, because the correlation of ΔG_1 and ΔG_4 with ΔG^f_{O*} captures the majority of the correlation of ΔG_1 and ΔG_4 with each other, it is safe to assume conditional independence for their respective probability distributions.

4.1.4 Structure and model parameter learning for the ORR PGM.

For structure (graph) learning of the ORR PGM, we use a constraint-based method [123] taking advantage of expert knowledge, in this case, multi-scale, micro-kinetic modeling and related causal relations.

Using the DFT computed data shown in Figure 14, through the statistical dependency test [130], we know both y_1 and y_2 depend on x and they are conditionally independent given x; x, y_1, and y_2 are shown in Figure 16. Therefore, we build part of the network structure with x, y_1, y_2 using a constraint-based method [123], which selects a desired structure based on constraints of dependency among variables. Subsequently, we add other nodes, ω_i's which represent different types of errors associated with statistical modeling, experiment, solvation, etc, using also any dependencies known from expert knowledge. Finally, we add the QoIs, x^*_O and r^*_O, whose evaluations depend on the values of y_i for each x_0 due to physics knowledge, see Figure 19. Overall, we combine available data and expert "physicochemical" knowledge to build the structure, see Figure 15.

Therefore, based on the discussion above, for the random variables $X_{1:n}$ taking values $X_{1:n} = x_{1:n}$ where

$$x_{1:n} = \{x, y_1, y_2, \omega_{e0}, \omega_{d0}, \omega_{s0}, \omega_{e1}, \omega_{d1}, \omega_{s1}, \omega_{c1}, \omega_{e2}, \omega_{d2}, \omega_{s2}, \omega_{c2}\},$$

and where the entries are defined in Table 3, the PGM corresponds to a Directed Acyclical Graph (DAG) and is defined as

$$\begin{aligned} & p(x, y_1, y_2, \omega_{e0}, \omega_{d0}, \omega_{s0}, \omega_{e1}, \omega_{d1}, \omega_{s1}, \omega_{c1}, \omega_{e2}, \omega_{d2}, \omega_{s2}, \omega_{c2} | x_0) \\ = & \prod_{i=1,2} p(y_i | x, \omega_{ei}, \omega_{di}, \omega_{si}, \omega_{ci}) \cdot p(x | \omega_{e0}, \omega_{d0}, \omega_{s0}, x_0) \cdot \prod_{\substack{j=e_k,d_k,s_k,c_1,c_2 \\ k=0,1,2}} p(\omega_j) \end{aligned} \tag{4.1}$$

where

$$y_i = \beta_{y_i,0} + \beta_{y_i,x} x + \omega_{ei} + \omega_{di} + \omega_{si} + \omega_{ci} \tag{4.2}$$

for $i = 1, 2$ and

$$x = x_0 + \omega_{e0} + \omega_{d0} + \omega_{s0} \tag{4.3}$$

Once we have obtained the structure of the graph from the previous step, we then learn the model

$$p(x|\theta) = \prod_i x(x_i | x_{\pi_i}, \theta_{i|\pi_i})$$

in the following steps:

(a) First, we select a parametric family for models $p(x_i | x_{\pi_i}, \theta_{i|\pi_i})$. For the ORR example, we select as our parametric family of PGMs, a family of Gaussian Bayesian Networks (GBN).

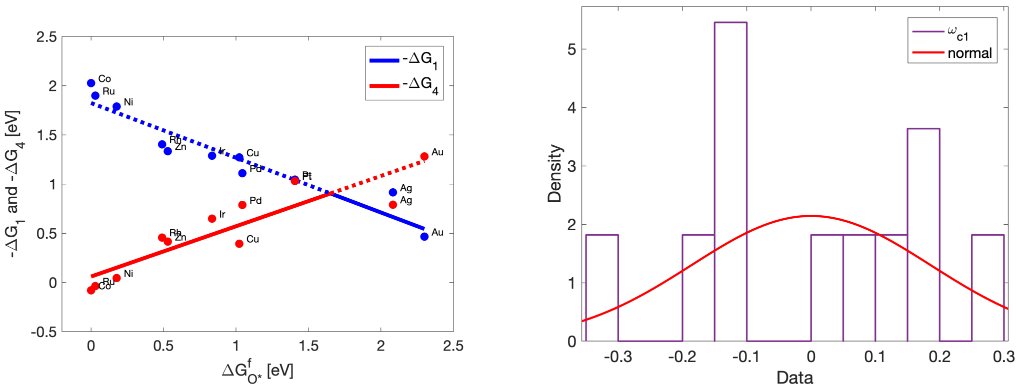

Figure 14. (L): DFT-computed data for reaction energies with respect to different metals/oxygen binding energies, which is used for structure learning with respect to x and y_i as shown in Figure 16. (R): Data representing the error in correlation/linear regression, used for parameter learning of ω_{c1} in Figure 16 by means of Maximum Likelihood, see (4.8)**.**

Gaussian Bayesian Networks (GBN), [66], is a special class of Probabilistic Graphical Models commonly used in natural and social sciences and where the CPDs (1.2) are linear and Gaussian. More specifically, for a GBN consisting of variables $X = X_{1:n}$, every node X_i is a linear Gaussian of its parents, i.e.,

$$p(x_i|x_{\pi_i}, \theta_{i|\pi_i}) = \mathcal{N}(\beta_{i0} + \beta_i^T x_{\pi_i}, \sigma_i^2)\,, \tag{4.4}$$

where $\theta_{i|\pi_i} = (\beta_{i0}, \beta_i, \sigma_i)^T$ for some constants β_{i0}, $\beta_i = (\beta_{i,i_1}, \ldots, \beta_{i,i_m})$, and variance σ_i which does not depend on X_{π_i}. Then by the conjugacy properties of Gaussians, the joint distribution in (1.1) becomes $p(x|\theta) = \mathcal{N}(\mu, \mathcal{C})$, i.e. it is also a Gaussian with parameters μ, $\mathcal{C}$, which can be calculated from β_{i0}, β_i, and σ_i (for more details and derivations see Appendix B).

So for the ORR example, the corresponding CPDs can be defined as,

$$p(y_i|x, \omega_{ei}, \omega_{di}, \omega_{si}, \omega_{ci}) = \mathcal{N}(\beta_{y_i,0} + \beta_{y_i,x}x + \omega_{ei} + \omega_{di} + \omega_{si} + \omega_{ci}, 0) \tag{4.5}$$

for $i = 1, 2$, and

$$p(x|\omega_{e0}, \omega_{d0}, \omega_{s0}, x_0) = \mathcal{N}(x_0 + \omega_{e0} + \omega_{d0} + \omega_{s0}, 0) \tag{4.6}$$

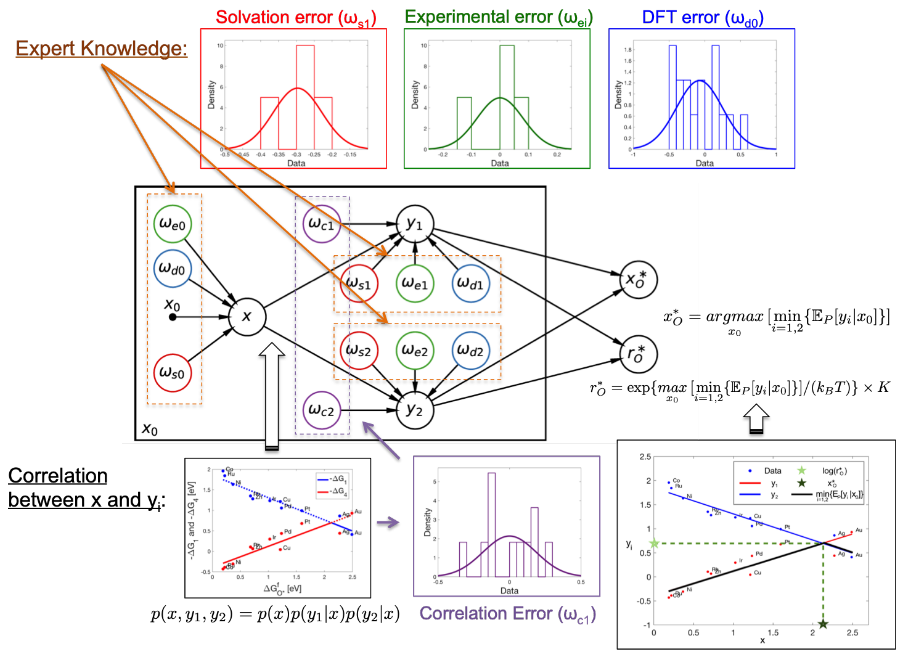

Figure 15. **PGMs allow us to combine heterogeneous data, expert knowledge and physical models: ORR PGM, where as (primary) output and QoI we construct the volcano curve between x_0 (oxygen binding energy) and y_i (reaction energies). We build the PGM via the following steps: (a) we construct a random variable x from the DFT data (using quantum calculations) for the oxygen binding energy given the real unknown value x_0; (b) we include statistical correlations between the DFT (quantum calculation) data x and y_i; (c) model the residual as an random error in correlation (random variable ω_{ci}); (d) we model as random variables and incorporate in the PGM different kinds of errors in x and y given by expert knowledge (see Section 4.1.3) from different sources (random variables ω_{ei}: error in experimental data, ω_{di}: error between quantum and experimental values, ω_{si}: error due to solvation effects which is calculated via DFT, i.e., we add these random variables into the PGM and build the connection/arrows with corresponded random variable x or y_i. Here we combine data from DFT computations (x, y_i, ω_{ci}, ω_{di}, ω_{si}, depicted in *blue*), with experimental data (ω_{ei}, ω_{di}, depicted in *green*); we fuse these heterogeneous experimental and computational data by taking advantage of the PGM formulation in Figure 16. Once the volcano curve between x and y_i is constructed, we obtain a prediction for the optimal oxygen binding energy x_O^* and optimal reaction rate r_O^* using physical modeling, i.e. that the optimal oxygen binding energy is identified when the two reaction energies are equal and the optimal reaction rate is proportional to $\exp\{\max[\min[y_1, y_2]]/(k_B T)\}$.**

$$p(\omega_j) = \mathcal{N}(\beta_{j0}, \sigma_j^2) \tag{4.7}$$

for all $j = e_0, d_0, s_0, e_1, d_1, s_1, c_1, e_2, d_2, s_2, c_2$.

(b) Once the graph is learned, we can select a parametric or semi-parametric family of PGMs (1.1), (1.2) and subsequently focus on parameter learning. Here we opt to use the global likelihood decomposition method, [66]. This approach is essentially a Maximum Likelihood Estimation (MLE) on PGMs, that exploits a fundamental scalability property that allows us to "divide and conquer" the parameter inference problem on the graph; Of course we can also employ a Bayesian approach instead of MLE, see for instance [66] for the case of PGMs.

In the MLE step, we take advantage of the Global Likelihood Decomposition [66],

$$L(\theta; \mathcal{D}) = \prod_i L_i(\theta_{i|\pi_i}|\mathcal{D}) = \prod_i \prod_m P(x_i[m]|x_{\pi_i}[m]; \theta_{i|\pi_i}) \tag{4.8}$$

where $L(\theta; \mathcal{D})$ is the likelihood given data $\mathcal{D} = \{\xi[1], \ldots, \xi[M]\}$ see Figure 14; noting that

$$\log L(\theta; \mathcal{D}) = \sum_i \log L_i(\theta_{i|\pi_i}|\mathcal{D}), \tag{4.9}$$

we observe that if we assume that $\theta_{i|\pi_i}$ are disjoint, i.e. that each conditional probability density, $p(x_i|x_{\pi_i}, \theta_{i|\pi_i})$, is parametrized by a separate set of parameters that do not overlap (this is a general assumption especially in our case, although we could extend all the results for shared parameters), we can pick the parameters $\hat{\theta}_{i|\pi_i}$ by solving

$$\hat{\theta}_{i|\pi_i} = \underset{\theta_{i|\pi_i}}{\operatorname{argmax}} \big[\log L_i(\theta_{i|\pi_i}|\mathcal{D}) \big] . \tag{4.10}$$

The formulas, (4.9) and (4.10), imply that we can "divide and conquer" our overall learning problem by learning the parameters $\theta_{i|\pi_i}$ for $p(x_i|x_{\pi_i}, \theta_{i|\pi_i})$ separately for each network node X_i using the corresponding parts of the data set $\mathcal{D}$ and (4.10).

Therefore, using MLE for the GBN (4.4) with given data to estimate the parameters as we describe above in (4.8)-(4.10), and the outcomes are shown in following table.

Table 2. Outcomes of MLE

$\beta_{y_1,0}$ = 0.0595	$\beta_{e0,0}$, $\beta_{ei,0}$ = 0
σ^2_{e0} = 0.0329	σ^2_{ei} = 0.0065
$\beta_{y_2,0}$ = 1.8231	$\beta_{d0,0}$ = -0.0754
$\beta_{di,0}$ = -0.0222	σ^2_{di} = 0.0354
β_{y_1,x_0} = 0.5111	σ^2_{d0} = 0.1032
$\beta_{s1,0}$ = -0.2967	σ^2_{s1} = 0.0046
β_{y_2,x_0} = -0.5564	$\beta_{s0,0}$ = 0.0067
$\beta_{s2,0}$ = -0.1209	σ^2_{s2} = 0.0054
$\beta_{ci,0}$ = 0	σ^2_{s0} = 0.0010
σ^2_{c1} = 0.0347	σ^2_{c2} = 0.0204

Software: In the ORR PGM case, since we only have a fairly small network, we can build the PGM component by component, essentially by hand. However, for more complex networks such as in medical or social science applications, there are numerous software which allow us to learn the structure and the parametric model from data or expert knowledge, for instance, BayesiaLab [19], Hugin [12, 89], Netica [133], Tetrad [59, 121] etc.

Although both the aforementioned learning tasks are well-studied in the PGM and AI literature, to our knowledge they have not been explored in physico-chemical applications. In such problems we are faced with a unique combination of challenges, such as multi-scale and multi-physics models, and the relatively sparse and heterogeneous data; some of the data can be expensive and coming from *different* sources and scales, such as experimental data and quantum, electronic structure computations. The overall ORR PGM combines data, multi-scale modeling and causal relationships, see Figure 15.

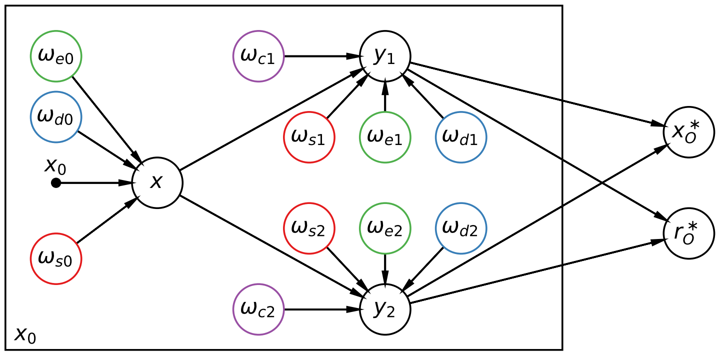

Figure 16. PGM for ORR where the QoI is a volcano curve, see Figure 13(c). The construction of the PGM is based on expert knowledge, physicochemical modeling and statistical analysis of data, see Table 3 for notation and Figure 15 for full details. In particular, here we consider a special class of PGMs, namely a Gaussian Bayesian Network, i.e., all CPDs are Gaussians (4.4) **which are fitted to available data using Maximum Likelihood Estimation. Note the conditional independence between the y-variables, assumed based on expert knowledge.**

Table 3. Notations used on the PGM in Figure 16

Notation	Meaning	Notation	Meaning
x_0	real oxygen binding energy ΔG^f_{O*}	x	ΔG^f_{O*} by electronic calculation
y_1	$-\Delta G_4 := \Delta G^f_{OH*}$	y_2	$-\Delta G_1 := -\Delta G^f_{OOH*} + \Delta G^f_{O_2}$
x^*_O	optimal ΔG^f_{O*}	r^*_O	optimal rate
ω_{ci}	error in correlation	ω_{ei}	error in experimental data
ω_{di}	error between electronic calculated values and experimental values	ω_{si}	error carried by solvation effect in water

In the next Sections we will assess the predictive capabilities of the PGM for ORR we built in Figure 16.

4.2 Post-Hoc Analysis of P: Uncertainty Quantification and Predictive Guarantees for PGMs

Once a baseline PGM model has been constructed as in Figure 16, we intend to use the resulting model for predictions of our Quantities of Interest (QoI). However, we first need to be convinced about the reliability and predictive capabilities of the model, given the uncertainties stemming from the sparse data and from multiple sources, all used in the construction of the model, as depicted in Figure 15.

In this direction, a proper Uncertainty Quantification (UQ) framework should provide quantitative insights into the reliability of our probabilistic model, for instance how much predictions can change by varying model parameters or more generally model features; specialized UQ methods such as Sensitivity Analysis (SA) should be capable to identify which parameters in a model have the most influence on predictions. In principle, one hopes to employ such UQ methods not only to assess the predictions of a model, but also to improve it by reducing its predictive uncertainty by reducing the uncertainty/error in the most influential parameters/mechanisms or by selecting more informative data, e.g. in Figure 15.

With such considerations in mind, we need to extend existing UQ methods to PGMs in order to handle the uncertainties caused by multiple sources of error, e.g. sparse data, lack of knowledge, incomplete modeling, and take advantage of the graph structure of the PGM, in particular correlations and causal relationships between model components and QoIs. To this end, when assessing the reliability of our predictions for our QoIs, there are two kinds of uncertainties arising with respect to the baseline PGM P we just built in Figure 16; we discuss them next.

A. Aleatoric Uncertainty for a given probabilistic model P: This type of model uncertainty is also known as statistical uncertainty and simply stems from the probabilistic nature of random variables described by a known probability distribution P. This type of UQ addresses questions of the following type: the QoI is a random variable (hence unknown from a deterministic perspective), however its' probabilistic model is known. For example, consider the QoI y in the volcano curve in Figure 15, given by

$$y|x_0 := \min(-\Delta G_1(\Delta G^f_{O^*}), -\Delta G_4(\Delta G^f_{O^*})) = \min(y_1|x_0, y_2|x_0) \tag{4.11}$$

for each x_0. Then, if we have a known baseline GBN model, see Section 4.1,

$$y_1|x_0 \sim \mathcal{N}(\alpha_1 x_0 + \beta_1, \sigma_1^2) \tag{4.12}$$

$$y_2|x_0 \sim \mathcal{N}(\alpha_2 x_0 + \beta_2, \sigma_2^2) \tag{4.13}$$

with some known constants α_i, β_i and σ_i, we obtain the probabilistic model P for the QoI (4.11):

$$P: \qquad y|x_0 \sim \min(\mathcal{N}(\alpha_1 x_0 + \beta_1, \sigma_1^2), \mathcal{N}(\alpha_2 x_0 + \beta_2, \sigma_2^2))\,, \tag{4.14}$$

which in turns provides the distribution of the QoI y for any fixed x_0, see Figure 18(L). In other words, we do not know the exact value of y, but we know the uncertainty it has with the baseline PGM P constructed in Figure 16; therefore we can calculate the mean value of the QoI (blue curve in Figure 18 (L),

$$\mathbb{E}_P\left[y|x_0\right] = \mu_1\Phi(\frac{\mu_2 - \mu_1}{\theta}) + \mu_2\Phi(\frac{\mu_1 - \mu_2}{\theta}) - \theta\phi(\frac{\mu_2 - \mu_1}{\theta}) \tag{4.15}$$

where $\mu_1 = \alpha_1 x_0 + \beta_1$, $\mu_2 = \alpha_2 x_0 + \beta_2$, $\theta = \sqrt{\sigma_1^2 + \sigma_2^2}$, and $\phi(\cdot)$, $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution respectively. Similarly we may consider any other statistics besides the mean, see also the full probability distribution function of the QoI random variable y in Figure 18.

B. Model-form Uncertainty around a "baseline" model P: This type of uncertainty quantification is also known as epistemic or systematic or structural uncertainty. Typically it stems from limited data and/or knowledge (e.g. the real model is too complex) available when building a baseline model P. Therefore model-form uncertainty can be characterized by the existence of many (or infinitely many!) alternative probabilistic models to P, see for instance Figure 17. In fact, the model P we initially construct based on the available data/expert knowledge is referred to as a "baseline" precisely because there many are alternative, possibly more predictive models to P for the QoIs we are interested in.

In this case, besides the aleatoric uncertainty of the QoI considered previously, model-form uncertainty is an additional uncertainty source for the probabilistic model P itself, here given by the PGM in Figure 16 and the corresponding GBN (4.4). For instance, in the example above, we consider the mean value of y (4.15) as our QoI, see also the blue curve in Figure 18 (Left)). Then for each x_0, (4.15) is deterministic for a given PGM P. However, the baseline model P is not the exact real model, i.e. there are "model-form uncertainties" around the baseline model P itself due to lack of data and/or knowledge regarding the probability distribution (see Figure 17). These additional uncertainties enter in a combined fashion from multiple sources in the PGM model P, see Figure 15, and propagate eventually to the QoI $\mathbb{E}_P\left[y|x_0\right]$; we refer to Figure 18 (R) for an initial demonstration and comparison to aleatoric uncertainties.

In this chapter, model form uncertainties are significant due to the limited amount of data available to build the PGM (4.4) in Figure 16, see for instance Figure 15. In fact, the second primary goal of this chapter–in addition to the introduction of PGMs in chemistry models–is to model, quantify and rank the impact of such model uncertainties, and provide predictive guarantees for the QoIs

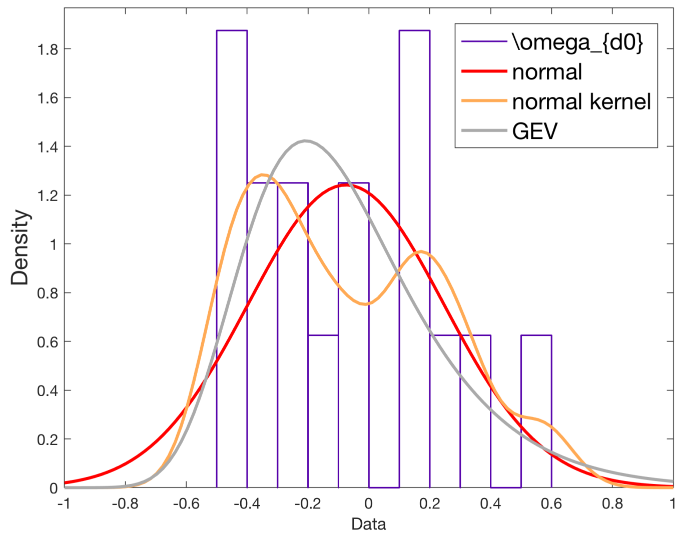

Figure 17. Example of single-source model-form uncertainty emanating from the CPD model (1.2) **for the PGM node** ω_{d0} **(see Figure 16 and Table 3). The model-form uncertainty stems from the different possible CPD models that can fit the depicted sparse data (histogram). Specifically, the dark blue curve is a Gaussian CPD and is part of the baseline model** P **for the predictive uncertainty analysis in** (4.26) **and in Figure 22; the red curve is a generalized extreme value (GEV) distribution (also parametric), which fits the data better than the Gaussian; the brown curve is a normal Kernel Density Estimator (KDE) of the histogram (non-parametric model) which fits the data better than both parametric models. Therefore the KDE can reduce model misspecification and eventually predictive uncertainty of QoIs (see Section 4.4). Depicted sparse data are due to the limited number of metals for catalysts in the periodic table and a small number of quantum calculations we can afford to perform; thus, sparsity of available data induces *model-form uncertainty*. This uncertainty from the PGM node** ω_{d0} **propagates through the graph to the QoIs** x^*, r^* **in Figure 16. Finally, each node in Figure 16 provides an additional source of model-form uncertainty. We rank the impact of all such uncertainties on the QoI in Section 4.3.3.**

in their presence. One such example of a QoI is x^*_O in Figure 16, i.e.,

$$x^*_O := \underset{x_0}{\operatorname{argmax}}[\min\{\mathbb{E}_P\left[y_1|x_0\right], \mathbb{E}_P\left[y_2|x_0\right]\}] \tag{4.16}$$

see also Figure 19 for a demonstration. We discuss these points in full detail in the next Sections and provide all mathematical details in the Appendix B.

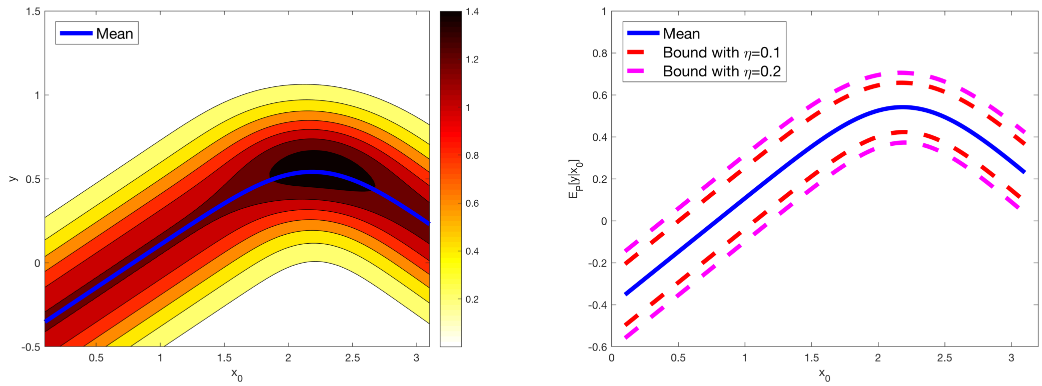

Figure 18. (L) Aleatoric Uncertainty: Contour plot of the probability distribution of $y = \min(y_1, y_2)$ (where $y_1 : -\Delta G_4$, $y_2 : -\Delta G_1$) as a function of $x_0 = \Delta G^f_{O*}$, capturing the randomness of the QoI y; the blue curve is the mean (expected) value $\mathbb{E}[y|x_0]$ for the ORR PGM P in Figure 16. (R) Model-form Uncertainty: The predictive guarantees (dotted lines) for the QoI $\mathbb{E}_P[\min(y_1|x_0, y_2|x_0)]$ if the alternative PGM model Q satisfies $R(Q||P) \leq 0.1$ or ≤ 0.2. The definition and details on $R(\cdot||\cdot)$ and predictive guarantees will be presented in subsequent Sections and the Appendices.

4.3 Model-form UQ & Sensitivity Analysis

The primary goal of this Section is using the concept of the model-form sensitivity index shown in Section 3.2 to quantify and rank the impact of model uncertainties from each component of the PGM–the components described mathematically by CPDs $p(x_l|x_{\pi_l})$ in (1.1), (1.2)–to the QoIs f.

4.3.1 Model misspecification on ORR PGM

For PGMs such as (1.1) are special because they are built based on individual CPDs (1.2), therefore each CPD $p(x_l|x_{\pi_l})$ needs to be associated with its' own

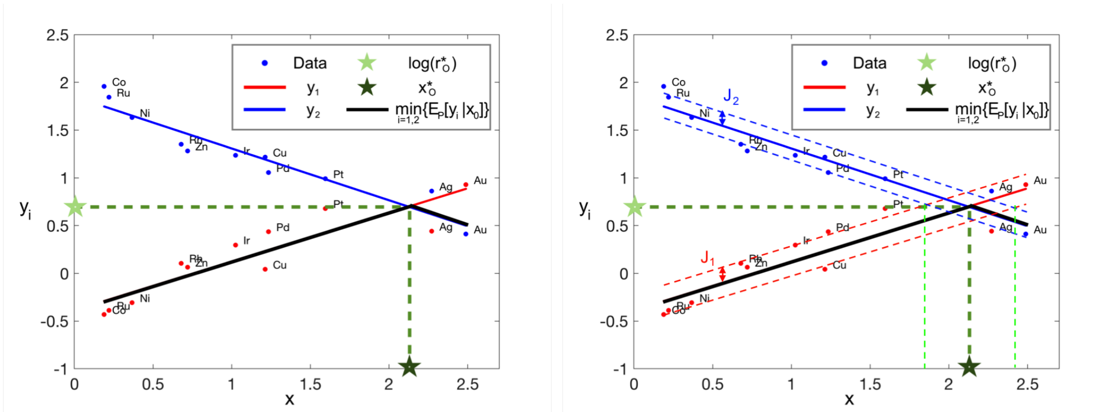

Figure 19. **(L) Aleatoric Uncertainty: QoIs of the ORR model shown in Figure 15, where the optimal oxygen binding energy x^*_O is identified when the two reaction energies are equal by *physical modeling*: we set it to be** $\mathrm{argmin}_{x_0}(\mathbb{E}_P\left[y_1|x_0\right], \mathbb{E}_P\left[y_2|x_0\right])$**; then the optimal reaction rate r^*_O is given by** $\exp\{\max[\min[y_1, y_2]]/(k_B T)\} \times K$. **(R) Model-form Uncertainty: The predictive guarantees for the average of the QoI x^*_O given by model P in Figure 15 are calculated in terms of guaranteed confidence bounds J_i, see Section 4.3.2. The predictive guarantees are depicted by the green dotted lines around the baseline prediction corresponding to $\mathbb{E}_P\left[x^*_O\right]$ calculated first on the Left panel. Note that not all QoIs are impacted (but not all the same!) from model-form uncertainties: compare blue, red and green confidence intervals in the Right panel, as well as in Figure 23.**

model misspecification parameter η_l: Figure 15 depicts the multiple model/data uncertainties that enter during the building of the baseline model at each component CPD of the PGM P. To this end, and in order to isolate and rank the impact of each individual model misspecisfication η_l, we consider the domain of all PGMs $\mathcal{D}^{\eta_l}_{l,P}$ which are identical to the entire PGM P except at the l-th component CPD and can be η_l away in KL from the baseline CPD $p(x_l|x_{\pi_l})$, while maintaining the same parents x_{π_l}:

$$\mathcal{D}^{\eta_l}_{l,P} = \left\{ \begin{array}{c} Q : R(Q_{l|\pi_l}||P_{l|\pi_l}) \leq \eta_l \text{ for all } x_{\pi_l}\text{of model } P, \\ q(x_j|x_{\pi_j}) \equiv p(x_j|x_{\pi_j}) \text{ for all } j \neq l \end{array} \right\} . \tag{4.17}$$

As we discuss in Section 3.2.3, this is an infinite dimensional set containing all possible models (non-parametric) which are η_l-"close" to P at the l-th component CPD of the PGM in KL divergence. Furthermore, η_l will be either calculated as the KL distance of the baseline model P_l from the available data, or η_l can take arbitrary fixed values that correspond to model P perturbations associated with sensitivity analysis; we will be discuss this latter point in Section 4.3.3.

4.3.2 Model-form sensitivity indices for ORR PGM

Here we quantify and rank the impact on the QoI f of model misspecisfication η_l for each CPD. This is a form of non-parametric sensitivity analysis for PGMs which will allow us to re-evaluate and improve our baseline models by comparing the contributions of each CPD to the overall predictive uncertainty in Section 4.4. Based on the definition of model-form sensitivity indices shown in Section 3.2.3, i.e.,

$$I^{\pm}(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) := \sup_{Q \in \mathcal{D}_{l,P}^{\eta_l}}/\inf \mathbb{E}_Q [f] - \mathbb{E}_P [f] \, , \tag{4.18}$$

it captures the impact of model-form uncertainties entering in any baseline CPD $p(x_l|x_{\pi_l})$, e.g. see Figure 17, to the QoIs of interest, as uncertainty propagates through the graph and the PGM; for instance, we refer to the QoIs $f = x^*$ or $f = r^*$ in Figure 16. In addition, we can also consider the corresponding relative bias

$$\frac{I^{+}(f(X), P; \mathcal{D}_{l,P}^{\eta_l})}{|\mathbb{E}_P [f]|} \, , \tag{4.19}$$

as a percentage relative to the baseline value of the QoI.

Moreover, based on Theorem 3.5 and Corollary 3.6, the indices $I^{\pm}(f(X), P; \mathcal{D}_{l,P}^{\eta_l})$ can be computed exactly using a variational formula for the KL divergence for our ORR example. In particular, we have

Theorem 4.1 *Let* $P = \mathcal{N}(\mu, \mathcal{C})$ *be the joint distribution of ORR PGM defined on* (4.1) *with given* x_0 *and QoI* $f(X) = y_i$. *Then:*

(a) The model-form sensitivity indices (4.18) *for the node* ω_l *with some* $\eta_l > 0$ *are given by*

$$I^{\pm}(f(X), P; \mathcal{D}^{\eta_l}_{l,P}) = \pm |\tilde{\beta}_{y_i,\omega_l}| \sqrt{2\sigma_l^2 \eta_l} \ , \tag{4.20}$$

where σ_l *is given in* (4.7) *and* $\tilde{\beta}_{y_i,\omega_l}$ *is given in Table 7 in Appendix B.*

(b) Furthermore, if we perturb each component with same η *for any given parents, i.e.,* $\eta_j \equiv \eta$ *for each* j *with any given* $X_{\pi_j} = x_{\pi_j}$, *then we can rank all PGM components by the relative magnitude of the sensitivity indices*

$$\frac{I^+(f(X), P; \mathcal{D}^{\eta_l}_{l,P})}{\sum_j I^+(f(X), P; \mathcal{D}^{\eta_j}_{j,P})} = \frac{|\tilde{\beta}_{y_i,\omega_l}| \sqrt{2\sigma_l^2}}{\sum_j |\tilde{\beta}_{y_i,\omega_j}| \sqrt{2\sigma_j^2}} \tag{4.21}$$

(c) More generally, let P *be any joint distribution (not necessarily a GBN) for ORR PGM defined on* (4.1). *For* $f(X) = y_i$, *the model-form sensitivity indices defined in* (4.18) *for the node* ω_l *with some* $\eta_l > 0$ *are given by*

$$I^+(f(X), P; \mathcal{D}^{\eta_l}_{l,P}) = \inf_{c>0} \left[\frac{1}{c} \log \int e^{c\bar{F}_l} P_l(dx_l) + \frac{\eta_l}{c} \right] \tag{4.22}$$

where $\bar{F}_l(X) = F_l(X) - \mathbb{E}_P[F_l(X)] = F_l(X) - \mathbb{E}_P[f(X)]$ *and*

$$F_l(x) = \int y_i \prod_{X_i \in \{\omega_l\}^c} P(dx_i | x_{\pi_i}) = \tilde{\beta}_{y_i,0} + \tilde{\beta}_{y_i,\omega_l} \omega_l \ . \tag{4.23}$$

Remark [On the choice of a non-parametric setting]: The proposed UQ tools in this Section are non-parametric in nature since our challenges can involve uncertainty in the probabilistic model itself, as depicted in Figure 17 and for the entire model in Figure 15. On the other hand, we need to also remark that the proposed indices (4.18) and (3.86) can be too pessimistic when considering uncertainty/sensitivity questions for models confined within a particular parametric family. Indeed, since the uncertainty and sensitivity indices proposed above are based on KL divergence, they are inherently non-parametric and thus the resulting family of distributions (4.17) allows for densities that may not be attainable within a particular parametric family. For example, if we already know the probabilistic models we need to consider lie exclusively within a fixed parametric family, e.g. Gaussians such as (4.4), our non-parametric bounds (4.18) can be too wide since the family (4.17) includes many other distributions outside the parametric family at hand, namely Gaussians.

However for the physico-chemical problems considered here and due to the sparsity of available experimental and electronic-structure computational data–see for instance Figure 17 and Figure 15–our resulting family of probabilistic models is intrinsically non-parametric and is built as a "neighborhood" around a baseline model P. For instance, here the baseline model P is selected to be a Gaussian fit to the histogram of the CPD in Figure 17. Furthermore, many alternative densities to P are possible, e.g. given by various choices of kernel density estimators of the histogram in Figure 17 or other parametric families. Therefore considering the non-parametric family of models (4.17) and the resulting sensitivity index (4.18) is a natural and in fact necessary choice.

4.3.3 Model misspecification parameter η_l and PGM components ranking

The model misspecification parameters η_l are necessary in the calculation of the model-form sensitivity indices (4.20), see also Figure 21. As we show in Section 3.3 they can be practically selected or estimated in at least two different ways:

1. First, η_l can be calculated as the KL distance of the CPD $p(x_l|x_{\pi_l})$ in the

baseline PGM P in (1.1) from the available data in the form of a histogram or a KDE, see Figure 17 and Figure 15; we refer to Section 3.3 for full details. The resulting estimated η_l values constitute *surrogates* for the distance of the baseline model from the unknown "real" model.

2. Alternatively, η_l can take arbitrary fixed values that correspond to model perturbations associated with *local* sensitivity analysis (small η_l's) or *global* sensitivity analysis (larger η_l's). Both types of sensitivity analysis are conducted in the same mathematical framework, therefore we have the flexibility to explore combinations of small and large model perturbations at different nodes of the PGM.

Once we have selected η_l values for the baseline PGM, the model-form sensitivity indices defined in Section 4.3.2 are defined as the expected bias when we perturb *only one part of the model* in the PGM within $\eta_l > 0$; therefore, they measure the impact of uncertainty in one specific component in the PGM on the QoI f. We use the model-form sensitivity indices to rank PGM components according to the percentages of sensitivity indices,

$$\frac{I^+(f(X), P; \mathcal{D}^{\eta_l}_{l,P})}{\sum_j I^+(f(X), P; \mathcal{D}^{\eta_j}_{j,P})} \tag{4.24}$$

For any QoI $f(X) = X_i$, as discussed above, we consider the second perspective in setting identical model misspecification values η_l in the indices $I^{\pm}(f(X), P; \mathcal{D}^{\eta_l}_{l,P})$ in (4.24) similarly to parametric sensitivity analysis. Thus we perturb each part of model, $p(x_l|x_{\pi_l})$, by the same amount of model disspecification η_l in the sensitivity indices (4.18). Then, the indices $I^{\pm}(f(X), P; \mathcal{D}^{\eta_l}_{l,P})$ in (3.86) will yield a ranking of the impact of each component on the model,

$$\frac{I^+(f(X), P; \mathcal{D}^{\eta_l}_{l,P})}{\sum_j I^+(f(X), P; \mathcal{D}^{\eta_j}_{j,P})} = \frac{|\tilde{\beta}_{il}|\sqrt{2\sigma_l^2}}{\sum_j |\tilde{\beta}_{ij}|\sqrt{2\sigma_j^2}}\,. \tag{4.25}$$

Here σ_j^2 is the variance of X_j under the conditional probability distribution $p(x_j|x_{\pi_j})$ as we defined in (4.4), $\tilde{\beta}_{ij}$ depends on the linear Gaussian coefficients of $X_1, \ldots, X_i$, which illustrate the linear dependency between X_i and X_j given the ancestors of X_j, and $\tilde{\beta}_{ij} = 0$ for $j \notin \rho_i$. For more details we refer to Appendix B. We can show that the ratio of indices will only depend on Gaussian coefficients and the covariance matrix, while the value of η_l will not affect the result in this case. A demonstration of the rankings (4.25) for the ORR PGM is shown in Figure 22(L).

We can also estimate η_l as the "distance" between data and our PGM P (1.1), where data is represented by a histogram or a KDE approximation of the histogram, or any given particular model Q from data or expert knowledge. In this case, η_l may be different for different PGM components l and thus we have from (4.20):

$$\frac{I^+(f(X), P; \mathcal{D}^{\eta_l}_{l,P})}{\sum_j I^+(f(X), P; \mathcal{D}^{\eta_j}_{j,P})} = \frac{|\tilde{\beta}_{il}|\sqrt{2\sigma_l^2\eta_l}}{\sum_j |\tilde{\beta}_{ij}|\sqrt{2\sigma_j^2\eta_j}}\,, \tag{4.26}$$

We refer to Figure 22(R) for a demonstration, while more details and derivations are included in Appendix B.

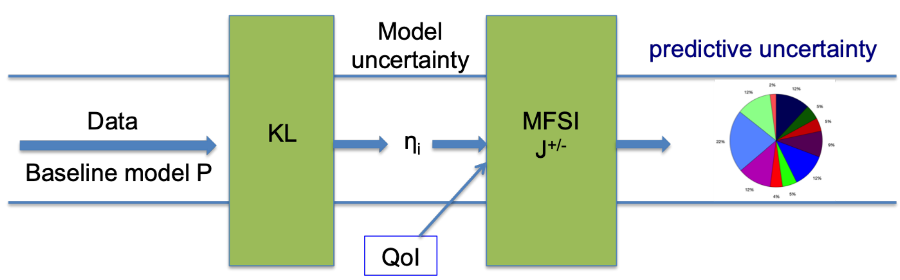

Figure 20. Schematic description of our proposed methodology: Predictive uncertainties of the QoI for each component on PGM (for the pie chart, see Figure 22) are calculated and are due to model-form uncertainties; inputs to our methodology are (sparse) DFT and experimental data and of course the baseline model P from Figure 16.

4.3.4 Model-form sensitivity indices for QoIs x_O^* and r_O^* in the ORR PGM

Here we demonstrate the model-form sensitivity indices and ranking of PGM components for our QoIs, namely the optimal oxygen binding energy $\Delta G^f_{O^*}$ and the optimal reaction rate, x_O^* and r_O^* in Figure 16. In this case,

$$\begin{aligned} I^{\pm}(x_O^*, P; \mathcal{D}^{\eta_l}_{l,P}) \quad := \quad & \sup_{Q\in\mathcal{D}^{\eta_l}_{l,P}}/\inf\{\underset{x_0}{argmax}\,[\min\{\mathbb{E}_Q[y_1|x_0], \mathbb{E}_Q[y_2|x_0]\}] - \\ & \underset{x_0}{argmax}\,[\min\{\mathbb{E}_P[y_1|x_0], \mathbb{E}_P[y_2|x_0]\}]\} \end{aligned} \quad (4.27)$$

Indeed, by solving the optimization problem for $x_O = x_O^*$, we have:

- if $l = ei, di, si, ci$ (various types of errors which affect y_i, see Table 3), $i = 1, 2$

$$I^{\pm}(x_O^*, P; \mathcal{D}^{\eta_l}_{l,P}) = \frac{I^{\pm}(y_i, P; \mathcal{D}^{\eta_l}_{l,P})}{\beta_{y_1,x} - \beta_{y_2,x}} = \frac{\pm\sqrt{2\sigma_l^2\eta_l}}{\beta_{y_1,x} - \beta_{y_2,x}}. \quad (4.28)$$

- if $l = e0, d0, s0$ (various types of errors which affect x, see Table 3)

$$\begin{aligned} & I^{\pm}(x_O^*, P; \mathcal{D}^{\eta_l}_{l,P}) = \\ & \qquad \frac{I^{\pm}(y_2, P; \mathcal{D}^{\eta_l}_{l,P}) - I^{\mp}(y_1, P; \mathcal{D}^{\eta_l}_{l,P})}{\beta_{y_1,x} - \beta_{y_2,x}} = \frac{\pm(|\beta_{y_1,x}| + |\beta_{y_2,x}|)\sqrt{2\sigma_l^2\eta_l}}{\beta_{y_1,x} - \beta_{y_2,x}}. \end{aligned} \quad (4.29)$$

Here σ_l^2 is the variance of X_l under the conditional probability distribution $p(x_l|x_{\pi_l})$, defined in (4.4), and $\beta_{y_i,x}$ are the coefficients given by $p(y_i|x)$, see Section 4.1.4. Similarly, we can compute the model-form sensitivity indices for r_O^*.

Remark [Propagation/Non-Propagation of Uncertainties to the QoIs]: We note the discrepancies in the propagation of model misspecification to the QoI between different PGM components, as demonstrated in Figure 22. In particular, in Figure 22(L) the same uncertainty (described by model misspecification η_l) is

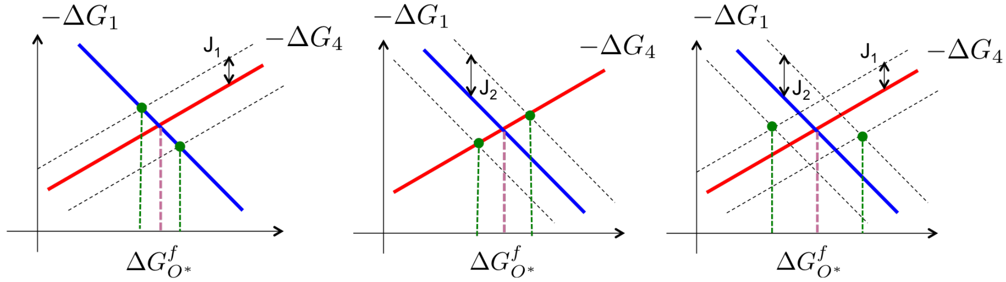

Figure 21. Predictive Uncertainty bounds $J_i, i = 1,2$ for the QoI x^*_O (see Figure 19(R)) for model misspecification η_l in $P(\omega_l)$: (a) for $l = e1, d1, s1, c1$, $J_1 = I^+(y_1, P; \mathcal{D}^{\eta_l}_{l,P}) = \sqrt{2\sigma_l^2\eta_l}$; (b) for $l = e2, d2, s2, c2$, $J_2 = I^+(y_2, P; \mathcal{D}^{\eta_l}_{l,P}) = \sqrt{2\sigma_l^2\eta_l}$; (c) for $l = e0, d0, s0$, $J_i = I^+(y_i, P; \mathcal{D}^{\eta_l}_{l,P}) = |\beta_{y_i,x}|\sqrt{2\sigma_l^2\eta_l}$, $i = 1,2$.

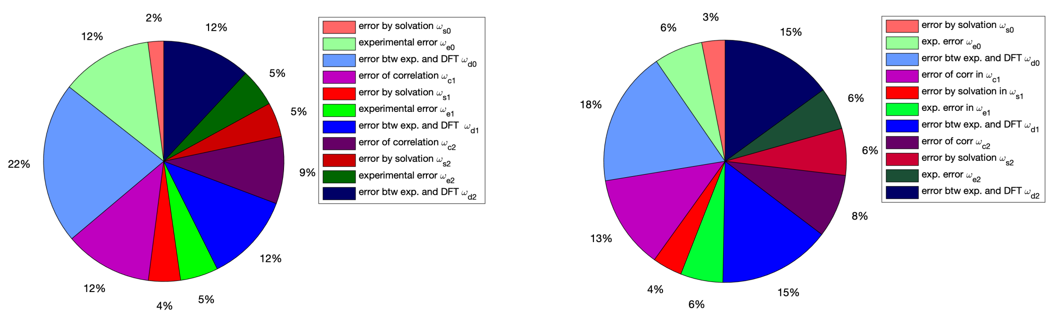

Figure 22. Relative percentage sizes of predictive uncertainty of x^*_O in each ORR PGM mechanism in Figure 16 using (4.26). (L): Here η_l has a fixed value for all l; the particular value does not matter since it is canceled out by the ratio, see (4.25). (R): In this case we select $\eta_l = R(data||P_l)$ as a distance of each CPD from the available data; for details and derivations we refer to the Section 3.3. The analysis brings together knowledge from data, physical models from different scales/mechanisms, including mechanisms and data from different expert groups.

applied on all ORR PGM nodes, however not all propagate and affect the same the QoI: see Figure 23 for examples of propagation (22%) and non-propagation (5% and 0%) of model misspecification to the QoI.

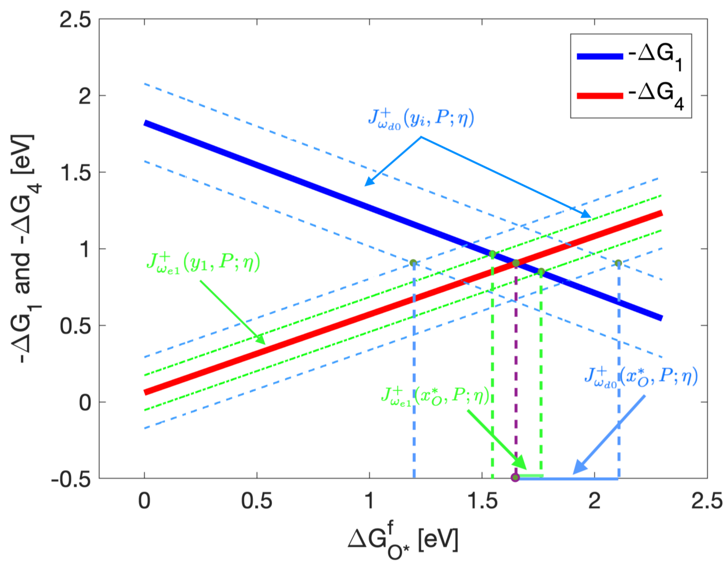

Figure 23. ***Propagation* vs. *Non-propagation* of model misspecification of the PGM nodes ω_{d0} and ω_{e1} respectively, to the predictions of the QoI x_O^*; misspecification is set to $\eta = 1$ for both PGM nodes. First, note that $I^+(y_2, P; \mathcal{D}^\eta_{\omega_{e1}}) = 0$ i.e., the model misspecification of ω_{e1} only affects the prediction of y_1, but not y_2, see Figure 21; therefore the uncertainty of ω_{e1} only propagates to x_O^* through y_1, while $I^+(y_1, P; \mathcal{D}^\eta_{\omega_{e1}})$ is small since ω_{e1} has a lower variance which is associated with more informative available data. Thus, it results in a small corresponding uncertainty in x_O^*. Meanwhile, the uncertainty of ω_{d0} propagates to x_O^* through both y_1 and y_2, (i.e., the model misspecification of ω_{d0} affects both the predictions of y_1 and y_2), and $I^+(y_i, P; \mathcal{D}^\eta_{\omega_{d0}})$ is larger since ω_{d0} has a higher variance (due to insufficient informative data available). Therefore we have a larger corresponding uncertainty in x_O^* predictions, as shown in the Figure.**

4.4 Improving Models via Predictive Uncertainty Reduction: Model Complexity vs. Data Acquisition

Given the already constructed baseline model P (see Section 4.1) and the sparse data set for each model component sampled from an unknown model Q (e.g. as shown in Fig 14), we can build an improved baseline model P for our ORR model through the procedure presented in Section 3.4 in Steps 1-3 below.

Step 1: Find suitable data-based η_l's:

$$\eta_l = \max_{x_{\pi_l}} R(Q(X_l|x_{\pi_l})||p(x_l|x_{\pi_l}))$$

where Q is the surrogate model given by KDE/histogram, using (3.94),(3.95).

Step 2: Calculate the model-form sensitivity indices (4.20):

$$I^{\pm}(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) \qquad \text{for all PGM nodes } l \ ,$$

for a given QoI f using (4.20).

Step 3: We target any l^*- component X_{l^*} on the PGM; usually we select the l's with the highest $I^{+}(f(X), P; \mathcal{D}_{l,P}^{\eta_l})$ values (we handle $I^{-}(f(X), P; \mathcal{D}_{l^*}^{\eta_{l^*}})$ in a similar fashion), see also Section 3.4. Then, we reduce $I^{+}(f(X), P; \mathcal{D}_{l^*}^{\eta_{l^*}})$ based on the result in (4.20), i.e., for $f(X) = X_i$ we have that

$$I^{\pm}(f(X), P; \mathcal{D}_{l^*}^{\eta_{l^*}}) = \pm\sqrt{2}|\beta_{il^*}|\sqrt{\sigma_{l^*}^2}\sqrt{\eta_{l^*}} \tag{4.30}$$

Two detailed methods of reducing $I^{+}(f(X), P; \mathcal{D}_{l^*}^{\eta_{l^*}})$ are shown below, and we can use either one (or both) of the methods, depending on which is easier to implement first.

4.4.1 Identifying additional "high quality data" – variance reduction

Based on Steps 1-3 above as starting point, we develop the following strategy for identifying and acquiring additional, useful data:

1. Using Steps 1-3 we target the l^*-components of the PGM with (some of) the higher values of predictive uncertainty determined by $I^{\pm}(f(X), P; \mathcal{D}_{l^*}^{\eta_{l^*}})$.

2. For the l^* components of the PGM we seek the most useful additional data, namely the data that reduce the predictive uncertainty (4.30), i.e. reduce the

combination of the variance $\sigma^2_{l^*}$ and the model misspecification η_{l^*}, where the latter is estimated from data Section 4.3.3.

In fact, this perspective relying on (4.30), identifies what is the right type of data and how to prioritize our focus on data retrieval on the nodes of PGM (pick the best l^*) as far as predictions for the QoI f are concerned. Specifically, we seek data that lead to the reduction of the variance $\sigma^2_{l^*}$, while the model misspecification η_{l^*} does not increase or the increment is much smaller than the reduction of $\sigma^2_{l^*}$. Notice that in this case the model remains a Gaussian Network.

For the ORR PGM it turns out that we can add more data using DFT calculations for Bimetallics to reduce the variance of the correlation errors ω_{ci}, σ^2_{ci}. Then the predictive uncertainty of y_i on ω_{ci}, $J^{\pm}_{\omega_{ci}}(y_i, P; \eta_{\omega_{ci}})$, is reduced according to (4.30), while the model misspecification $\eta_{\omega_{ci}}$ is also reduced in this case. Same for the predictive uncertainty of QoI x_{O^*}, see (4.28). All results are collected in Figure 24.

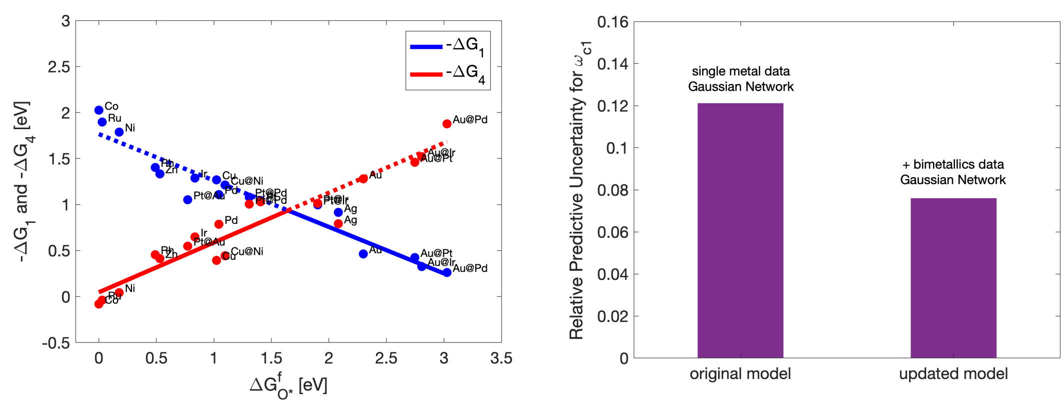

Figure 24. (L): DFT-computed data for reaction energies with respect to different metals/oxygen binding energies. Here we also include Bimetallics data in addition to the single metals in Figure 14. (R): Reduction of predictive uncertainty (4.20) **of** x^*_O **by reducing the model uncertainties of** ω_{ci} **where here we set** $\eta_{c1} = R(data||P_{c1})$**, see also Section 4.3.3.**

4.4.2 Improving the baseline model P – model misspecification reduction

Based on (4.30), an alternative route is to reduce the model misspecification η_{l^*} by picking a better model, $\tilde{P}_{l^*}$, than the baseline model P_{l^*}; the new model should represents the (fixed) available data more accurately by using a kernel-based method; in this case the new model is a "hybrid" Bayesian Network, i.e. it is a mixture model of Gaussian and kernel-based networks. For example, if we replace the linear, Gaussian model for ω_{c1} in Figure 16 with a linear, kernel-based model as shown in Figure 25 (Left), we can reduce the predictive uncertainty by reducing the model misspecification η_i.

Moreover, we can combine the approaches above to reduce the predictive uncertainty, e.g., after adding more bimetallics data, we can further reduce the uncertainty by replacing the corresponding component of the baseline model for ω_{c1} (Gaussian model) by normal kernel density estimator as shown in Figure 25. We can compute the model-form sensitivity indices $J_l^{\pm}$ for the updated hybrid model, where P_l could be KDE or another distribution, using Theorem 3.5 and in particular (4.22).

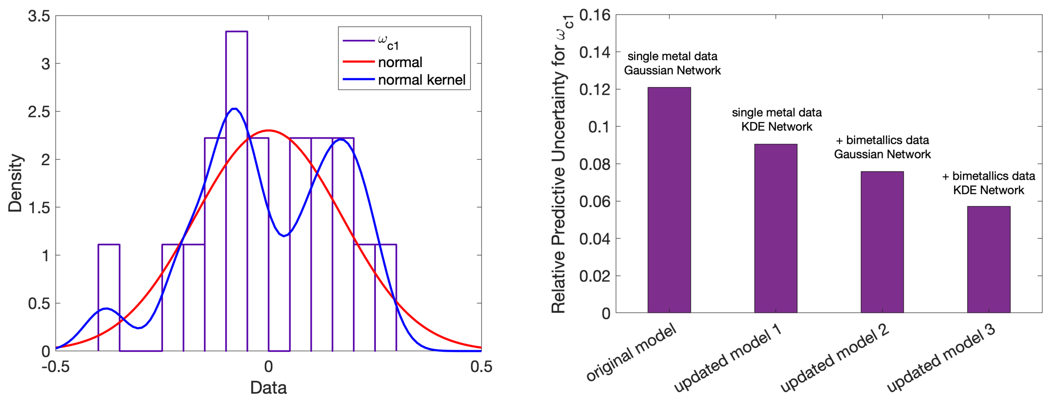

Figure 25. (L): Baseline model (Gaussian) of ω_{c1} (red curve) and the updated model (normal-kernel density estimation, blue curve) and additional Bimetallics data from Figure 24. (R): Different relative predictive uncertainties (4.19) when we: only perturb the model of ω_{c1} by $\eta_{c1} = R(data||P_{c1})$ when P_{c1} is Gaussian with the original single-metal data; or using a KDE with the original data (updated model 1); or using a Gaussian with the additional Bimetallics data (updated model 2); or using both KDE and Bimetallics data (updated model 3).

CHAPTER 5

SENSITIVITY ANALYSIS FOR PARAMETRIC PROBABILISTIC GRAPHICAL MODELS

In this chapter, we provide a new sensitivity analysis method for parametric PGMs. In the cases where we are confident about the parametric family that our model should follow, there is no need to consider the model-from uncertainty and the uncertainty indices we introduced above may be too pessimistic (as they are inherently non-parametric) when studying uncertainty/sensitivity questions for models *confined within* a given parametric family; e.g. if we have confidence in the "physics" involved, e.g. PGMs in medical diagnostics with binomial CPDs, since a test can be only positive/negative, [76]. . Therefore, once the parametric structure of the PGM is already established, we need a set of UQ and SA tools suitable for parametric PGMs. Existing UQ and SA methods, such as gradient or ANOVA methods, [122, 114, 31] are not clearly taking advantage of the graphical, causal structure in PGMs. In this direction, we will explore SA methods for parametric sensitivity analysis for PGM using Likelihood Ratio and Fisher Information Matrix, and compare it with the model-form sensitivity indices we introduce above.

5.1 Likelihood Ratio Method and Score Function

Considering a parametric distribution P^θ embedded on a PGM which could be written as

$$p^\theta(x) = \prod_{i=1}^{n} p^{\theta_i}(x_i|x_{\pi_i}) \tag{5.1}$$

where X_{π_i} is the parents of node X_i, and $\theta = (\theta_1, \ldots, \theta_n)$ with θ_i are the parameters of conditional distribution $P_{i|\pi_i}$ (θ_i can be a vector and may depend on X_{π_i}), and we assume they satisfy the following assumptions:

Assumption 1 *For any $l \in \pi_i$, we have $l < i$, i.e., $X_{\pi_i} \subset \{X_1, \ldots, X_{i-1}\}$ for all i.*

Assumption 2 *for any $i \neq j$, θ_i are disjoint with θ_j, i.e. each conditional probability density, $p^{\theta_i}(x_i|x_{\pi_i})$, is parameterized by a separate set of parameters that do not overlap.*

Note that these are general assumptions, we could always get Assumption 1 by reordering $(X_1, \ldots, X_n)$, and extend all the results for the models which have shared parameters, i.e. do not satisfy Assumption 2 ([66] Theorem 7.5). Then for a parametric PGM, we give the following definition:

Definition 5.1 *For a parametric PGM, P^θ, as defined in* (5.1)*, we define the* ***score function of the PGM*** *by*

$$W^\theta(x) \quad := \quad \nabla_\theta \log p^\theta(x) = \nabla_\theta \sum_{i=1}^{n} \log p^{\theta_i}(x_i|x_{\pi_i}) \tag{5.2}$$

Furthermore, we have

$$\begin{aligned} W^\theta(x) &= \nabla_\theta \sum_{i=1}^{n} \log p^{\theta_i}(x_i|x_{\pi_i}) \\ &= (\nabla_{\theta_1} \log p^{\theta_1}(x_1|x_{\pi_1}), \ldots, \nabla_{\theta_n} \log p^{\theta_n}(x_n|x_{\pi_n}))^T \\ &= (W^{\theta_1}(x_1), \ldots, W^{\theta_n}(x_n))^T \end{aligned} \tag{5.3}$$

where

$$W^{\theta_i}(x_i) = \nabla_{\theta_i} \log p^{\theta_i}(x_i|x_{\pi_i}) \tag{5.4}$$

is the score function of conditional distribution $P_{i|\pi_i}$ with CPD $p^{\theta_i}(x_i|x_{\pi_i})$ in (5.1) and given parents x_{π_i} for X_i [131]. And it satisfies

$$\begin{aligned}
\mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[W^{\theta_i}(X)\right] &= \mathbb{E}_{P^{\theta_i}_{i\pi_i}}\left[\nabla_{\theta_i} \log p^{\theta_i}(x_i|x_{\pi_i})\right] \\
&= \mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[\frac{\nabla_{\theta_i} p^{\theta_i}(x_i|x_{\pi_i})}{p^{\theta_i}(x_i|x_{\pi_i})}\right] \\
&= \int \nabla_{\theta_i} p^{\theta_i}(x_i|x_{\pi_i}) dx_i \\
&= \nabla_{\theta_i} \int p^{\theta_i}(x_i|x_{\pi_i}) dx_i = 0
\end{aligned} \tag{5.5}$$

Then for any QoI $f(X) = f(X_1, \ldots, X_n)$, the gradient based sensitivity index $\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X)\right]$ can be represented using the score function:

$$\begin{aligned}
\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X)\right] &= \nabla_\theta \int f(x) p^\theta(x) dx \\
&= \int f(x) \nabla_\theta p^\theta(x) dx \\
&= \int f(x) \nabla_\theta \log p^\theta(x) p^\theta(x) dx \\
&= \mathbb{E}_{P^\theta}\left[f(X) W^\theta(X)\right] \\
&= \left(\mathbb{E}_{P^\theta}\left[f(X) W^{\theta_1}(X_1)\right], \ldots, \mathbb{E}_{P^\theta}\left[f(X) W^{\theta_n}(X_n)\right]\right)^T
\end{aligned} \tag{5.6}$$

and we call $\mathbb{E}_{P^\theta}\left[f(X) W^\theta(X)\right]$ the Likelihood Ratio (LR) estimator for the gradient based sensitivity index $\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X)\right]$ since it can be evaluated exactly with Monte Carlo sampling [44, 5].

Moreover, for the special case that $f(X) = f(X_l)$, for some $1 \leq l \leq n$, we have

$$\begin{aligned}
\mathbb{E}_{P^\theta}\left[f(X_l) W^{\theta_i}(X_i)\right] &= \mathbb{E}_{P^\theta_{\pi_i}}\left[\mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[f(X_l) W^{\theta_i}(X_i)\right]\right] \\
&= \mathbb{E}_{P^\theta_{\pi_i}}\left[f(X_l) \mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[W^{\theta_i}(X)\right]\right] = 0
\end{aligned} \tag{5.7}$$

for any $i > l$, where $P^{\theta_i}_{i|\pi_i}$ is the conditional probability with density function $p^{\theta}_i(x_i|x_{\pi_i},)$ with given $X_{\pi_i} = x_{\pi_i}$, $P^{\theta}_{\pi_i}$ is the marginal distribution of P^{θ} for X_{π_i}. Therefore,

$$\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X_l)\right] = (\mathbb{E}_{P^{\theta}_{1:l}}\left[f(X_l)W^{\theta_1}(X_1)\right], \ldots, \mathbb{E}_{P^{\theta}_{1:l}}\left[f(X_l)W^{\theta_l}(X_l)\right], 0, \ldots, 0)^T \tag{5.8}$$

where $P^{\theta}_{1:l}$ is the marginal distribution for $(X_1, \ldots, X_l)$. More specifically, we have $\mathbb{E}_{P^\theta}\left[f(X_l)W^{\theta_i}(X_i)\right] = 0$ if X_i is not an ancestor of X_l.

We summarize all the results above in the following Theorem that allows to describe local sensitivities of PGMs in terms of the score function of PGMs:

Theorem 5.2 *(a) For any PGM $p(x|\theta) = \prod_{i=1}^{n} p^{\theta_i}(x_i|x_{\pi_i})$ that satisfies Assumption 1 and 2, and a given QoI $f(X)$, the gradient based sensitivity index $\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X)\right]$ can be estimated by the Likelihood Ratio (LR) estimator, i.e.*

$$\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X)\right] = (\mathbb{E}_{P^\theta}\left[f(X)W^{\theta_1}(X_1)\right], \ldots, \mathbb{E}_{P^\theta}\left[f(X)W^{\theta_n}(X_n)\right])^T \tag{5.9}$$

(b) In the special case when $f(X) = X_l$ for some $1 \leq l \leq n$, we have

$$\nabla_\theta \mathbb{E}_{P^\theta}\left[f(X_l)\right] = (\mathbb{E}_{P^{\theta}_{1:l}}\left[f(X_l)W^{\theta_1}(X_1)\right], \ldots, \mathbb{E}_{P^{\theta}_{1:l}}\left[f(X_l)W^{\theta_l}(X_l)\right], 0, \ldots, 0)^T \tag{5.10}$$

where $P^{\theta}_{1:l}$ is the marginal distribution for $(X_1, \ldots, X_l)$, i.e . $\mathbb{E}_{P^\theta}\left[f(X_l)W^{\theta_i}(X_i)\right] = 0$ for all X_i not an ancestor of X_l, $i \notin \rho_l$, where W is the score function defined in (5.2) and (5.4).

5.2 Fisher Information Matrices and Cramer-Rao Type Bounds for PGMs

Definition 5.3 *For a parametric PGM, P^θ, as defined in* (5.1)*, we define the* ***Fisher information matrix (FIM)*** *of the PGM by*

$$\mathcal{I}(P^\theta) = \mathbb{E}_{P^\theta}\left[W^\theta (W^\theta)^T\right] \tag{5.11}$$

Furthermore, the FIM satisfies the property given by the following Lemma:

Lemma 5.1 *For any PGM $p(x|\theta) = \prod_{i=1}^n p^{\theta_i}(x_i|x_{\pi_i})$ that satisfies Assumption 2, and the FIM of P^θ defined in* (5.11)*, we have*

$$\mathcal{I}(P^\theta) = diag(\mathbb{E}_{P^\theta_{\pi_1}}\left[\mathcal{I}(P^{\theta_1}_{1|\pi_1})\right], \mathbb{E}_{P^\theta_{\pi_2}}\left[\mathcal{I}(P^{\theta_2}_{2|\pi_2})\right], \cdots, \mathbb{E}_{P^\theta_{\pi_n}}\left[\mathcal{I}(P^{\theta_n}_{n|\pi_n})\right]) \tag{5.12}$$

where $\mathcal{I}(P^{\theta_i}_{i|\pi_i}) = \mathbb{E}_{P^\theta_{i|\pi_i}}\left[W^{\theta_i}(W^{\theta_i})^T\right]$ is the FIM of $p^{\theta_i}(x_i|x_{\pi_i})$ with given $X_{\pi_i} = x_{\pi_i}$, and $P^\theta_{\pi_i}$ is the marginal distribution of P^θ for X_{π_i}.

Proof of Lemma 5.1: For the FIM of P^θ, which defined by $\prod_{k=1}^n p^{\theta_k}(x_k|x_{\pi_k})$, $\mathcal{I}(P^\theta) = \mathbb{E}_{P^\theta}\left[W^\theta(W^\theta)^T\right]$, we have

$$\begin{aligned}
\mathcal{I}_{ii}(P^\theta) &= \mathbb{E}_{P^\theta}\left[\nabla_{\theta_i}(\sum_{k=1}^n \log p^{\theta_k}(x_k|x_{\pi_k}))\nabla_{\theta_i}(\sum_{i=k}^n \log p^{\theta_k}(x_k|x_{\pi_k}))^T\right] \\
&= \mathbb{E}_{P^\theta}\left[\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_i})\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_i})^T\right] \\
&= \mathbb{E}_{P^\theta_{\pi_i}}\left[\mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_i})\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_j})^T\right]\right] \\
&= \mathbb{E}_{P^\theta_{\pi_i}}\left[\mathcal{I}(P^{\theta_i}_{i|\pi_i})\right]
\end{aligned} \tag{5.13}$$

where $\mathcal{I}_{ij}(P^\theta)$ is a sub-matrix on $\mathcal{I}(P^\theta)$ that corresponding to θ_i, θ_j, $\mathcal{I}(P^{\theta_i}_{i|\pi_i})$ is the FIM of $p^{\theta_i}(x_i|x_{\pi_i})$ with fixed x_{π_i}, and for $i \neq j$, since $\mathbb{E}_{P^\theta}\left[W^{\theta_i}\right] = 0$, without loss of generality, assume $i < j$, we have

$$\begin{aligned}
& \mathcal{I}_{ij}(P^\theta) \\
= \ & \mathbb{E}_{P^\theta}\left[\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_i})\nabla_{\theta_j}\log p^{\theta_j}(x_j|x_{\pi_j})^T\right] \\
= \ & \mathbb{E}_{P^\theta_{1:i-1}}\left[\mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_i})\mathbb{E}_{P^\theta_{i+1:j-1}}\left[\mathbb{E}_{P^{\theta_j}|x_{\pi_j}}\left[\nabla_{\theta_j}\log p^{\theta_j}(x_j|x_{\pi_j})^T\right]\right]\right]\right] \\
= \ & \mathbb{E}_{P^\theta_{1:i-1}}\left[\mathbb{E}_{P^{\theta_i}_{i|\pi_i}}\left[\nabla_{\theta_i}\log p^{\theta_i}(x_i|x_{\pi_i})\mathbb{E}_{P^\theta_{i+1:j-1}}[0]\right]\right] \\
= \ & 0
\end{aligned} \tag{5.14}$$

therefore

$$\mathcal{I}(P^\theta) = \begin{pmatrix} \mathbb{E}_{P^\theta_{\pi_1}}\left[\mathcal{I}(P^{\theta_1}_{1|\pi_1})\right] & & & \\ & \mathbb{E}_{P^\theta_{\pi_2}}\left[\mathcal{I}(P^{\theta_2}_{2|\pi_2})\right] & & \emptyset \\ \emptyset & & \ddots & \\ & & & \mathbb{E}_{P^\theta_{\pi_n}}\left[\mathcal{I}(P^{\theta_n}_{n|\pi_n})\right] \end{pmatrix} \tag{5.15}$$

Example:

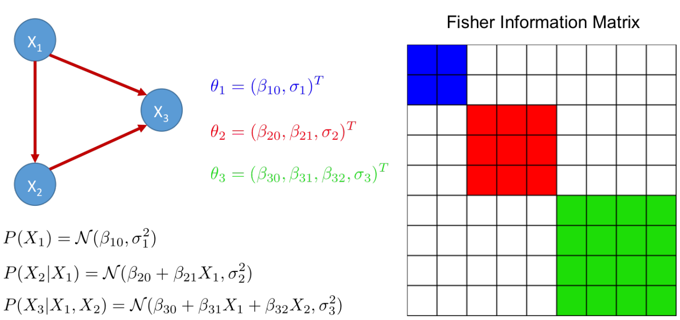

Figure 26. An example of simple Gaussian Bayesian Network with its parameters and corresponding block-diagonal structure of FIM.

Especially, for some models with parameters differ by orders of magnitude, a reasonable option for carrying out sensitivity analysis is to perform perturbations which are proportional to the parameter magnitude. This can be carried out by perturbing the logarithm of the model parameters instead of the parameters itself.

Using the chain rule $\nabla_{\log\theta} f(\theta) = \nabla_\theta f(\theta) \cdot \nabla_{\log\theta}\theta = \theta \cdot \nabla_\theta f(\theta)$ where '$\cdot$' is defined as the element by element multiplication, then we obtain the logarithmically-scaled FIM:

$$\mathcal{I}_{ii}(P^{\log\theta}) = \theta_i \mathcal{I}_{ii}(P^\theta)\theta_i^T \tag{5.16}$$

Moreover, based on the FIM we defined on PGMs, we can have Cramer-Rao type bounds for PGMs, and the results are concluded in the following Theorem:

Theorem 5.4 *(a) For any PGM $p(x|\theta) = \prod_{i=1}^n p^{\theta_i}(x_i|x_{\pi_i})$ with a given QoI $f(X)$, the gradient based sensitivity index $\nabla_\theta \mathbb{E}_{P^\theta}[f]$ satisfies*

$$|v^T \nabla_\theta \mathbb{E}_{P^\theta}[f]| \leq \sqrt{Var_{P^\theta}(f)}\sqrt{v^T \mathcal{I}(P^\theta)v} \tag{5.17}$$

where $v \in \mathbb{R}^n$, and $\mathcal{I}(P^\theta)$ is the Fisher information matrix for the PGM, P^θ, defined in (5.11).

(b) If P^θ satisfies the Assumption 1 and $f(X) = f(X_l)$, then

$$|v^T \nabla_\theta \mathbb{E}_{P^\theta}[f(X_l)]| \leq \sqrt{Var_{P_l^\theta}(f)}\sqrt{v^T \mathcal{I}_{1:l}(P^\theta)v} \tag{5.18}$$

where P_l^θ is the marginal distribution of X_l and

$$\mathcal{I}_{1:l}(P^\theta) = \begin{pmatrix} \mathbb{E}_{P^\theta_{\pi_1}}\left[\mathcal{I}(P^{\theta_1}_{1|\pi_1})\right] & & & & \emptyset & \\ & \ddots & & & & \\ & & \mathbb{E}_{P^\theta_{\pi_l}}\left[\mathcal{I}(P^{\theta_l}_{l|\pi_l})\right] & & & \\ & & & 0 & & \\ \emptyset & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \tag{5.19}$$

The proof of Theorem 5.4 (a) is given in [29] (Theorem 2.13) and Theorem 5.4 (b) is a direct derivation with Lemma 5.1.

5.3 Connection with The Model-form UQ Indices

If we consider the ambiguity set $\mathcal{Q}$ defined by

$$\mathcal{Q} := \{\text{all PGMs } Q : q(x) = p(x|\theta + \epsilon v) \text{ with } v \in \mathbb{R}^k \text{ and } \epsilon \in \mathbb{R}\} \tag{5.20}$$

where $p(x|\theta)$ is the density of a parametric PGM P^θ as defined in (5.1), then in the case that $\epsilon \to 0$, we have the following Theorem that recover FIM as the Hessian of KL divergence $R(P^{\theta+v}||P^\theta)$:

Theorem 5.5 *(a) Let P^θ be a parametric family of probability measures, where $\theta \in \mathbb{R}^k$, and let $v \in \mathbb{R}^k$, then*

$$\eta = R(P^{\theta+v}||P^\theta) = \frac{1}{2} v^T \mathcal{I}(P^\theta) v + \mathcal{O}(|v|^3) \tag{5.21}$$

where $\mathcal{I}(P^\theta)$ is the Fisher Information Matrix (FIM) given by

$$\mathcal{I}(P^\theta) = \int \nabla_\theta \log p^\theta(\omega) (\nabla_\theta \log p^\theta(\omega))^T P^\theta(d\omega)\,. \tag{5.22}$$

(b) Therefore, we have

$$I^{\pm}(f(X), P; \mathcal{D}^{R(P^{\theta+\epsilon v}||P^\theta)}) = \pm\sqrt{Var_P(f)}\sqrt{v^T \mathcal{I}(P^\theta) v}\ \epsilon + \mathcal{O}(|\epsilon|^{3/2}) \tag{5.23}$$

The proof of Theorem 5.5 (a) is stated in [29] (Lemma 2.21), and Theorem 5.5 (b) can be easily derived using the linearization form of our UQ index when $\eta \to 0^+$ which is proved in [29, 75], i.e.

Theorem 5.6 (Linearization of UQ indices) *Let P be a probability measure and let $f(X)$ be such that its MGF is finite in a neighborhood of the original. Considering any Q in the family of probability measures $\mathcal{D}^\eta = \{Q : R(Q||P) \le \eta\}$, then when $\eta \to 0^+$, the UQ indices defined as* (3.84) *or equivalently* (3.3) *satisfy*

$$I^{\pm}(f, P; \eta) = \pm\sqrt{2Var_P(f)}\eta^{1/2} + \frac{1}{3}\frac{\kappa_3(f)}{Var_P(f)}\eta + \mathcal{O}(\eta^{3/2}) \tag{5.24}$$

Remark: Based on Lemma 5.1, we can write all the eigenvalues and eigenvectors of $\mathcal{I}(P^\theta)$ as λ_{il} and $e^{il} = (0, \ldots, e_{il}^T, \ldots, 0)^T$, where λ_{il} and e_{il} are the corresponding eigenvalue and eigenvector of $\mathcal{I}_{ii}(P^\theta)$. Then by Theorem 5.5, we have

$$R(P^{\theta+v}||P^\theta) = \sum_{i=1}^{n} \mathbb{E}_{P^\theta_{\pi_i}} \left[R(P^{\theta_i+v_i}_{i|\pi_i}||P^{\theta_i}_{i|\pi_i}) \right] = \frac{1}{2} v^T \mathcal{I}(P^\theta) v + \mathcal{O}(|v|^3) \tag{5.25}$$

thus, for $v^i = (0, \ldots, v_i, \ldots, 0)$, we have

$$\begin{aligned} R(P^{\theta+v^i}||P^\theta) &= \mathbb{E}_{P^\theta_{\pi_i}} \left[R(P^{\theta_i+v_i}_{i|\pi_i}||P^{\theta_i}_{i|\pi_i}) \right] \\ &= \frac{1}{2} v_i^T \mathcal{I}_{ii}(P^\theta) v_i + \mathcal{O}(|v_i|^3) \\ &= \mathbb{E}_{P^\theta_{\pi_i}} \left[\frac{1}{2} v_i^T \mathcal{I}(P^{\theta_i}_{i|\pi_i}) v_i \right] + \mathcal{O}(|v_i|^3) \end{aligned} \tag{5.26}$$

Especially, when $v^i = e^{il}$, an eigenvector of $\mathcal{I}_{ii}(P^\theta)$ in (5.15), we have

$$R(P^{\theta+e^{il}}||P^\theta) = \frac{\lambda_{ij}}{2} (e^{il})^T e^{il} + \mathcal{O}(|e^{il}|^3) \tag{5.27}$$

and the eigenvetor with the largest eigenvalue is corresponded to the most influential direction/components for P^θ.

5.4 Chest Clinic Example

Here we apply the sensitivity analysis methods proposed above to a parametric PGM, which Lauritzen and Spiegelhalter proposed in [77] by fictitious qualitative medical 'knowledge':

"Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to some geometric region X may increase the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does

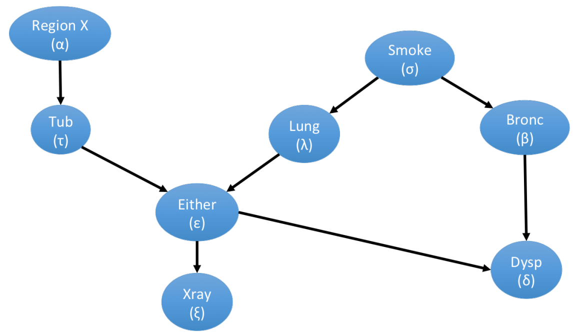

Figure 27. Chest clinic example.

the presence or absence of dyspnoea." From the causal network, we could get the joint distribution of our model P, with density

$$p(\xi, \epsilon, \tau, \lambda, \alpha, \sigma) = p(\xi|\epsilon)p(\epsilon|\tau, \lambda)p(\tau|\alpha)p(\alpha)p(\lambda|\sigma)p(\sigma)$$

where each random variable following a Bernoulli distribution given by the table below, and assume for the node α 'visit region X?', let $p(a) = p_\alpha$ to stand for $Pr(\alpha = a)$ with parameter p_α, similarly, t stands for the presence of 'tuberculosis' with parameters p_{τ_1}, p_{τ_0}, which are corresponded with the cases $\alpha = a$ and $\alpha = \bar{a}$; s, 'smoker', with parameter p_σ; 1, 'lung cancer', with parameters p_{λ_1}, p_{λ_0}; b, 'bronchitis'; e, 'lung cancer or bronchitis'; x, 'positive X-ray'; and d, 'dyspnoea', with parameters p_{ξ_1}, p_{ξ_0}.

Table 4. Conditional probability table given in [77]

α :	$p_\alpha = 0.01$	σ :	$p_\sigma = 0.5$
τ :	$p_{\tau_0} = 0.01$	λ :	$p_{\lambda_0} = 0.01$
	$p_{\tau_1} = 0.05$		$p_{\lambda_1} = 0.1$
ξ :	$p_{\xi_0} = 0.05$		
	$p_{\xi_1} = 0.98$		

Note that the CPDs of all the nodes are fixed to be Bernoulli distributed in nature, therefore, we can do sensitivity analysis for the model by looking at the FIM for the parameters in P with normal and logarithmically scale, then by (5.13) and (5.16), we can compute the results as shown in Figure 28.

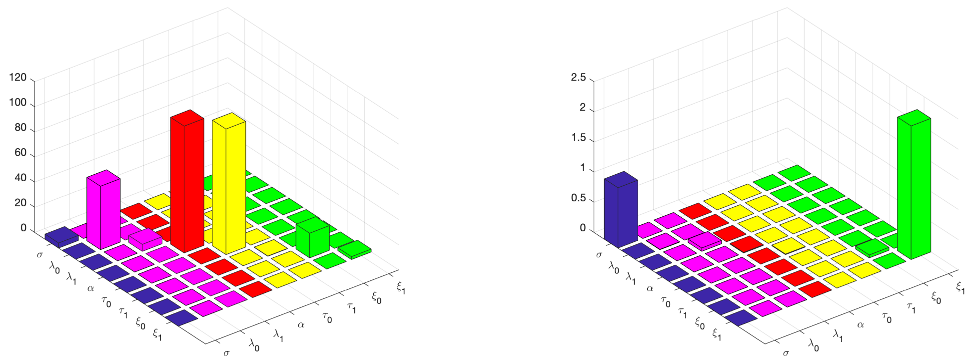

Figure 28. normal/logarithmically-scaled FIM.

Moreover, if we are interested in $f(X)$, which is define by

$$f = \begin{cases} 1 & \text{if } \xi = x \\ 0 & \text{if } \xi = \bar{x} \end{cases} \tag{5.28}$$

then we have

$$\mathbb{E}_P\left[f(X)\right] = \sum_{(\epsilon,\tau,\lambda,\alpha,\sigma)} p(\xi = x|\epsilon)p(\epsilon|\tau,\lambda)p(\tau|\alpha)p(\alpha)p(\lambda|\sigma)p(\sigma) = 0.1103 \tag{5.29}$$

$$\mathrm{Var}_P(f) = \mathbb{E}_P\left[f(X)\right]\left(1 - \mathbb{E}_P\left[f(X)\right]\right) = 0.0981 \tag{5.30}$$

and by (5.6), we could compute the LR estimators for the gradient based sensitivity index $\nabla_\theta \mathbb{E}_{P^\theta}\left[f\right]$, which can be bounded by the Cramer-Rao type bounds based on Theorem 5.4 as shown in Figure 29.

Similarly, we can also consider the logarithmically-scaled sensitivity index $\nabla_{\log\theta}\mathbb{E}_{P^\theta}\left[f\right] = \theta \cdot \nabla_\theta \mathbb{E}_{P^\theta}\left[f\right]$ as shown in Figure 30.

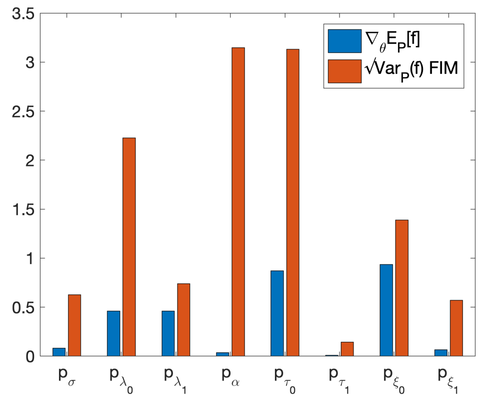

Figure 29. Likelihood Ratio (LR) estimators (5.6) **and Cramer-Rao type bounds** (5.18) **for the gradient based sensitivity index** $\nabla_\theta \mathbb{E}_{P^\theta}[f]$. **The results are consistent with our finding in Theorem 5.4.**

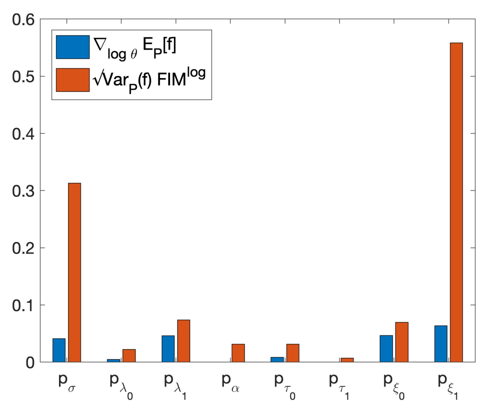

Figure 30. LR estimators and Cramer-Rao type bounds for the logarithmically-scaled gradient based sensitivity index in $\nabla_{\log\theta} \mathbb{E}_{P^\theta}[f]$ **(where** FIM^{log} **is the logarithmically-scaled FIM given by** (5.16)**).**

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 1

A.1 Derivation of the Langmuir bimolecular adsorption model

By considering competitive adsorption of hydrogen and oxygen on a catalyst surface in the form of (18), the net rates of adsorption can be obtained

$$r_{H_2} = r_{H_2}^{ads} - r_{H_2}^{des} = k_{H_2}^{ads} P_{H_2} C_*^2 - k_{H_2}^{des} C_{H*}^2,$$
$$\tag{A1-1}$$
$$r_{O_2} = r_{O_2}^{ads} - r_{O_2}^{des} = k_{O_2}^{ads} P_{O_2} C_*^2 - k_{O_2}^{des} C_{O*}^2,$$

where

$$K_{H_2} = \frac{k_{H_2}^{ads}}{k_{H_2}^{des}}, \qquad K_{O_2} = \frac{k_{O_2}^{ads}}{k_{O_2}^{des}}, \tag{A1-2}$$

and k^{ads} and k^{des} are the adsorption and desorption rate constants, r^{ads} and r^{des} represent the adsorption and desorption rate, P is the partial pressure and the H and O denote hydrogen and oxygen, respectively. The site balance gives

$$C_t = C_* + C_{H^*} + C_{O^*}, \tag{A1-3}$$

where C_t, C_*, C_{H^*} and C_{O^*} are the concentrations of total active sites, vacant sites and occupied sites by hydrogen and oxygen, respectively [35]. For simplicity, the chemical reaction between hydrogen and oxygen atoms is not accounted for here.

By defining the hydrogen and oxygen coverages

$$\hat{\theta}_{H^*} = \frac{C_{H^*}}{C_t}, \qquad \hat{\theta}_{O^*} = \frac{C_{O^*}}{C_t}, \tag{A1-4}$$

and considering the site balance of (A1-3), the set of governing ordinary differential equations are formulated by

$$\begin{aligned} \frac{d\hat{\theta}_{H^*}}{dt} &= k^{ads}_{H^*} P_{H_2}(1 - \hat{\theta}_{H^*} - \hat{\theta}_{O^*})^2 - k^{des}_{H_2}\hat{\theta}^2_{H^*}, \quad \theta^0_{H^*} = \hat{\theta}_{H^*}(0), \\ \frac{d\hat{\theta}_{O^*}}{dt} &= k^{ads}_{O_2} P_{O_2}(1 - \hat{\theta}_{H^*} - \hat{\theta}_{O^*})^2 - k^{des}_{O_2}\hat{\theta}^2_{O^*}, \quad \theta^0_{O^*} = \hat{\theta}_{O^*}(0), \end{aligned} \tag{A1-5}$$

where $\theta^0_{H^*}$ and $\theta^0_{O^*}$ represent the initial hydrogen and oxygen coverages, respectively.

Equilibrium hydrogen and oxygen coverages can be calculated as

$$\begin{aligned} \hat{\theta}_{H^*} &= \frac{(K_{H_2}P_{H_2})^{\frac{1}{2}}}{1 + (K_{H_2}P_{H_2})^{\frac{1}{2}} + (K_{O_2}P_{O_2})^{\frac{1}{2}}}, \\ \hat{\theta}_{O^*} &= \frac{(K_{O_2}P_{O_2})^{\frac{1}{2}}}{1 + (K_{H_2}P_{H_2})^{\frac{1}{2}} + (K_{O_2}P_{O_2})^{\frac{1}{2}}}, \end{aligned} \tag{A1-6}$$

where the equilibrium constants can be described as follows,

$$\begin{aligned} K_{H_2} &= \exp\Big(- \frac{\Delta G_{H_2 \to 2H^*}}{k_B T} \Big) P_o^{-1}, \\ K_{O_2} &= \exp\Big(- \frac{\Delta G_{O_2 \to 2O^*}}{k_B T} \Big) P_o^{-1}, \end{aligned} \tag{A1-7}$$

and $k_B = 1.38065\, J/K$ is the Boltzmann constant, T is temperature at which the adsorption occurs, and we set it to be 298.15 K in this chapter. $\Delta G_{H_2 \to 2H^*}$ and $\Delta G_{O_2 \to 2O^*}$ denote the hydrogen and oxygen Gibbs free energies of adsorption, respectively. P_o is the standard state pressure. In this system the standard state pressure is taken to be the total pressure. The Gibbs free energies of adsorption are

$$\begin{aligned} \Delta G_{H_2 \to 2H^*} &= \Delta H_{H_2 \to 2H^*} - T\Delta S_{H_2 \to 2H^*}, \\ \\ \Delta G_{O_2 \to 2O^*} &= \Delta H_{O_2 \to 2O^*} - T\Delta S_{O_2 \to 2O^*}, \end{aligned} \tag{A1-8}$$

where $\Delta H_{H_2\to 2H^*}$ and $\Delta H_{O_2\to 2O^*}$ denote enthalpies of adsorption, and $\Delta S_{H_2\to 2H^*}$ and $\Delta S_{O_2\to 2O^*}$ are the entropies of adsorption.

The enthalpy of adsorption can also be obtained as

$$\begin{aligned}\Delta H_{H_2\to 2H^*} &= 2H_{H^*} - H_{H_2} - 2E_{slab},\\ H_{O_2\to 2O^*} &= 2H_{O^*} - H_{O_2} - 2E_{slab},\end{aligned} \tag{A1-9}$$

where H_{H_2} and H_{O_2} are the enthalpies of H_2 and O_2 in the gas phase, and H_{H^*} and H_{O^*} denote the enthalpies of H^* and O^* metal-adsorbate complexes, respectively. The energy of the metal slab (E_{slab}) is the same as its enthalpy as there are no pressure-volume effects. The enthalpies of H^* and O^* on the surface in (A1-7) can be computed by

$$\begin{aligned}H_{H^*} &= E_{H*}^{DFT} + \sum_{i=1}^{3}\Big(\frac{h\nu_{H^*}^i}{2} + \frac{h\nu_{H^*}^i}{\exp\big(\frac{h\nu_{H^*}^i}{k_BT}\big) - 1}\Big),\\ H_{O^*} &= E_{O*}^{DFT} + \sum_{i=1}^{3}\Big(\frac{h\nu_{O^*}^i}{2} + \frac{h\nu_{O^*}^i}{\exp\big(\frac{h\nu_{O^*}^i}{k_BT}\big) - 1}\Big),\end{aligned} \tag{A1-10}$$

.

In above equation E_{H*}^{DFT} and E_{O*}^{DFT} denote the electronic energies of the hydrogen and oxygen adsorbate-metal complex, as calculated by density functional theory (DFT). For $i = 1, \ldots, 3$ $\nu_{H^*}^i$ and $\nu_{O^*}^i$ represent the harmonic vibrational frequencies of adsorbed species on Pt(111) in the hollow site and $h = 6.626\times10^{-34} J.s$ is Planck's constant.

The enthalpies of the molecular gas species are calculated with the following thermodynamic equations

$$H_{H_2} = \frac{7}{2}k_BT + E_{H_2}^{DFT} + \left(\frac{h\nu_{H_2}}{2} + \frac{h\nu_{H_2}}{\exp\left(\frac{h\nu_{H_2}}{k_BT}\right) - 1}\right),$$

$$H_{O_2} = \frac{7}{2}k_BT + E_{O_2}^{DFT} + \left(\frac{h\nu_{O_2}}{2} + \frac{h\nu_{O_2}}{\exp\left(\frac{h\nu_{O_2}}{k_BT}\right) - 1}\right), \tag{A1-11}$$

.

where $E_{H_2}^{DFT}$ and $E_{O_2}^{DFT}$ are the DFT electronic energies, and ν_{H_2} and ν_{O_2} are the respective diatomic fundamental frequencies.

The entropies of adsorption in (A1-8) can then be calculated by

$$\begin{aligned}\Delta S_{H_2\to 2H^*} = \Delta S_{H_2\to 2H^*}^{vib} - \frac{7}{2}k_B - k_B\ln\left[\left(\frac{2\pi m_{H_2}k_BT}{h^2}\right)^{\frac{3}{2}}\frac{k_BT}{P}\right] \\ -k_B\ln\left(\tfrac{T}{2\Theta_{R,H_2}}\right) - k_B\ln(\omega_{e1,H_2}), \\ \Delta S_{O_2\to 2O^*} = \Delta S_{O_2\to 2O^*}^{vib} - \frac{7}{2}k_B - k_B\ln\left[\left(\frac{2\pi m_{O_2}k_BT}{h^2}\right)^{\frac{3}{2}}\frac{k_BT}{P}\right] \\ -k_B\ln\left(\tfrac{T}{2\Theta_{R,O_2}}\right) - k_B\ln(\omega_{e1,O_2}),\end{aligned} \tag{A1-12}$$

where m is the molecular mass, and ΔS^{vib} is the change in vibrational contribution to entropy. The rotational temperatures, denoted by Θ_R, are 85.3K and and 2.07K for H_2 and O_2 respectively [88]. The degeneracy of their first electronic energy levels, denoted by ω_{e1}, are 1 and 3.

Finally from (A1-8)-(A1-12) we conclude

$$\begin{aligned}\Delta G_{H_2\to 2H^*} = \Delta H_{H_2\to 2H*}^{DFT} + \Delta H_{H_2\to 2PtH}^{vib} - \frac{7}{2}k_BT - \\ T\left(\Delta S_{H_2\to 2PtH}^{vib} - k_B\ln\left[\left(\frac{2\pi m_{H_2}k_BT}{h^2}\right)^{\frac{3}{2}}\frac{k_BT}{P}\right] - \frac{7}{2}k_B - k_B\ln\left[\frac{T}{2\Theta_{R,H_2}}\right]\right), \\ \Delta G_{O_2\to 2O^*} = \Delta H_{O_2\to 2O*}^{DFT} + \Delta H_{O_2\to 2PtO}^{vib} - \frac{7}{2}k_BT - T\left(\Delta S_{O_2\to 2PtO}^{vib} - \right. \\ \left. k_B\ln\left[\left(\frac{2\pi m_{O_2}k_BT}{h^2}\right)^{\frac{3}{2}}\frac{k_BT}{P}\right] - \frac{7}{2}k_B - k_B\ln\left[\frac{T}{2\Theta_{R,O_2}}\right] - k_B\ln[3]\right),\end{aligned} \tag{A1-13}$$

then by grouping terms into those that are involved in the scaling relations, ΔE_H and ΔE_O, and those that are not a function of metal surface, we obtain

$$
\begin{aligned}
\Delta G_{H_2\rightarrow 2H^*} &= -2\Delta E_H + \Bigg[D_{0,H} + \Delta G^{vib}_{H_2\rightarrow 2PtH} + \\
&\qquad k_BT\Bigg(\ln\Bigg[\Big(\frac{2\pi m_{H_2}k_BT}{h^2}\Big)^{\frac{3}{2}}\frac{k_BT}{P}\Bigg] + \ln\Bigg[\frac{T}{2\Theta_{R,H_2}}\Bigg]\Bigg)\Bigg],\\
\Delta G_{O_2\rightarrow 2O^*} &= -2\Delta E_O + \Bigg[D_{0,O} + \Delta G^{vib}_{O_2\rightarrow 2PtO} + \\
&\qquad k_BT\Bigg(\ln\Bigg[\Big(\frac{2\pi m_{O_2}k_BT}{h^2}\Big)^{\frac{3}{2}}\frac{k_BT}{P}\Bigg] + \ln\Bigg[\frac{T}{2\Theta_{R,O_2}}\Bigg] + \ln[3]\Bigg)\Bigg].
\end{aligned}
\tag{A1-14}
$$

In Equations (A1-13)-(A1-14) the vibrational contributions to ΔG are assumed to be independent of the metal substrate. Frequencies calculated for atomic hydrogen and oxygen on platinum are used in calculating vibrational contributions to Gibbs Energy for all adsorbate-metal systems.

A.2 Derivations of the LSIs

The LSIs can be derived by direct differentiation of coverages with respect to the electronic part of the binding energies using the chain rule and recognizing that $\Delta E_i = -\frac{1}{2}(\Delta H^{DFT}_{i_2\rightarrow 2i*} + D_0)$,

$$
\begin{aligned}
\frac{\partial\hat{\theta}_{H^*}}{\partial(\Delta E_H)} &= \frac{\partial\hat{\theta}_{H^*}}{\partial K_{H_2}}\frac{\partial K_{H_2}}{\partial(\Delta G_{H_2\rightarrow 2H^*})}\frac{\partial(\Delta G_{H_2\rightarrow 2H^*})}{\partial(\Delta E_H)},\\
\frac{\partial\hat{\theta}_{H^*}}{\partial(\Delta E_O)} &= \frac{\partial\hat{\theta}_{H^*}}{\partial K_{O_2}}\frac{\partial K_{O_2}}{\partial(\Delta G_{O_2\rightarrow 2O^*})}\frac{\partial(\Delta G_{O_2\rightarrow 2O^*})}{\partial(\Delta E_O)},\\
\frac{\partial\hat{\theta}_{O^*}}{\partial(\Delta E_H)} &= \frac{\partial\hat{\theta}_{O^*}}{\partial K_{H_2}}\frac{\partial K_{H_2}}{\partial(\Delta G_{H_2\rightarrow 2H^*})}\frac{\partial(\Delta G_{H_2\rightarrow 2H^*})}{\partial(\Delta E_H)},\\
\frac{\partial\hat{\theta}_{O^*}}{\partial(\Delta E_O)} &= \frac{\partial\hat{\theta}_{O^*}}{\partial K_{O_2}}\frac{\partial K_{O_2}}{\partial(\Delta G_{O_2\rightarrow 2O^*})}\frac{\partial(\Delta G_{O_2\rightarrow 2O^*})}{\partial(\Delta E_O)},
\end{aligned}
\tag{A2-1}
$$

where

$$\frac{\partial \hat{\theta}_{H^*}}{\partial K_{H_2}} = \frac{P_{H_2}\Big(1+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)}{2(K_{H_2}P_{H_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2},$$

$$\frac{\partial \hat{\theta}_{H^*}}{\partial K_{O_2}} = -\frac{P_{O_2}(K_{H_2}P_{H_2})^{\frac{1}{2}}}{2(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2},$$

$$\frac{\partial \hat{\theta}_{O^*}}{\partial K_{H_2}} = -\frac{P_{H_2}(K_{O_2}P_{O_2})^{\frac{1}{2}}}{2(K_{H_2}P_{H_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}, \tag{A2-2}$$

$$\frac{\partial \hat{\theta}_{O^*}}{\partial K_{O_2}} = \frac{P_{O_2}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}\Big)}{2(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2},$$

and

$$\frac{\partial K_{H_2}}{\partial(\Delta G_{H_2\to 2H^*})} = -\frac{1}{RT}\exp\Big(-\frac{\Delta G_{H_2\to 2H^*}}{RT}\Big),$$

$$\frac{\partial K_{O_2}}{\partial(\Delta G_{O_2\to 2O^*})} = -\frac{1}{RT}\exp\Big(-\frac{\Delta G_{O_2\to 2O^*}}{RT}\Big), \tag{A2-3}$$

and

$$\frac{\partial(\Delta G_{H_2\to 2H^*})}{\partial(\Delta E_H)} = \frac{\partial(\Delta G_{O_2\to 2O^*})}{\partial(\Delta E_O)} = -2. \tag{A2-4}$$

Then the LSIs with respect to ΔE_H are formulated accordingly

$$\begin{aligned}
& S_H^H(\Delta E_H, \Delta E_O) \\
= \;& \frac{\partial(\ln\hat{\theta}_{H^*})}{\partial(\Delta E_H)} = \frac{1}{\hat{\theta}_{H^*}}\frac{\partial\hat{\theta}_{H^*}}{\partial K_{H_2}}\frac{\partial K_{H_2}}{\partial(\Delta G_{H_2\to 2H^*})}\frac{\partial(\Delta G_{H_2\to 2H^*})}{\partial(\Delta E_H)} \\
= \;& \frac{2}{\hat{\theta}_{H^*}}\frac{P_{H_2}\Big(1+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)}{2(K_{H_2}P_{H_2})^{\frac{1}{2}}\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}\frac{1}{k_BTP_o}\exp\Big(-\frac{\Delta G_{H_2\to 2H^*}}{k_BT}\Big) \\
= \;& \frac{1+(P_{O_2}K_{O_2})^{\frac{1}{2}}}{k_BT\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)} \\
= \;& \frac{P_o^{\frac{1}{2}}+P_{O_2}^{\frac{1}{2}}\exp\Big(\frac{\Delta E_O - C_{O_2}}{k_BT}\Big)}{k_BT\Bigg(P_o^{\frac{1}{2}}+P_{H_2}^{\frac{1}{2}}\exp\Big(\frac{\Delta E_H - C_{H_2}}{k_BT}\Big)+P_{O_2}^{\frac{1}{2}}\exp\Big(\frac{\Delta E_O - C_{O_2}}{k_BT}\Big)\Bigg)}
\end{aligned} \tag{A2-5}$$

$$
\begin{aligned}
& S_H^O(\Delta E_H, \Delta E_O) \\
= \ & \frac{\partial(\ln \hat{\theta}_{O^*})}{\partial(\Delta E_H)} = \frac{1}{\hat{\theta}_{O^*}} \frac{\partial \hat{\theta}_{O^*}}{\partial K_{H_2}} \frac{\partial K_{H_2}}{\partial(\Delta G_{H_2 \to 2H^*})} \frac{\partial(\Delta G_{H_2 \to 2H^*})}{\partial(\Delta E_H)} \\
= \ & -\frac{2}{\hat{\theta}_{O^*}} \frac{P_{H_2}(K_{O_2}P_{O_2})^{\frac{1}{2}}}{2(K_{H_2}P_{H_2})^{\frac{1}{2}}\Big(1 + (K_{H_2}P_{H_2})^{\frac{1}{2}} + (K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2} \frac{1}{k_B T P_o} \exp\Big(-\frac{\Delta G_{H_2 \to 2H^*}}{k_B T}\Big) \\
= \ & -\frac{(P_{H_2}K_{H_2})^{\frac{1}{2}}}{k_B T\Big(1 + (K_{H_2}P_{H_2})^{\frac{1}{2}} + (K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)} \\
= \ & -\frac{P_{H_2}^{\frac{1}{2}} \exp\Big(\frac{\Delta E_H - C_{H_2}}{k_B T}\Big)}{k_B T\left(P_o + P_{H_2}^{\frac{1}{2}} \exp\Big(\frac{\Delta E_H - C_{H_2}}{k_B T}\Big) + P_{O_2}^{\frac{1}{2}} \exp\Big(\frac{\Delta E_O - C_{O_2}}{k_B T}\Big)\right)}
\end{aligned}
\tag{A2-6}
$$

where

$$
\begin{aligned}
C_{H_2} &= \frac{1}{2}\Delta G_{H_2 \to 2H^*} + \Delta E_H, \\
C_{O_2} &= \frac{1}{2}\Delta G_{O_2 \to 2O^*} + \Delta E_O,
\end{aligned}
\tag{A2-7}
$$

are both constants.

And the relevant LSIs with respect to ΔE_O can also be computed similarly using the following equations,

$$
\begin{aligned}
S_O^H = \frac{\partial(\ln \hat{\theta}_{H^*})}{\partial(\Delta E_O)} = \frac{1}{\hat{\theta}_{H^*}} \frac{\partial \hat{\theta}_{H^*}}{\partial K_{O_2}} \frac{\partial K_{O_2}}{\partial(\Delta G_{O_2 \to 2O^*})} \frac{\partial(\Delta G_{O_2 \to 2O^*})}{\partial(\Delta E_O)}, \\
S_O^O = \frac{\partial(\ln \hat{\theta}_{O^*})}{\partial(\Delta E_O)} = \frac{1}{\hat{\theta}_{O^*}} \frac{\partial \hat{\theta}_{O^*}}{\partial K_{O_2}} \frac{\partial K_{O_2}}{\partial(\Delta G_{O_2 \to 2O^*})} \frac{\partial(\Delta G_{O_2 \to 2O^*})}{\partial(\Delta E_O)},
\end{aligned}
\tag{A2-8}
$$

For the CLSIs in the deterministic case, we have

$$
\frac{\partial(\Delta E_O)}{\partial(\Delta E_H)} = \frac{1}{\frac{\partial(\Delta E_H)}{\partial(\Delta E_O)}} = a,
\tag{A2-9}
$$

by the correlation of (24). Then the relevant CLSIs with respect to ΔE_H can be obtained considering the parameter correlation in the chain rule,

$$
\begin{aligned}
& S^{H}_{H,\,corr}(\Delta E_H) \\
= \;& \left[\frac{\partial(\ln\hat{\theta}_{H^*})}{\partial(\Delta E_H)}\right]_{corr} = \frac{1}{\hat{\theta}_{H^*}}\left[\frac{\partial\hat{\theta}_{H^*}}{\partial(\Delta E_H)} + \frac{\partial\hat{\theta}_{H^*}}{\partial(\Delta E_O)}\frac{\partial(\Delta E_O)}{\partial(\Delta E_H)}\right] \\
= \;& \frac{2}{\hat{\theta}_{H^*}}\left[\frac{P_{H_2}(K_{H_2}P_{H_2})^{-\frac{1}{2}}\Big(1+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)}{2\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}\frac{1}{k_BTP_o}\exp\Big(-\frac{\Delta G_{H_2\to 2H^*}}{k_BT}\Big)\right. \\
& \left. -a\frac{P_{O_2}(K_{O_2}P_{O_2})^{-\frac{1}{2}}(K_{H_2}P_{H_2})^{\frac{1}{2}}}{2\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}\frac{1}{k_BTP_o}\exp\Big(-\frac{\Delta G_{O_2\to 2O^*}}{k_BT}\Big)\right] \\
= \;& \frac{1+(P_{O_2}K_{O_2})^{\frac{1}{2}}}{k_BT\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)} - \frac{a(P_{O_2}K_{O_2})^{\frac{1}{2}}}{k_BT\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)} \\
= \;& \frac{\frac{1-a}{k_BT}P_{O_2}^{\frac{1}{2}}exp\Big(\frac{a\Delta E_H+b-C_{O_2}}{k_BT}\Big)+\frac{P_o^{\frac{1}{2}}}{k_BT}}{\Big(P_o^{\frac{1}{2}}+P_{H_2}^{\frac{1}{2}}\exp\Big(\frac{\Delta E_H-C_{H_2}}{k_BT}\Big)+P_{O_2}^{\frac{1}{2}}\exp\Big(\frac{a\Delta E_H+b-C_{O_2}}{k_BT}\Big)\Big)}
\end{aligned}
\tag{A2-10}
$$

$$
\begin{aligned}
& S^{O}_{H,\,corr}(\Delta E_H) \\
= \;& \left[\frac{\partial(\ln\hat{\theta}_{O^*})}{\partial(\Delta E_H)}\right]_{corr} = \frac{1}{\hat{\theta}_{O^*}}\left[\frac{\partial\hat{\theta}_{O^*}}{\partial(\Delta E_H)} + \frac{\partial\hat{\theta}_{O^*}}{\partial(\Delta E_O)}\frac{\partial(\Delta E_O)}{\partial(\Delta E_H)}\right] \\
= \;& -\frac{2}{\hat{\theta}_{O^*}}\left[\frac{P_{H_2}(K_{H_2}P_{H_2})^{-\frac{1}{2}}(K_{O_2}P_{O_2})^{\frac{1}{2}}}{2\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}\frac{1}{k_BTP_o}\exp\Big(-\frac{\Delta G_{H_2\to 2H^*}}{K_BT}\Big)\right. \\
& \left. -a\frac{P_{O_2}(K_{O_2}P_{O_2})^{-\frac{1}{2}}(1+K_{H_2}P_{H_2})^{\frac{1}{2}}}{2\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)^2}\frac{1}{k_BTP_o}\exp\Big(-\frac{\Delta G_{O_2\to 2O^*}}{k_BT}\Big)\right] \\
= \;& -\frac{(P_{H_2}K_{H_2})^{\frac{1}{2}}}{k_BT\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)} + \frac{a\Big(1+(P_{H_2}K_{H_2})^{\frac{1}{2}}\Big)}{k_BT\Big(1+(K_{H_2}P_{H_2})^{\frac{1}{2}}+(K_{O_2}P_{O_2})^{\frac{1}{2}}\Big)} \\
= \;& \frac{\frac{a-1}{k_BT}P_{H_2}^{\frac{1}{2}}exp\Big(\frac{\Delta E_H-C_{H_2}}{k_BT}\Big)+\frac{aP_o^{\frac{1}{2}}}{K_BT}}{\Big(P_o^{\frac{1}{2}}+P_{H_2}^{\frac{1}{2}}\exp\Big(\frac{\Delta E_H-C_{H_2}}{k_BT}\Big)+P_{O_2}^{\frac{1}{2}}\exp\Big(\frac{a\Delta E_H+b-C_{O_2}}{k_BT}\Big)\Big)}
\end{aligned}
\tag{A2-11}
$$

From (A2-10) and (A2-11), we can conclude that when ΔE_H small, $S^{H}_{H,\,corr}$ goes to $1/(k_BT)$, $S^{O}_{H,\,corr}$ goes to $a/(k_BT)$, and when ΔE_H large, $S^{H}_{H,\,corr}$ goes to $(1-a)/(k_BT)$, $S^{O}_{H,\,corr}$ goes to 0. The plot of $S^{H}_{H,\,corr}$ and $S^{O}_{H,\,corr}$, shown in Figure 31, is consistent with this result.

The CLSIs with respect to ΔE_O can also be computed similarly using the following

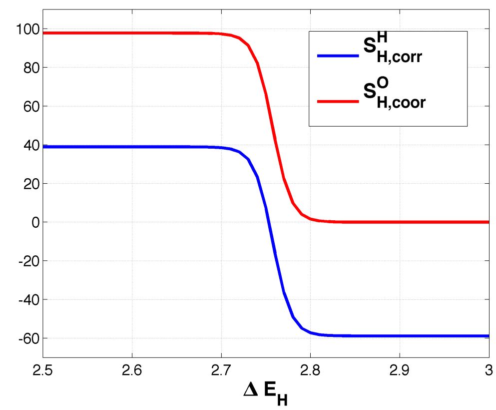

Figure 31. Correlated LSI of $S^H_{H,\,corr}$ and $S^O_{H,\,corr}$ respected to ΔE_H (eV) in the deterministic case, according to (A2-10) and (A2-11). When ΔE_H is small, less than -2.1, $S^H_{H,\,corr}$ goes to $1/(k_BT) = 38.9218$, $S^O_{H,\,corr}$ goes to $a/(k_BT) = 97.7639$, and when ΔE_H is large, greater than 2.8, $S^H_{H,\,corr}$ goes to $(1-a)/(k_BT) = -58.8421$, $S^O_{H,\,corr}$ goes to 0. And for Pt, $\Delta E_H = 2.6581(eV)$, $S^H_{H,\,corr} = 38.9021$ and $S^O_{H,\,corr} = 97.7442$.

equations,

$$\begin{aligned} S^H_{O,\,corr} &= \left[\frac{\partial(\ln\hat{\theta}_{H^*})}{\partial(\Delta E_O)}\right]_{corr} = \frac{1}{\hat{\theta}_{H^*}}\left[\frac{\partial\hat{\theta}_{H^*}}{\partial(\Delta E_O)} + \frac{\partial\hat{\theta}_{H^*}}{\partial(\Delta E_H)}\frac{\partial(\Delta E_H)}{\partial(\Delta E_O)}\right], \\ S^O_{O,\,corr} &= \left[\frac{\partial(\ln\hat{\theta}_{O^*})}{\partial(\Delta E_O)}\right]_{corr} = \frac{1}{\hat{\theta}_{O^*}}\left[\frac{\partial\hat{\theta}_{O^*}}{\partial(\Delta E_O)} + \frac{\partial\hat{\theta}_{O^*}}{\partial(\Delta E_H)}\frac{\partial(\Delta E_H)}{\partial(\Delta E_O)}\right]. \end{aligned} \tag{A2-12}$$

A.3 Correlated parametric models

As we said in Section VII B, besides normal distribution, we can fit the data or the adjusted data using some other parametric models to determine the distribution of ω. In Table 5, we consider the data $\omega + 1$ and give the fitting results of four different parametric distributions by MLE, we can see the Extreme Value distribution is the best fit model according to the goodness of fit Log-likelihood value, shown in Figure 32.

These results are given in MLE sense, we can also try Moment Matching Es-

Table 5. Fit results by different parametric models

Data set	Model of fitting *	Log-likelihood value
$\omega + 1$	Normal distribution	-3.32336
	Gamma distribution	-5.02882
	t Location-Scale distribution	-2.76145
	Extreme Value distribution	-1.60086

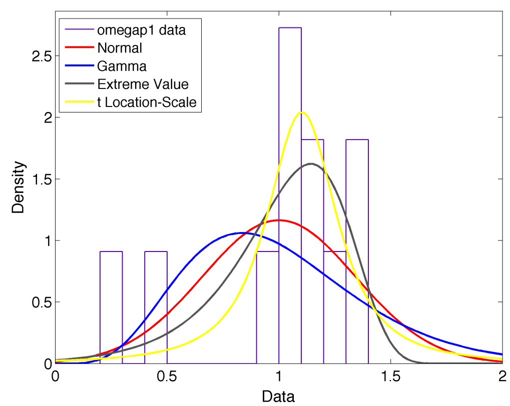

Figure 32. Fits of $\omega + 1$ using Normal, Gamma, t Location-Scale and Extreme Value distributions, where Extreme Value distribution is the best approximation of them using the maximum likelihood method.

timation method (MME) or other methods [15]. Moreover, following the steps in Section VII B, we can compute the CLSIs for the corresponding parametric models. The results are shown in Figure 33.

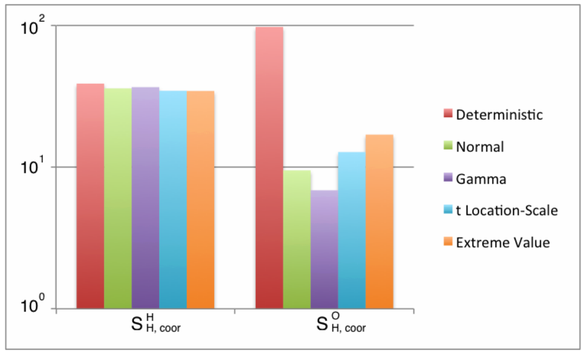

Figure 33. The correlated LSI results, $S^{H}_{H,\,corr}$ and $S^{O}_{H,\,corr}$, for different parametric models, computed by (37) and (38). Although the results of $S^{H}_{H,\,corr}$ are almost the same for different models, the results of $S^{O}_{H,\,corr}$ using uncertain models are much smaller than the deterministic model. Moreover, we can find the order of CLSI values is matched with the order of log-likelihood values for parametric models.

A.4 Correlated non-parametric models

In Figure 34, we show the histogram and some kernel density estimators for our data, ω, using different kernel or bandwidth. The Log-likelihood values of each model are presented in Table 6. Comparing values in Table 5, we find all the Log-likelihood values of non-parametric models to be much higher than the parametric models because they capture the second mode of the data on the left, between -1 and -0.5, while the parametric ones do not. The normal kernel density distribution with small bandwidth is the best fit of those three.

Table 6. Non-parametric models with different kernel or bandwidth

Kernel function	bandwidth	Log-likelihood value
uniform distribution	0.1	1.64177699
normal distribution	0.1	1.168341941
normal distribution	0.05	3.078962095

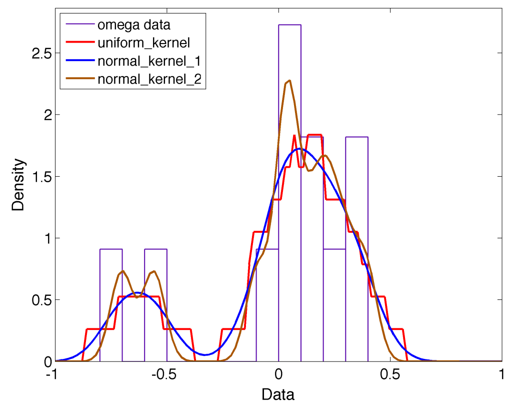

Figure 34. Fit of ω using non-parametric distributions with different kernels or bandwidth described in Table 6.

The results of $S^H_{H,corr}$ and $S^O_{H,corr}$ for Pt using histogram, uniform and normal kernel density function with different bandwidths are shown in Figure 35.

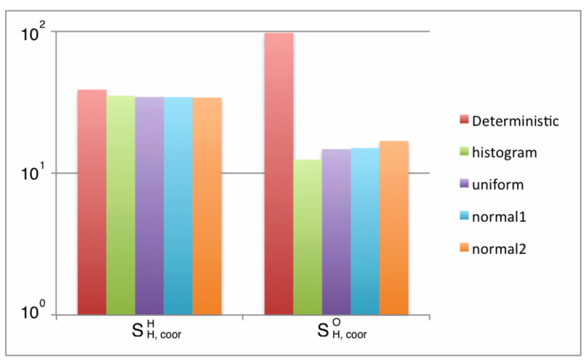

Figure 35. The correlated LSI results, $S^H_{H,corr}$ and $S^H_{O,corr}$, of Pt for different non-parametric models, computed by (43) and (38). The bandwidth of the histogram, uniform and normal1 is 0.1 and the bandwidth of the normal2 is 0.05. As with the uncertain parametric models, the results of $S^H_{H,corr}$ are almost the same for these different models, but the results of $S^O_{H,corr}$ using uncertain non-parametric models are much smaller than the deterministic model.

A.5 Computational implementation

To compute the proposed CGSI of (13) we employ a standard Monte Carlo sampling method. By applying this method we can bypass the direct integration of (13) by sampling independent and identically distributed random vectors of $\lambda_1^{(1)}, \lambda_1^{(2)}, \ldots, \lambda_1^{(n)}$ from the PDF of $p(\lambda_1)$ where n denotes the sufficiently large number of samples required for convergence.

Following such an approach requires two Monte Carlo sampling loops; (i) an internal loop to calculate the integration of (11) and (ii) an external loop to compute the CGSA index of (13). The algorithmic implementation of the proposed approach can be summarized as follows:

1. **Draw a sample for λ_1 from the marginal probability distribution of $p(\lambda_1)$ described in (10).**

   The required marginal distribution (uniform, normal, ...) is dictated by how the parameter varies over its entire range. For many physico-chemical systems a normal distribution may apply whose mean and standard deviation are computed.

2. **Draw many samples for λ_2 from the conditional probability distribution of $p(\lambda_2|\lambda_1)$ for each λ_1.**

   By sampling from such conditional probability distribution we account for parameter correlation in GSI calculation.

3. **Calculate $\ln F(\lambda_1)$ from (11), then estimate the gradient $\nabla_{\lambda_1} \ln F(\lambda_1)$ using centered finite difference approximation considering a sufficiently small perturbation ϵ [43],**

$$\nabla_{\lambda_1} \ln F(\lambda_1) \approx \frac{\ln F(\lambda_1 + \epsilon) - \ln F(\lambda_1 - \epsilon)}{2\epsilon} \qquad \text{(A5-1)}$$

all by Monte Carlo integration, and then compute $|\nabla_{\lambda_1} \ln F(\lambda_1)|^q$.

4. **Repeat the previous steps until convergence of the estimator (13).** This step computes the CGSI.

We give an elementary example below to pin down the notation.

Example: By sampling λ_1 from a normal marginal distribution with mean of μ_{λ_1} and standard deviation of σ_{λ_1},

$$\lambda_1 \in \Lambda_1, \qquad \Lambda_1 \sim \mathcal{N}(\mu_{\Lambda_1}, \sigma^2_{\Lambda_1}), \tag{A5-2}$$

with the PDF of

$$p(\lambda_1) = \frac{1}{\sigma_{\Lambda_1}\sqrt{2\pi}}\, \mathrm{e}^{-\dfrac{(\lambda_1 - \mu_{\Lambda_1})^2}{2\sigma^2_{\Lambda_1}}},$$

we can draw samples directly from conditional distribution of $p(\lambda_2|\lambda_1)$ by sampling λ_2 from a normal distribution

$$\lambda_2 \in \Lambda_2, \qquad \Lambda_2 \sim \mathcal{N}(\mu, \sigma^2), \tag{A5-3}$$

with the conditional PDF of $p(\lambda_2|\lambda_1)$ for the given λ_1

$$p(\lambda_2|\lambda_1) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\dfrac{(\lambda_2 - \mu)^2}{2\sigma^2}},$$

where

$$\mu = \mu_{\Lambda_2} + \frac{\sigma_{\Lambda_2}}{\sigma_{\Lambda_1}}\rho(\lambda_1 - \mu_{\Lambda_1}), \qquad \sigma^2 = (1 - \rho^2)\sigma^2_{\Lambda_2}, \tag{A5-4}$$

and ρ is the correlation coefficient, and μ_{Λ_2} and σ_{Λ_2} are the mean and standard deviation of the normal distribution of Λ_2, respectively [40, 130].

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 2 & 3

B.1 Properties of Gaussian Bayesian Networks

Notations:

- $X_{1:k}$: vector of random variables $XX_{1:k} = (X_1, \ldots, X_k)$, and
- $x_{1:k}$: vector of values of the corresponding random variables $X_{1:k}$.
- $P(\cdot)$, $Q(\cdot)$: probability measure for random variables X.
- $p(\cdot)$, $q(\cdot)$: denote the probability density function (PDF) corresponding to P, Q.
- $\mu_{1:k}$: mean vector of $X_{1:k}$ where we use the notations $\mu_{1:k} = (\mu_1, \ldots, \mu_k)$.
- $\mathcal{C}_{1:k}$, $\mathcal{C}$: $\mathcal{C}_{1:k}$ is the covariance matrix of $X = X_{1:k}$ for any $k \leq n$, and $\mathcal{C} = \mathcal{C}_{1:n}$. Furthermore, $\mathcal{C}_{1:k}$ is also the sub matrix that consists of first k rows and k columns of matrix $\mathcal{C}$

In order to simplify the proofs and notations, we assume that for any $l \in \pi_i$, we have that $l < i$, i.e.,

$$X_{\pi_i} \subset \{X_1, \ldots, X_{i-1}\}, \quad \text{for all } i \leq n\,.$$

Note that this is a general assumption which can be satisfied by reordering ($X_1, \ldots,$ X_n), see [66][Theorem 7.5]. The ORR PGM shown in Fig 16 automatically satisfies this assumption. Then we can rewrite $p(x_i|x_{\pi_i})$ as

$$p(x_i|x_{\pi_i}) = p(x_i|x_{1:i-1}) = \mathcal{N}(\beta_{i0} + \beta_i^T x_{1:i-1}, \sigma_i^2)\,, \quad \text{for } i \leq n\,, \tag{B1-1}$$

with $\beta_i = (\beta_{i1}, \ldots, \beta_{i,i-1})$ and where $\beta_{ij} = 0$ if $j \notin \pi_i$. Given the parameters in (4.4) for each X_i, we can compute the joint distribution for X iteratively by the following:

Lemma B.1 *For any X_i in the GBN* (B1-1)*, $X_{1:i} = (X_1, \ldots, X_i)$ are jointly Gaussian with distribution $\mathcal{N}(\mu_{1:i}, \mathcal{C}_{1:i})$ where $\mu_{1:i}$, $\mathcal{C}_{1:i}$ can be computed iteratively through $\mu_{1:i-1}$, $\mathcal{C}_{1:i-1}$ by*

$$\mu_{1:i} = (\mu_{1:i-1}, \mu_i)^T \tag{B1-2}$$

$$\mathcal{C}_{1:i} = \begin{bmatrix} & & & \mathcal{C}_{1i} \\ & \mathcal{C}_{1:i-1} & & \vdots \\ & & & \mathcal{C}_{i-1,i} \\ \mathcal{C}_{1i} & \ldots & \mathcal{C}_{i-1,i} & \mathcal{C}_{ii} \end{bmatrix} \tag{B1-3}$$

where μ_i, $\mathcal{C}_{ii}$ are the mean and variance of the marginal Gaussian distribution of X_i, denoted by P_i, given iteratively by

$$\mu_i = \beta_{i0} + \beta_i^T \mu_{1:i-1} \tag{B1-4}$$

$$\mathcal{C}_{ii} = \sigma_i^2 + \beta_i^T \mathcal{C}_{1:i-1} \beta_i \tag{B1-5}$$

and

$$\mathcal{C}_{ji} = \sum_{k=1}^{i-1} \beta_{ik} \mathcal{C}_{jk} \qquad \textit{for } j = 1, \ldots, i-1,$$

where $\mathcal{C}_{jk} = Cov(X_j, X_k) = \mathbb{E}\left[(X_j - \mu_j)(X_k - \mu_k)\right]$ are the elements in $\mathcal{C}_{1:i-1}$. Finally, $p(x_1)$ follows $\mathcal{N}(\mu_1, \mathcal{C}_{11})$ with $\mu_1 = \beta_{10}$, $\mathcal{C}_{11} = \sigma_1^2$.

Proof. This is a general result for multivariate Gaussian distribution, see [66] Theorem 7.3. ■

Conversely to Lemma 1, if we are given a joint distribution of a GBN, we can readily obtain the conditional distribution of X_i given any $X_{1:l}$ for any $l < i$ by the following:

Lemma B.2 *Consider the GBN* (B1-1) *with joint distribution* $p(x) = \mathcal{N}(\mu, \mathcal{C})$. *Then for any* X_i,

$$p(x_i|x_{1:l}) = \mathcal{N}(\tilde{\beta}_{i0} + \tilde{\beta}_i x_{1:l}, \tilde{\sigma}_i^2)\,, \quad \textit{for any } l < i\,,$$

where

$$\tilde{\beta}_{i0} = \mu_i - \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1}\mu_{1:l} \tag{B1-6}$$

$$\tilde{\beta}_i = \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1} \tag{B1-7}$$

$$\tilde{\sigma}_i^2 = \mathcal{C}_{ii} - \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1}\mathcal{C}_{1:l,i} \tag{B1-8}$$

and $\mathcal{C}_{i,1:l} = \mathcal{C}_{1:l,i}^T = (\mathcal{C}_{i1}, \ldots, \mathcal{C}_{il})$, $\mathcal{C}_{ii}$ *is the variance of* X_i, $\mathcal{C}_{1:l}$ *is the covariance matrix of* $X_{1:l}$. *All these variances and covariances are included as sub matrices in* $\mathcal{C}$. *Note that* $\tilde{\beta}_{ij} = \beta_{ij}$ *if* $j \in \pi_i$ *and* X_j *is not an ancestor of other variables in* X_{π_i}; $\tilde{\beta}_{ij} = 0$ *if* X_j *is not an ancestor of* X_i.

Proof. Given that $p(x) = \mathcal{N}(\mu, \mathcal{C})$, then by the properties of multivariate Gaussians, [130], we know the density of marginal distribution for $(X_1, \ldots, X_l, X_i)$, $l < i$, is

$$p(x_{1:l}, x_i) = \mathcal{N}\left(\begin{pmatrix} \mu_{1:l} \\ \mu_i \end{pmatrix}, \begin{pmatrix} \mathcal{C}_{1:l} & \mathcal{C}_{1:l,i} \\ \mathcal{C}_{i,1:l} & \mathcal{C}_{ii} \end{pmatrix}\right) \tag{B1-9}$$

where $\mu_{1:l} = (\mu_1, \ldots, \mu_l)^T$, $\mathcal{C}_{1:l}$ is the sub matrix of $\mathcal{C}$ consisting of the first l rows and columns; furthermore, $\mathcal{C}_{i,1:l} = \mathcal{C}_{1:l,i}^T = (\mathcal{C}_{i1}, \ldots, \mathcal{C}_{il})$. Therefore, by the Gaussian properties of conditional distribution, [130], we have

$$p(x_i|x_{1:l}) := p(x_i|x_1, \dots, x_l) = \mathcal{N}(\mu_{i|1:l}, \mathcal{C}_{i|1:l}) \tag{B1-10}$$

where

$$\mu_{i|1:l} = \mu_i + \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1}(x_{1:l} - \mu_{1:l}) = \mu_i - \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1}\mu_{1:l} + \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1}x_{1:l} \tag{B1-11}$$

and

$$\mathcal{C}_{i|1:l} = \mathcal{C}_{ii} - \mathcal{C}_{i,1:l}\mathcal{C}_{1:l}^{-1}\mathcal{C}_{1:l,i} \tag{B1-12}$$

Thus, we obtain $p(x_i|x_{1:l}) = \mathcal{N}(\tilde{\beta}_{i0} + \tilde{\beta}_i x_{1:l}, \tilde{\sigma}_i^2)$, where $\tilde{\beta}_{i0} + \tilde{\beta}_i x_{1:l} = \mu_{i|1:l}$ yields (B1-6), (B1-7), and $\tilde{\sigma}_i^2 = \mathcal{C}_{i|1:l}$ yields (B1-8), for all $i \leq n$. ■

Therefore, for the ORR PGM, applying Lemma B.2, we can compute $\tilde{\beta}_{y_i,\omega_j}$ in $p(y_i|\omega_j, x_0)$ for all j. Indeed, based on the model shown in Section 4.1.4 in Main Text, for $j = e1, d1, s1, c1$ we have the following CPDs that model different errors which affect y_1 (see Table 3 and Figure 16 in Main Text), and

$$\begin{aligned} & p(y_1|\omega_j, x_0) \\ = \; & \int p(y_1|x, \omega_{e1}, \omega_{d1}, \omega_{s1}, \omega_{c1}) p(x|\omega_{e0}, \omega_{d0}, \omega_{s0}, x_0) \prod_{all\ \{\omega_k\}\setminus\omega_j} p(\omega_k) d\omega_k dx \\ = \; & \mathcal{N}(\tilde{\beta}_{y_1,0} + \tilde{\beta}_{y_1,\omega_j}\omega_j, \tilde{\sigma}_{y_1}^2) = \mathcal{N}(\tilde{\beta}_{y_1,0} + \omega_j, \tilde{\sigma}_{y_1}^2)\,, \end{aligned} \tag{B1-13}$$

where

$$\tilde{\beta}_{y_1,0} = \beta_{y_1,0} + \beta_{y_1,x}(\beta_{e0,0} + \beta_{d0,0} + \beta_{s0,0}) + \sum_{\omega_k \in \{\omega_{e1},\omega_{d1},\omega_{s1},\omega_{c1}\}\setminus\omega_j} \beta_{k0} \tag{B1-14}$$

$$\tilde{\sigma}_{y_1}^2 = \beta_{y_1,x}^2(\sigma_{e0}^2 + \sigma_{d0}^2 + \sigma_{s0}^2) + \sum_{\omega_k \in \{\omega_{e1},\omega_{d1},\omega_{s1},\omega_{c1}\}\setminus\omega_j} \sigma_k^2 \tag{B1-15}$$

and $\tilde{\beta}_{y_1,\omega_j} = 1$. We recall that all β values are already calculated from MLE, see Table 2.

Similarly for $j = e2, d2, s2, c2$, which only affect y_2, see Table 3 in Main Text, we have

$$p(y_1|\omega_j, x_0) = \mathcal{N}(\tilde{\beta}_{y_1,0} + \tilde{\beta}_{y_1,\omega_j}\omega_j, \tilde{\sigma}^2_{y_1}) = \mathcal{N}(\tilde{\beta}_{y_1,0}, \tilde{\sigma}^2_{y_1}), \tag{B1-16}$$

where

$$\tilde{\beta}_{y_1,0} = \beta_{y_i,0} + \beta_{y_1,x}(\beta_{e0,0} + \beta_{d0,0} + \beta_{s0,0}) + \sum_{\omega_k \in \{\omega_{e1},\omega_{d1},\omega_{s1},\omega_{c1}\}} \beta_{k0} \tag{B1-17}$$

$$\tilde{\sigma}^2_{y_1} = \beta^2_{y_1,x}(\sigma^2_{e0} + \sigma^2_{d0} + \sigma^2_{s0}) + \sum_{\omega_k \in \{\omega_{e1},\omega_{d1},\omega_{s1},\omega_{c1}\}} \sigma^2_k \tag{B1-18}$$

and $\tilde{\beta}_{y_1,\omega_j} = 0$.

Finally, for $j = e0, d0, s0$ which affect x, see Table 3 in Main Text, we have

$$p(y_1|\omega_j, x_0) = \mathcal{N}(\tilde{\beta}_{y_1,0} + \tilde{\beta}_{y_1,\omega_j}\omega_j, \tilde{\sigma}^2_{y_1}) = \mathcal{N}(\tilde{\beta}_{y_1,0} + \beta_{y_1,\omega_j}\omega_j, \tilde{\sigma}^2_{y_1}) \tag{B1-19}$$

where where

$$\tilde{\beta}_{y_1,0} = \beta_{y_1,0} + \beta_{y_1,x}(\sum_{\omega_k \in \{\omega_{e0},\omega_{d0},\omega_{s0}\}\backslash\omega_j} \beta_{k0}) + \beta_{e1,0} + \beta_{d1,0} + \beta_{s1,0} + \beta_{c1,0} \tag{B1-20}$$

$$\tilde{\sigma}^2_{y_1} = \beta^2_{y_1,x}(\sum_{\omega_k \in \{\omega_{e0},\omega_{d0},\omega_{s0}\}\backslash\omega_j} \sigma^2_k) + \sigma^2_{e1} + \sigma^2_{d1} + \sigma^2_{s1} + \omega^2_{c1} \tag{B1-21}$$

and $\tilde{\beta}_{y_1,\omega_j} = \beta_{y_1,x}$. Similar constructions are carried out for the conditionals of y_2. We summarize all our results for $\tilde{\beta}_{y_i,\omega_j}$ in the following table:

Table 7. Different $\tilde{\beta}_{y_i,\omega_j}$ in $p(y_i|\omega_j, x_0) = \mathcal{N}(\tilde{\beta}_{y_i,0} + \tilde{\beta}_{y_i,\omega_j}\omega_j, \tilde{\sigma}^2_{y_i})$

	$\omega_j = \omega_{e0}, \omega_{d0}, \omega_{s0}$	$\omega_j = \omega_{e1}, \omega_{d1}, \omega_{s1}, \omega_{c1}$	$\omega_j = \omega_{e2}, \omega_{d2}, \omega_{s2}, \omega_{c2}$
$f = y_1$	$\tilde{\beta}_{y_1,\omega_j} = \beta_{y_1,x}$	$\tilde{\beta}_{y_1,\omega_j} = 1$	$\tilde{\beta}_{y_1,\omega_j} = 0$
$f = y_2$	$\tilde{\beta}_{y_2,\omega_j} = \beta_{y_2,x}$	$\tilde{\beta}_{y_2,\omega_j} = 0$	$\tilde{\beta}_{y_2,\omega_j} = 1$

Remark: We recall that β's were calculated in Table 2. The values of $\tilde{\beta}_{y_i,\omega_j}$'s have a physical meaning for the ORR PGM since they capture dependence via the DAG structure: $\tilde{\beta}_{y_i,\omega_j} = 0$ implies ω_j does not affect the prediction of y_i, and $\tilde{\beta}_{y_i,\omega_j} = \beta_{y_1,x}$ shows how much the uncertainty of ω_j propagates to y_i through the linear regression in Figure 14.

B.2 Predictive Uncertainty Indices

B.2.1 Proof of Theorem 3.1

To prove the theorem, we first show two lemmas which are presented in [29, 50], and we present the proof here for completeness.

Lemma B.3 *Let P be a probability measure and let $f(X)$ be such that its MGF is finite in a neighborhood of the origin. Then for any Q with $R(Q||P) < \infty$, we have*

$$-\inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_P\left[e^{-c\bar{f}(X)}\right]+\frac{\eta}{c}\right] \leq \mathbb{E}_Q\left[f(X)\right]-\mathbb{E}_P\left[f(X)\right] \leq \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_P\left[e^{c\bar{f}(X)}\right]+\frac{\eta}{c}\right] \tag{B2-1}$$

Proof of Lemma B.3: For any general QoI $f(X)$ which has finite moment generating function (MGF), $\mathbb{E}_P\left[e^{\pm c\bar{f}(X)}\right] := \mathbb{E}_P\left[e^{c(f(X)-\mathbb{E}_P[f(X)])}\right]$, in a neighborhood of the origin, there is a known fact in statistics and large deviation theory [27, 29] that

$$\log \mathbb{E}_P\left[e^{f(X)}\right] = \sup_{Q\ll P}\{\mathbb{E}_Q\left[f(X)\right] - R(Q||P)\} \,. \tag{B2-2}$$

Changing $f(X)$ to $c\bar{f}(X) = c(f(X) - \mathbb{E}_P\left[f(X)\right])$, we get

$$\mathbb{E}_P\left[e^{\pm c\bar{f}(X)}\right] = \sup_{Q\ll P}\{c(\mathbb{E}_Q\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right]) - R(Q||P)\} \tag{B2-3}$$

which gives us the following upper and lower bounds with $c > 0$,

$$-\inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_P\left[e^{-c\bar{f}(X)}\right]+\frac{\eta}{c}\right] \leq \mathbb{E}_Q\left[f(X)\right]-\mathbb{E}_P\left[f(X)\right] \leq \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_P\left[e^{c\bar{f}(X)}\right]+\frac{\eta}{c}\right] \tag{B2-4}$$

where $\eta = R(Q||P)$.

Lemma B.4 *Let P be a probability measure and $f(X)$ to be a non-constant function such that its moment generating function $\mathbb{E}_P\left[e^{\pm c\bar{f}(X)}\right]$ is finite in a neighborhood of 0. Let Q be such that $R(Q||P) = \eta$.*

(a) For any $\eta \geq 0$ the optimization problems

$$\pm \inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_P \left[e^{\pm c \bar{f}(X)} \right] + \frac{\eta}{c} \right] \tag{B2-5}$$

have unique minimizers $c_\pm \in [0, +\infty]$. Moreover there exists $0 < \eta^\pm \leq \infty$ such that the minimizers $c_\pm = c_\pm(\eta)$ are finite for $\eta \leq \eta^\pm$ and $c_\pm(\eta) = +\infty$ if $\eta > \eta^\pm$.

(b) If $c^\pm = c_\pm(\eta)$ is finite

$$\pm \inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_P \left[e^{\pm c \bar{f}(X)} \right] + \frac{\eta}{c} \right] = \pm \left(\mathbb{E}_{P^{\pm c_\pm}} [f(X)] - \mathbb{E}_P [f(X)] \right), \tag{B2-6}$$

where $P^{\pm c_\pm}$ is defined by

$$\frac{dP^{\pm c_\pm}}{dP} = \frac{e^{\pm c_\pm f(x)}}{\mathbb{E}_P \left[e^{\pm c_\pm f(X)} \right]} \tag{B2-7}$$

and $c_\pm(\eta)$ is strictly increasing in η and is determined by the equation

$$R(P^{\pm c_\pm} || P) = \eta \,. \tag{B2-8}$$

(c) If $\eta^\pm < \infty$ then $f(X)$ is necessarily P almost surely bounded above/bounded below respectively with upper/lower bound $f_\pm$. For $\eta > \eta^\pm$ we have that $c_\pm(\eta) = +\infty$ and

$$\pm \inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_P \left[e^{\pm c \bar{f}(X)} \right] + \frac{\eta}{c} \right] = \pm \left(f_\pm(X) - \mathbb{E}_P [f(X)] \right). \tag{B2-9}$$

Proof of the Lemma B.4: For notational ease, in the proof, let us set $H(c) = \log \mathbb{E}_P \left[e^{\pm c \bar{f}(X)} \right]$ and note that since $\bar{f}$ is centered we have $H(0) = H'(0) = 0$. We have $H'(c) = \mathbb{E}_{P^c}[f] - \mathbb{E}_P[f(X)]$ and $H''(c) = \mathrm{Var}_{P^c}(f) > 0$ since $f(X)$ is not constant P almost surely.

If $d_+ < \infty$ then we have $\lim_{c \to d_+} H(c) = \infty$ and $\lim_{c \to d_+} H'(c) = \infty$. If $d_+ = \infty$ then

$$\lim_{c \to \infty} H'(c) = \begin{cases} f_+ - \mathbb{E}_P[f(X)] & \text{if } f \text{ is bounded} \\ +\infty & \text{otherwise} \end{cases} \,. \tag{B2-10}$$

Since $c^{-1}H(c) = c^{-1}\int_0^c H'(t)dt$ and $H'(c)$ is strictly increasing $c^{-1}H(c)$ is a strictly increasing function and we have $\lim_{c\to\infty} c^{-1}H(c) = \lim_{c\to\infty} H'(c)$ which is finite if only if $f(X)$ is bounded. Let us set

$$B(c;\eta) = \frac{\mathbb{E}_P\left[e^{\pm c\bar{f}(X)}\right] + \eta}{c} = \frac{H(c) + \eta}{c}$$

and then distinguish two cases:

(a) If $d_+ \leq \infty$ or if $d_+ = \infty$ and $f(X)$ is unbounded then we have $\lim_{c\to 0} B(c;\eta) = \lim_{c\to d_+} B(c;\eta) = +\infty$ and thus $B(c;\eta)$ has at least one minimum for some $0 < c < d_+$. By calculus the minimum must be a solution of

$$0 = \frac{\partial}{\partial c}B(c;\eta) = \frac{cH'(c) - H(c) - \eta}{c^2}$$

that is me must have $cH'(c) - H(c) = \eta$. Since $\frac{\partial}{\partial c}(cH'(c) - H(c)) = cH''(c) > 0$ the function $cH'(c) - H(c)$ is strictly increasing and thus there is a unique minimizer c_+ for $B(c;\eta)$.

(b) If $d_+ = \infty$ but $f(X)$ is bounded, since $cH'(c) - H(c)$ is strictly increasing we have $\lim_{c\to\infty} cH'(c) - H(c) = M_+$ which may or may not be finite depending on P. If $\eta \leq \eta_+$ we can proceed as in (a) to find a unique minimizer for a finite c_+, while if $\eta > M\eta_+$, $B(c;\eta)$ is strictly decreasing and thus the minimizer is attained at $c_+ = \infty$.

To conclude the proof we note that if $c_+ < \infty$ then $c_+H'(c_+) - H(c_+) = \eta$ and thus

$$B(c_+, \eta) = H'(c_+)$$

which proves (B2-6). On the other hand a simple computation shows that for any c

$$R(P^c||P) = cH'(c) - H(c)$$

and this establishes (B2-8). Finally if $c_+ = \infty$ the infimum is equal to $\lim_{c\to\infty} \frac{H(c)}{c}$ and this establishes (B2-9).

The proof of Theorem 3.1 follows immediately from the two lemmas, since by Lemma B.3,

$$\sup_{Q\in\mathcal{D}^\eta}/\inf \mathbb{E}_Q\left[f(X)\right] - \mathbb{E}_P\left[f(X)\right]$$

is bounded by

$$\pm \inf_{c>0}\left[\frac{1}{c}\log \mathbb{E}_P\left[e^{\pm c\bar{f}(X)}\right] + \frac{\eta}{c}\right]$$

then by Lemma B.4, we can find $Q^{\pm} \in \mathcal{D}^\eta$ which achieves the equality of the bounds by setting $Q^{\pm} := P^{\pm c_{\pm}}$ defined by

$$\frac{dP^{\pm c_{\pm}}}{dP} = \frac{e^{\pm c_{\pm} f(x)}}{\mathbb{E}_P\left[e^{\pm c_{\pm} f(X)}\right]} \tag{B2-11}$$

where $c_{\pm}$ is determined by the equation

$$R(P^{\pm c_{\pm}}||P) = \eta \tag{B2-12}$$

B.3 Model-Form Sensitivity Indices for PGMs

B.3.1 Proof of Theorem 3.4 & 3.5

Proof of Theorem 3.4: **Step 1: Bounds for the predictive uncertainty:** Since for any $Q \in \mathcal{D}_l^{\eta_l}$, we have $\pi_j^Q \equiv \pi_j^P = \pi_j$ and $Q_{j|\pi_j} \equiv P_{j|\pi_j}$ for all $j \neq l$, therefore, we can rewrite the bias as

$$\begin{aligned}
& \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\
= & \int_X f(x_k) \prod_{i=1}^{n} Q(dx_i|x_{\pi_i^Q}) - \int_X f(x_k) \prod_{i=1}^{n} P(dx_i|x_{\pi_i^P}) \\
= & \int_X f(x_k) \prod_{X_i \in \{X_k \cup \rho_k^Q)\}} Q(dx_i|x_{\pi_i^Q}) - \int_X f(x_k) \prod_{X_i \in \{\rho_k^P \cup \{k\})\}} P(dx_i|x_{\pi_i^P}) \\
= & \mathbb{E}_{Q_{\{k\}}}\left[f(X_k)\right] - \mathbb{E}_{P_{\{k\}}}\left[f(X_k)\right] \qquad \text{(B3-1)}
\end{aligned}$$

If $l \notin \rho_k^P \cup \{k\}$, we have $\pi_i^Q \equiv \pi_i^P =: \pi_i$ and $Q(dx_i|x_{\pi_i}) \equiv P(dx_i|x_{\pi_i})$ for all $i \in \rho_k \cup \{k\}$, therefore $Q_{\{k\}} \equiv P_{\{k\}}$, and thus $\mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] = 0$. Based on this calculation for $Q \in \mathcal{D}_l^{\eta_l}$, notice that our indices capture the graph structure correctly, e.g. perturbations on disconnected nodes do not affect the QoI $f = f(X_k)$.

On the other hand, for $l \in \rho_k^P \cup \{k\}$, consider $\pi_l := \pi_l^Q \cup \pi_l^P$, and $\rho_i := \rho_i^Q \cup \rho_i^P$ for all i, and define

$$Q(dx_l|x_{\pi_l}) := Q(dx_l|x_{\pi_l^Q}) \text{ for all } x_{\pi_l} \qquad \text{(B3-2)}$$

$$P(dx_l|x_{\pi_l}) := P(dx_l|x_{\pi_l^P}) \text{ for all } x_{\pi_l} \qquad \text{(B3-3)}$$

Since $Q(dx_j|x_{\pi_j}) \equiv P(dx_j|x_{\pi_j})$ for all $j \neq l$, we have

$$\begin{aligned}
& \mathbb{E}_Q\left[f(X_k)\right] - \mathbb{E}_P\left[f(X_k)\right] \\
= & \int_X f(x_k) \prod_{i\in\rho_k\cup\{k\}\setminus\rho_l\cup\{l\}} Q(dx_i|x^Q_{\pi_i}) \cdot Q(dx_l|x^Q_{\pi_l}) \cdot \prod_{i\in\rho_l} Q(dx_i|x^Q_{\pi_i}) \\
& - \int_X f(x_k) \prod_{i\in\rho_k\cup\{k\}\setminus\rho_l\cup\{l\}} P(dx_i|x^P_{\pi_i}) \cdot P(dx_l|x^P_{\pi_l}) \cdot \prod_{i\in\rho_l} P(dx_i|x^P_{\pi_i}) \\
= & \int_X f(x_k) \prod_{i\in\rho_k\cup\{k\}\setminus\rho_l\cup\{l\}} P(dx_i|x_{\pi_i}) \cdot Q(dx_l|x_{\pi_l}) \cdot \prod_{i\in\rho_l} P(dx_i|x_{\pi_i}) \\
& - \int_X f(x_k) \prod_{i\in\rho_k\cup\{k\}\setminus\rho_l\cup\{l\}} P(dx_i|x_{\pi_i}) \cdot P(dx_l|x_{\pi_l}) \cdot \prod_{i\in\rho_l} P(dx_i|x_{\pi_i}) \\
= & \int \left[\int F(x_l, x_{\rho_l}) Q(dx_l|x_{\pi_l}) - \int F(x_l, x_{\rho_l}) P(dx_l|x_{\pi_l})\right] \prod_{i\in\rho_l} P(dx_i|x_{\pi_i}) \\
= & \mathbb{E}_{P_{\rho_l}}\left[\mathbb{E}_{Q_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right] - \mathbb{E}_{P_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right]\right]
\end{aligned} \tag{B3-4}$$

where

$$F(x_l, x_{\rho_l}) = \int f(x_k) \prod_{i\in\rho_k\cup\{k\}\setminus\rho_l\cup\{l\}} P(dx_i|x_{\pi_i}) = \mathbb{E}_{P_{\{k\}|\rho_l^P\cup\{l\}}}\left[f(X_k)\right] \tag{B3-5}$$

therefore

$$\begin{aligned}
& \sup_{Q\in\mathcal{D}_l^{\eta_l}} \mathbb{E}_{P_{\rho_l}}\left[\mathbb{E}_{Q_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right] - \mathbb{E}_{P_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right]\right] \\
\leq & \mathbb{E}_{P_{\rho_l}}\left[\sup_{Q\in\mathcal{D}_l^{\eta_l}} \mathbb{E}_{Q_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right] - \mathbb{E}_{P_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right]\right] \\
= & \mathbb{E}_{P_{\rho_l}}\left[\sup_{Q_l\in\mathcal{E}_l^{\eta_l}} \mathbb{E}_{Q_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right] - \mathbb{E}_{P_{l|\pi_l}}\left[F(X_l, X_{\rho_l})\right]\right]
\end{aligned} \tag{B3-6}$$

where we define the ambiguity set for CPDs at l, namely

$$\mathcal{E}_l^{\eta_l} := \{\text{all CPD } Q_{l|\pi_l} = Q_l(\cdot|x_{\pi_l}) : R(Q_{l|\pi_l}||P_{l|\pi_l}) \leq \eta_l \text{ for all } x_{\pi_l} = x^P_{\pi_l} \cup x^Q_{\pi_l}\} \tag{B3-7}$$

Using Lemma B.3, for any given $X_{\rho_l} = x_{\rho_l}$, we have

$$\sup_{Q_l \in \mathcal{E}_l^{\eta_l}} \mathbb{E}_{Q_{l|\pi_l}} \left[F(X_l, X_{\rho_l}) \right] - \mathbb{E}_{P_{l|\pi_l}} \left[F(X_l, X_{\rho_l}) \right] \leq \inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{l|\pi_l}} \left[e^{c\bar{F}(X_l, X_{\rho_l})} \right] + \frac{\eta_l}{c} \right] \tag{B3-8}$$

thus (B3-6) implies

$$\sup_{Q \in \mathcal{D}_l^{\eta_l}} \mathbb{E}_Q \left[f(X_k) \right] - \mathbb{E}_P \left[f(X_k) \right] \leq \mathbb{E}_{P_{\rho_l}} \left[\inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{l|\pi_l}} \left[e^{c\bar{F}(X_l, X_{\rho_l})} \right] + \frac{\eta_l}{c} \right] \right] \tag{B3-9}$$

Step 2: Tightness of the bounds: As in Theorem 3.2, for any given $x_{\rho_l^P}$, we can consider the conditional measure $P^{c+}_{l|\rho_l^P}$ defined by

$$\frac{dP^{c+}_{l|\rho_l^P}}{dP_{l|\pi_l^P}} = \frac{e^{c_+(x_{\rho_l^P}) F(x_l, x_{\rho_l^P})}}{\mathbb{E}_{P_{l|\pi_l^P}} \left[e^{c_+(x_{\rho_l^P}) F(X_l, x_{\rho_l^P})} \right]} \tag{B3-10}$$

where $c_+(x_{\rho_l^P})$ is a function of $x_{\rho_l^P}$ which is determined by $R(P^{c+}_{l|\pi_l^P} || P_{l|\pi_l^P}) = \eta_l$, i.e.,

$$\int c_+(x_{\rho_l^P}) F(x_l, x_{\rho_l^P}) \frac{e^{c_+(x_{\rho_l^P}) F(x_l, x_{\rho_l^P})}}{\mathbb{E}_{P_{l|\pi_l^P}} \left[e^{c_+(x_{\rho_l^P}) F(X_l, x_{\rho_l^P})} \right]} P(dx_l | x_{\pi_l^P}) - \log \mathbb{E}_{P_{l|\pi_l^P}} \left[e^{c_+(x_{\rho_l^P}) F(X_l, x_{\rho_l^P})} \right] = \eta_l \tag{B3-11}$$

for any $x_{\rho_l^P}$. Using Lemma B.4, define

$$q_l^+(x_l | x_{\pi_l^{Q+}}) := P^{c+}_{l|\rho_l^P} \propto e^{c_+(x_{\rho_l^P}) F(x_l, x_{\rho_l^P})} p(x_l | x_{\pi_l^P}) \quad \text{for all } x_{\pi_l^{Q+}} . \tag{B3-12}$$

Note that $\pi_l^{Q^+}$ depends on π_l^P and $F(x_l, x_{\rho_l^P})$, hence $\pi_l^P \subset \pi_l^{Q^+} \subset \rho_l^P$, and $\rho_l^{Q^+} = \rho_l^P$. Therefore, using the same notation as in Step 1, for $\pi_l = \pi_l^{Q^+}$, $\rho_l = \rho_l^{Q^+}$, we have

$$\mathbb{E}_{Q^+_{l|\pi_l}} \left[F(X_l, X_{\rho_l}) \right] - \mathbb{E}_{P_{l|\pi_l}} \left[F(X_l, X_{\rho_l}) \right] = \inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{l|\pi_l}} \left[\bar{F}(X_l, X_{\rho_l}) \right] + \frac{\eta_l}{c} \right] . \tag{B3-13}$$

Furthermore,

$$R(Q^+_{l|\pi_l} || P_{l|\pi_l}) \leq \eta_l \text{ for all } x_{\pi_l} \quad \text{implies that} \quad Q_l^+ \in \mathcal{E}_l^{\eta_l} . \tag{B3-14}$$

Thus, let $q^+(x) = q_l^+(x_l | x_{\pi_l}) \prod_{i \neq l} p(x_i | x_{\pi_i})$, we have $Q^+ \in \mathcal{D}_l^{\eta_l}$, and

$$\mathbb{E}_{Q^+} \left[f(X_k) \right] - \mathbb{E}_P \left[f(X_k) \right] = \mathbb{E}_{P_{\rho_l}} \left[\inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{l|\pi_l}} \left[e^{c\bar{F}(X_l, X_{\rho_l})} \right] + \frac{\eta_l}{c} \right] \right] \tag{B3-15}$$

so we can conclude that

$$\sup_{Q\in\mathcal{D}_l^{\eta_l}} \mathbb{E}_Q\left[f(X_k)\right]-\mathbb{E}_P\left[f(X_k)\right] = \mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log\mathbb{E}_{P_{l|\pi_l}}\left[e^{c\bar{F}(X_l,X_{\rho_l})}\right]+\frac{\eta_l}{c}\right]\right] \quad \text{(B3-16)}$$

The calculations are similar for $\inf_{Q\in\mathcal{D}_l^{\eta_l}} \mathbb{E}_Q\left[f(X_k)\right]-\mathbb{E}_P\left[f(X_k)\right]$. ■

Proof of Theorem 3.5: **Step 1: Bounds for the predictive uncertainty:** The proof is the same as the proof in Theorem 3.4, noting that $\mathcal{D}_{l,P}^{\eta_l}\subset\mathcal{D}_l^{\eta_l}$. Therefore, we have

$$\sup_{Q\in\mathcal{D}_{l,P}^{\eta_l}} \mathbb{E}_Q\left[f(X_k)\right]-\mathbb{E}_P\left[f(X_k)\right] \le \mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log\mathbb{E}_{P_{l|\pi_l}}\left[e^{c\bar{F}(X_l,X_{\rho_l})}\right]+\frac{\eta_l}{c}\right]\right] \quad \text{(B3-17)}$$

Step 2: Tightness of the bounds: If $F(x_l,x_{\rho_l})=F(x_l,x_{\pi_l})$, it is the same as the proof in Theorem 3.4 (b). Indeed, let

$$q_l^+(x_l|x_{\pi_l^Q}) := P_{l|\rho_l}^{c+} \propto e^{c_+(x_{\pi_l})F(x_l,x_{\pi_l})}p(x_l|x_{\pi_l}) \quad \text{for all } x_{\pi_l^Q}\,. \quad \text{(B3-18)}$$

where c_+ only depends on x_{π_l} since F only depends on x_l and x_{π_l}, and we have $\pi_l^Q=\pi_l$, therefore, $Q_l^+\in\mathcal{Q}_{l,P}^{\eta_l}$, and let $q^+(x)=q_l^+(x_l|x_{\pi_l})\prod_{i\neq l}p(x_i|x_{\pi_i})$, we have $Q^+\in\mathcal{D}_{l,P}^{\eta_l}$, and

$$\mathbb{E}_{Q^+}\left[f(X_k)\right]-\mathbb{E}_P\left[f(X_k)\right] = \mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log\mathbb{E}_{P_{l|\pi_l}}\left[e^{c\bar{F}(X_l,X_{\rho_l})}\right]+\frac{\eta_l}{c}\right]\right]\,. \quad \text{(B3-19)}$$

Therefore we can conclude that

$$\sup_{Q\in\mathcal{D}_{l,P}^{\eta_l}} \mathbb{E}_Q\left[f(X_k)\right]-\mathbb{E}_P\left[f(X_k)\right] = \mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log\mathbb{E}_{P_{l|\pi_l}}\left[e^{c\bar{F}(X_l,X_{\rho_l})}\right]+\frac{\eta_l}{c}\right]\right] \quad \text{(B3-20)}$$

same for $\inf_{Q\in\mathcal{D}_{l,P}^{\eta_l}} \mathbb{E}_Q\left[f(X_k)\right]-\mathbb{E}_P\left[f(X_k)\right]$. ■

B.3.2 Proof of Theorem 4.1

Note that this theorem can be directly derive from Theorem 3.5 with the computation (B3-36), but here we still give a complete proof.

Proof: First, we proved parts (a) & (c) of the Theorem:

Step 1: Bounds for $I^{\pm}$: We first consider (4.18) for any general QoI $f = f(X)$ which has finite moment generating function (MGF), $\mathbb{E}_P\left[e^{c\bar{f}}\right]$ in a neighborhood of the origin. Then, for $Q \in \mathcal{D}^{\eta_l}_{l,P}$, and since by definition $Q_j \equiv P_j$ for all $j \neq l$ for $Q \in \mathcal{D}^{\eta_l}_{l,P}$, we have:

$$
\begin{aligned}
& I^{\pm}(f(X), P; \mathcal{D}^{\eta_l}_{l,P}) = \sup/\inf_{Q\in\mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_Q\left[f\right] - \mathbb{E}_P\left[f\right] \\
= \;& \sup/\inf_{Q\in\mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_{Q_{\rho_l}}\left[\mathbb{E}_{Q_l}\left[\mathbb{E}_{Q_{\{l\cup\rho_l\}^c}}\left[f\right]\right]\right] - \mathbb{E}_{P_{\rho_l}}\left[\mathbb{E}_{P_l}\left[\mathbb{E}_{P_{\{l\cup\rho_l\}^c}}\left[f\right]\right]\right] \\
= \;& \sup/\inf_{Q\in\mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_{P_{\rho_l}}\left[\mathbb{E}_{Q_{l|\pi_l}}\left[F_l\right]\right] - \mathbb{E}_{P_{\rho_l}}\left[\mathbb{E}_{P_{l|\pi_l}}\left[F_l\right]\right] \qquad \text{(B3-21)}
\end{aligned}
$$

where

$$
F_l(x) = F_l(x_l, x_{\rho_l}) = \mathbb{E}_{P_{\{l\cup\rho_l\}^c}}\left[f\right] = \mathbb{E}_P\left[f|X_l = x_l, X_{\rho_l} = x_{\rho_l}\right]. \qquad \text{(B3-22)}
$$

We use the notation ρ_l to denote the set of indices of ancestors for X_l, x_{ρ_l} are the corresponded values for these random variables X_{ρ_l}, and P_{ρ_l} is the marginal distribution of X_{ρ_l} with respect to P; similarly we define Q_{ρ_l}. Finally, we use the notation $\{\cdot\}^c$ to denote all the indices of random variables on the PGM except the ones inside the curly bracket $\{\cdot\}$. Thus,

$$
I^{\pm}(f(X), P; \mathcal{D}^{\eta_l}_{l,P}) \;=\; \sup/\inf_{Q\in\mathcal{D}^{\eta_l}_{l,P}} \mathbb{E}_{P_{\rho_l}}\left[\mathbb{E}_{Q_{l|\pi_l}}\left[F_l\right] - \mathbb{E}_{P_{l|\pi_l}}\left[F_l\right]\right]. \qquad \text{(B3-23)}
$$

Now let $\mathcal{E}^{\eta_l}_l$ be defined as

$$
\mathcal{E}^{\eta_l}_l := \{Q_{l|\pi_l} = Q_l(\cdot|x_{\pi_l}) : R(Q_{l|\pi_l}||P_{l|\pi_l}) \leq \eta_l \text{ for all } x_{\pi_l}\} \qquad \text{(B3-24)}
$$

which contains all alternative models Q_l with density $q_l(\cdot|x_{\pi_l})$ for the l-th component $P_{l|\pi_l}$ of the PGM (1.1) within KL tolerance η_l and with same structure, i.e. the same parents π_l. Considering the maximization of the first term in (B3-23), since the second term is independent of Q_l), we have

$$\sup_{Q_l \in \mathcal{E}_l^{\eta_l}} \mathbb{E}_{P_{\rho_l}} \left[\mathbb{E}_{Q_{l|\pi_l}} [F_l] \right] \leq \mathbb{E}_{P_{\rho_l}} \left[\sup_{Q_l \in \mathcal{E}_l^{\eta_l}} \mathbb{E}_{Q_{l|\pi_l}} [F_l] \right] . \tag{B3-25}$$

Therefore,

$$I^+(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) \leq \mathbb{E}_P \left[\sup/\inf_{Q_l \in \mathcal{E}_l^{\eta_l}} \mathbb{E}_{Q_{l|\pi_l}} [F_l] - \mathbb{E}_{P_{l|\pi_l}} [F_l] \right] . \tag{B3-26}$$

By applying Theoremn 3.1 to the right hand side of (B3-26), we have for any $Q_l \in \mathcal{E}_l^{\eta_l}$,

$$\mathbb{E}_{Q_{l|\pi_l}} [F_l] - \mathbb{E}_{P_{l|\pi_l}} [F_l] \leq \inf_{c>0} \left[\frac{1}{c} \log \int e^{c\bar{F}_l(x)} P(dx_l|x_{\pi_l}) + \frac{\eta_l}{c} \right] , \tag{B3-27}$$

hence,

$$I^+(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) \leq \mathbb{E}_P \left[\inf_{c>0} \left[\frac{1}{c} \log \int e^{c\bar{F}_l(x)} P(dx_l|x_{\pi_l}) + \frac{\eta_l}{c} \right] \right] , \tag{B3-28}$$

where $\bar{F}_l(X) = F_l(X) - \mathbb{E}_P [F_l(X)] = F_l(X) - \mathbb{E}_P [f(X)]$.

We note that for our ORR PGM, and due to the results shown in Table 7 and since $\rho_{\omega_l} = \emptyset$ (see Figure 15), we have for $f(X) = y_i$,

$$F_l(x) = \int y_i \prod_{X_i \in \{\omega_l\}^c} P(dx_i|x_{\pi_i}) = \tilde{\beta}_{y_i,0} + \tilde{\beta}_{y_i,\omega_l} \omega_l \tag{B3-29}$$

and

$$I^+(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) = \max_{Q_l \in \mathcal{E}_l^{\eta_l}} \mathbb{E}_{Q_l} [F_l] - \mathbb{E}_{P_l} [F_l] \tag{B3-30}$$

where P_l is the distribution of ω_l in (4.7), and

$$\mathcal{E}_l^{\eta_l} := \{Q_l = Q_l(\cdot) : R(Q_l(\cdot)||P_l(\cdot)) \leq \eta_l\} . \tag{B3-31}$$

Step 2: Tightness of the bounds: Consider the probability measure P_l^{c+} defined as the tilted measure with respect to P_l:

$$\frac{dP_l^{c+}}{dP_l} = \frac{e^{c_+ F_l}}{\mathbb{E}_{P_l} [e^{c_+ F_l}]} , \tag{B3-32}$$

where c_+ is selected as the solution of $R(P_l^{c+}||P_l) = \eta_l$ and σ_l is given in (4.7). Then, we have

$$\mathbb{E}_{P_l^{c+}}[F_l] - \mathbb{E}_{P_l}[F_l] = \frac{1}{c_+}\log\int e^{c_+\bar{F}_l} P_l(dx) + \frac{R(P_l^{c+}||P_l)}{c_+} \tag{B3-33}$$

where

$$\bar{F}_l(x) = F_l(x) - \mathbb{E}_P[F_l(x)] = \tilde{\beta}_{y_i,\omega_l}(x_l - \beta_{l0}) \quad \text{for} \quad \omega_l = x_l\,, \tag{B3-34}$$

and β_{l0} is the mean of ω_l, given in (4.7). Thus, letting $Q_+ = P_l^{c+}$ and since F_l only depends on x_l we obtain that Q_+ has *the same parents* as P_l. Therefore, $Q_+ \in \mathcal{E}_l^{\eta_l}$, and allows us to reach equality in (B3-26). We can now conclude that

$$I^+(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) = \inf_{c>0}\left[\frac{1}{c}\log\int e^{c\bar{F}_l} P_l(dx_l) + \frac{\eta_l}{c}\right] \tag{B3-35}$$

We carry out the same proof for $I^-(f(X), P; \mathcal{D}_{l,P}^{\eta_l})$.

Step 3: Computation of the bounds: For $P = \mathcal{N}(\mu, \mathcal{C})$, we have

$$\begin{aligned}
I^\pm(f(X), P; \mathcal{D}_{l,P}^{\eta_l}) &= \pm\inf_{c>0}\left[\frac{1}{c}\log\int e^{\pm c\bar{F}_l(x_l)} P_l(dx_l) + \frac{\eta_l}{c}\right] \\
&= \pm\inf_{c>0}\left[\frac{1}{c}\log\int e^{\pm c(\tilde{\beta}_{y_i,\omega_l}(x_l-\beta_{l0}))} P_l(dx_l) + \frac{\eta_l}{c}\right] \\
&= \pm\inf_{c>0}\left[\frac{1}{c}\log e^{\pm c\tilde{\beta}_{y_i,\omega_l}(\beta_{l0}-\beta_{l0}) + \frac{1}{2}\tilde{\beta}^2_{y_i,\omega_l}\sigma_l^2 c^2} + \frac{\eta_l}{c}\right] \\
&= \pm\inf_{c>0}\left[\frac{1}{2}\tilde{\beta}^2_{y_i,\omega_l}\sigma_l^2 c + \frac{\eta_l}{c}\right] \\
&= \pm|\tilde{\beta}_{y_i,\omega_l}|\sqrt{2\sigma_l^2\eta_l}
\end{aligned} \tag{B3-36}$$

(b) The result follows immediately from (a). ■

Remark: In our case, for the ORR PGM constructed in Section 4.1.4 in Main Text, the optimizing probabilities $Q_\pm$ for $I^\pm(f(X), P; \mathcal{D}_{l,P}^{\eta_l})$, are given by

$$Q_\pm(\omega_l) \sim \mathcal{N}(\beta_{l0} \pm \sqrt{2\sigma_l^2\eta_l}, \sigma_l^2)\,. \tag{B3-37}$$

Remark: Theorem 4.1 is a special case of a more general Theorem (Theorem 3.5), which shows that model-form sensitivity indices (4.18) are computable under some constraints on the graph of the PGM, here we show the result only for our specific ORR PGM example.

B.4 Complexity of the Model-Form Indices

Here we discuss briefly the complexity of the proposed model-form indices. Note that we focus on the complexity with respect to the structure of PGMs, and ignore the complexity of computing the expectation [66]. For the model misspecification between two PGMs, i.e. $\eta = R(Q||P)$, by (3.49), we have

$$\begin{aligned} R(Q||P) &= \sum_{i=1}^{n} \mathbb{E}_{\pi_i}\left[R(Q_{i|\pi_i}||P_{i|\pi_i})\right] \\ &= \sum_{i=1}^{n} \int\int \log\frac{Q(dx_i|x_{\pi_i})}{P(dx_i|x_{\pi_i})} Q(dx_i|x_{\pi_i}) \prod_{k\in\rho_i} Q(dx_k|x_{\pi_k}) \end{aligned} \tag{B3-38}$$

therefore, the complexity of the calculation of η depends on the complexity of the model misspecification on each component, i.e., $\eta_i = \mathbb{E}_{\pi_i}\left[R(Q_{i|\pi_i}||P_{i|\pi_i})\right]$. First, we note that if we have an explicit formula for η_l, which has complexity $O(1)$ (see the GBN example below (B3-47)), then the complexity of η is $O(n)$. In general, if we know the density functions of P and Q, we can compute/estimate η_i by Monte Carlo method with samples or given data set $\mathcal{S}$, i.e.,

$$\eta_i \approx \frac{1}{|\mathcal{S}|} \sum_{(x_{\rho_i}, x_i)\in\mathcal{S}} \log\frac{q(x_i|x_{\pi_i})}{p(x_i|x_{\pi_i})} q(x_i|x_{\pi_i}) \prod_{k\in\rho_i} q(x_k|x_{\pi_k}) \tag{B3-39}$$

which has complexity $O(|\rho_i|)$ ($|\rho_i|$ is the number of indices in set ρ_i), so in all, the complexity of η is $O(\sum_{i=1}^{n} |\rho_i|)$. Moreover, with given baseline model P, we can also estimate the model misspecification η between the unknown exact/real model

Q^* and P (as shown in previous subsection) by the given data [105], e.g. if $\pi_i = \emptyset$, considering the empirical distribution

$$Q_i(x) = \frac{1}{|\mathcal{S}|} \sum_{x_i \in \mathcal{S}} U(x - x_i) \tag{B3-40}$$

where $U(x)$ is the unit-step function with $U(0) = 0.5$, we can estimate η_i by

$$\eta_i = \frac{1}{m} \sum_{k=1}^{m} \log \frac{(Q_i(x_{i,k}) - Q_i(x_{i,k-1}))/(x_{i,k} - x_{i,k-1})}{p(x_i)} \tag{B3-41}$$

where $\{x_{i,k}\}_{k=1}^{m}$ is the samples of X_i sorted in increasing order.

Then for the model-form UQ indices defined in (3.84),

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta}) = \pm \inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{\{k\}}} \left[e^{\pm c \bar{f}(X_k)} \right] + \frac{\eta}{c} \right] \tag{B3-42}$$

the complexity of the calculation of the indices (with given η) depends on the complexity of the moment generating function (MGF) $\mathbb{E}_{P_{\{k\}}} \left[e^{\pm c \bar{f}(X_k)} \right]$. Therefore, if there is an explicit form for the MGF (e.g. (B3-48)), then we can solve the minimization problem for the indices explicitly, and the complexity is $O(1)$. In general, we can evaluate the MGF by Monte Carlo methods as we discuss above, which has complexity $O(|\rho_k|)$, then the complexity of the calculation for the model-form UQ indices is $O(|\rho_k|)$.

Similarly, for the model-form sensitivity indices,

$$I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) = \pm \mathbb{E}_{P_{\rho_l}} \left[\inf_{c>0} \left[\frac{1}{c} \log \mathbb{E}_{P_{l|\pi_l}} \left[e^{\pm c \bar{F}(X_l, X_{\rho_l})} \right] + \frac{\eta_l}{c} \right] \right] \tag{B3-43}$$

then the complexity of the calculation with a given explicit MGF of $F(X_l, X_{\rho_l})$ could be $O(|\rho_l|)$, or $O(1)$ if it is independent of x_{ρ_k} (see the GBN example below (B3-50)). In general, since

$$F(X_l, X_{\rho_l}) = \mathbb{E}_{P_{\{k\}|\rho_l \cup \{l\}}} \left[f(X_k) \right] \tag{B3-44}$$

the complexity of evaluation for $F(X_l, X_{\rho_l})$ could be $O(|\rho_k|-|\rho_l|)$, and the comlexity of the model-form sensitivity indices would be $O(|\rho_l|(|\rho_k| - |\rho_l|))$.

Example (Gaussian Bayesian Network): Consider two GBNs P, Q defined as in (3.29), where

$$p(x_i|x_{\pi_i}) = \mathcal{N}(\beta_{i0} + \beta_i^T x_{\pi_i}, \sigma_i^2) \tag{B3-45}$$

and

$$q(x_i|x_{\pi_i}) = \mathcal{N}(\beta_{i0} + \beta_i^T x_{\pi_i}, \tilde{\sigma}_i^2) \ . \tag{B3-46}$$

Then

$$\eta_i = \mathbb{E}_{\pi_i}\left[R(Q_{i|\pi_i}||P_{i|\pi_i})\right] = \mathbb{E}_{\pi_i}\left[\log(\frac{\sigma_i}{\tilde{\sigma}_i}) + \frac{\tilde{\sigma}_i^2}{2\sigma_i^2} - \frac{1}{2}\right] = \log(\frac{\sigma_i}{\tilde{\sigma}_i}) + \frac{\tilde{\sigma}_i^2}{2\sigma_i^2} - \frac{1}{2} \tag{B3-47}$$

which has complexity $O(1)$, so the complexity of η is $O(n)$ by (B3-38). Furthermore for $f(X_k) = X_k$, we have

$$\mathbb{E}_{P_{\{k\}}}\left[e^{\pm c\bar{f}(X_k)}\right] = e^{\frac{1}{2}c^2\mathcal{C}_{kk}} \tag{B3-48}$$

where $\mathcal{C}_{kk}$ is the variance of the marginal distribution $P_{\{k\}}$ for X_k. Then for the model-form UQ indices, we can solve the minimization problem explicitly,

$$I^{\pm}(f(X_k), P; \mathcal{D}^{\eta}) = \pm \inf_{c>0}\left[\frac{1}{c}\log(e^{\frac{1}{2}c^2\mathcal{C}_{kk}}) + \frac{\eta}{c}\right] = \pm\sqrt{2\mathcal{C}_{kk}\eta} \tag{B3-49}$$

so the complexity is $O(1)$. Moreover, for the model-form sensitivity indices, consider the case $F(X_l, X_{\rho_l}) = \beta_{kl}X_l + \tilde{\beta}_{k0}$ for some constants β_{kl}, $\tilde{\beta}_{k0}$, as shown in a previous example (3.71). Then we have

$$\mathbb{E}_{P_{l|\pi_l}}\left[e^{\pm c\bar{F}(X_l, X_{\rho_l})}\right] = e^{\frac{1}{2}\beta_{kl}^2 c^2 \sigma_l^2} \tag{B3-50}$$

which has an explicit form and independent of x_{ρ_k}, therefore, we can also solve the minimization problem explicitly,

$$I^{\pm}(f(X_k), P; \mathcal{D}_l^{\eta_l}) = \pm\mathbb{E}_{P_{\rho_l}}\left[\inf_{c>0}\left[\frac{1}{c}\log(e^{\frac{1}{2}\beta_{kl}^2 c^2\sigma_l^2}) + \frac{\eta_l}{c}\right]\right] = \pm|\beta_{kl}|\sqrt{2\sigma_l^2\eta_l} \tag{B3-51}$$

and the complexity of the calculation is $O(1)$.

BIBLIOGRAPHY

[1] Brian M Adams, Mohamed Salah Ebeida, Michael S Eldred, John Davis Jakeman, Laura Painton Swiler, John Adam Stephens, Dena M Vigil, Timothy Michael Wildey, William J Bohnhoff, John P Eddy, et al. Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis. Technical report, Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States), 2014.

[2] H. Agarwal, J.E. Renaud, E.L. Preston, and D. Padmanabhan. Uncertainty quantification using evidence theory in multidisciplinary design optimization. *Reliability Engineering & System Safety*, 85(13):281–294, 2004.

[3] DR Anderson and K Burnham. Model selection and multimodel inference: a practical information-theoretic approach. *Springer Science & Business Media*, 2003.

[4] Olivier Antoine, Yann Bultel, and Robert Durand. Oxygen reduction reaction kinetics and mechanism on platinum nanoparticles inside nafion®. *Journal of Electroanalytical Chemistry*, 499(1):85–94, 2001.

[5] Georgios Arampatzis, Markos A Katsoulakis, and Luc Rey-Bellet. Efficient estimators for likelihood ratio sensitivity indices of complex stochastic dynamics. *The Journal of Chemical Physics*, 144(10):104107, 2016.

[6] Dominik Aronsky and Peter J Haug. Diagnosing community-acquired pneumonia with a bayesian network. In *Proceedings of the AMIA Symposium*, page 632. American Medical Informatics Association, 1998.

[7] A. Ayeb, W.M. Otten, A.J.G. Mankb, and P.H.L. Notten. The hydrogen evolution and oxidation kinetics during overdischarging of sealed nickel-metal hydride batteries. *Journal of the Electrochemical Society*, 153(11):A2055–A2065, 2006.

[8] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Computing in Science and Engineering*, 4(3):56–66, 2002.

[9] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[10] George Blyholder. Molecular orbital view of chemisorbed carbon monoxide. *The Journal of Physical Chemistry*, 68(10):2772–2777, 1964.

[11] P. E. Blchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, 1994. PRB.

[12] J. Bromley, N. A. Jackson, O. Clymer, A. M. Giacomello, and F. V. Jensen. The use of hugin® to develop bayesian networks as an aid to integrated water resource planning. *Environmental Modelling & Software*, 20(2):231–242, 2005.

[13] Federico Calle-Vallejo, Jakub Tymoczko, Viktor Colic, Quang Huy Vu, Marcus D Pohl, Karina Morgenstern, David Loffreda, Philippe Sautet, Wolfgang Schuhmann, and Aliaksandr S Bandarenka. Finding optimal surface sites on heterogeneous catalysts by counting nearest neighbors. *Science*, 350(6257):185–189, 2015.

[14] Rogan Carr, Jeffrey Comer, Mark D. Ginsberg, and Aleksei Aksimentiev. Microscopic perspective on the adsorption isotherm of a heterogeneous surface. *The Journal of Physical Chemistry Letters*, 2(14):1804–1807, 2011.

[15] George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, 2001.

[16] Hei Chan. Sensitivity analysis of bayesian networks and its application for service engineering. In *Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of*, pages 551–556. IEEE, 2009.

[17] M.W. Chase Jr. Nist-janf thermochemical tables, 2016. Accessed: 2016-05-31.

[18] I. Chorkendorff and J.W. Niemantsverdriet. *Concepts of Modern Catalysis and Kinetics*. Wiley, 2006.

[19] S. Conrady and L. Jouffe. Introduction to bayesian networks & bayesialab. *September*, 3(201):3, 2013.

[20] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[21] Thierry Crestaux, Olivier Le Maıtre, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, 2009.

[22] S. Da Veiga, F. Wahl, and F. Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463, 2009.

[23] B. deB. Darwent. Bond dissociation energies in simple molecules. *Natl. Stand. Ref. Data Ser. (U.S., Natl. Bur. Stand.)*, 31:32,41, 1970.

[24] M.E. Davis and R.J. Davis. *Fundamentals of chemical reaction engineering*. McGraw-Hill, New York, NY, 2003.

[25] G. N. Derry and P. N. Ross. A work function change study of oxygen adsorption on pt(111) and pt(100). *The Journal of Chemical Physics*, 82(6):2772–2778, 1985.

[26] Pham Luu Trung Duong, Wahid Ali, Ezra Kwok, and Moonyong Lee. Uncertainty quantification and global sensitivity analysis of complex chemical process using a generalized polynomial chaos approach. *Computers & Chemical Engineering*, 90:23–30, 2016.

[27] Paul Dupuis and Richard S Ellis. *A weak convergence approach to the theory of large deviations*, volume 902. John Wiley & Sons, 2011.

[28] Paul Dupuis, Markos A. Katsoulakis, Yannis Pantazis, and Petr Plecháč. Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):80–111, 2016.

[29] Paul Dupuis, Markos A Katsoulakis, Yannis Pantazis, and Petr Plechác. Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):80–111, 2016.

[30] J Durst, A Siebel, C Simon, F Hasche, J Herranz, and HA Gasteiger. New insights into the electrochemical hydrogen oxidation and evolution reaction mechanism. *Energy & Environmental Science*, 7(7):2255–2260, 2014.

[31] Michael S. Eldred and Brian M. Adams and David M. Gay and Laura P. Swiler and Karen Haskell and William J. Bohnhoff and John P. Eddy and William E. Hart and Jean-paul Watson and Patricia D. Hough and Tammy G. Kolda. DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.0 User's Manual. *Sandia*, 2009.

[32] H. A. Engelhardt and D. Menzel. Adsorption of oxygen on silver single crystal surfaces. *Surface Science*, 57(2):591–618, 1976.

[33] Jinchao Feng, Joshua Lansford, Alexander Mironenko, Davood Babaei Pourkargar, Dionisios G Vlachos, and Markos A Katsoulakis. Non-parametric correlative uncertainty quantification and sensitivity analysis: Application to a langmuir bimolecular adsorption model. *AIP Advances*, 8(3):035021, 2018.

[34] Uwe Fink, T. A. Wiggins, and D. H. Rank. Frequency and intensity measurements on the quadrupole spectrum of molecular hydrogen. *Journal of Molecular Spectroscopy*, 18(4):384–395, 1965.

[35] H.S. Fogler. *Elements of Chemical Reaction Engineering*. Prentice Hall PTR, Upper Saddle River, NJ, 2006.

[36] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning, Chapter 8*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[37] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using bayesian networks to analyze expression data. *Journal of computational biology*, 7(3-4):601–620, 2000.

[38] Robert Fung and Brendan Del Favero. Applying bayesian networks to information retrieval. *Communications of the ACM*, 38(3):42–ff, 1995.

[39] Hubert A Gasteiger and Nenad M Marković. Just a dreamor future reality? *science*, 324(5923):48–49, 2009.

[40] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, 2014.

[41] Z. Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, volume 521, pages 452459 (2015)

[42] Roger G Ghanem and Pol D Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[43] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, NY, 2004.

[44] P.W. Glynn *Likelihood ratio gradient estimation: an overview*. STANFORD UNIV CA DEPT OF OPERATIONS RESEARCH, 1987

[45] C.F. Goldsmith, A.S. Tomlin, and Klippenstein S.J. Uncertainty propagation in the derivation of phenomenological rate coefficients from theory: A case study of n-propyl radical oxidation. *Proc. Combust. Inst.*, 34(1):177–185, 2013.

[46] Miguel A Gómez-Villegas, Paloma Maín, and Rosario Susi. Sensitivity analysis in gaussian bayesian networks using a divergence measure. *Communications in StatisticsTheory and Methods*, 36(3):523–539, 2007.

[47] Miguel A Gómez-Villegas, Paloma Main, and Rosario Susi. The effect of block parameter perturbations in gaussian bayesian networks: Sensitivity and robustness. *Information Sciences*, 222:439–458, 2013.

[48] Miguel Ángel Gómez-Villegas, Paloma Main, Hilario Navarro, and Rosario Susi. Evaluating the difference between graph structures in gaussian bayesian networks. *Expert Systems with Applications*, 38(10):12409–12414, 2011.

[49] Miguel Angel Gómez-Villegas, Paloma Main, and Paola Viviani. Sensitivity to evidence in gaussian bayesian networks using mutual information. *Information Sciences*, 275:115–126, 2014.

[50] Konstantinos Gourgoulias, Markos A Katsoulakis, Luc Rey-Bellet, and Jie Wang. How biased is your model? concentration inequalities, information and model bias. *arXiv preprint arXiv:1706.10260*, 2017.

[51] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132(15):154104, 2010.

[52] Axel Gro. Coverage effects in the adsorption of h2 on pd(100) studied by ab initio molecular dynamics simulations. *The Journal of Chemical Physics*, 135(17):174707, 2011.

[53] Axel Gro and Arezoo Dianat. Hydrogen dissociation dynamics on precovered pd surfaces: Langmuir is still right. *Physical Review Letters*, 98(20):206107, 2007.

[54] S. Gu, B. Xu, and Y. Yan. Electrochemical energy engineering: A new frontier of chemical engineering innovation. *Annual Review of Chemical and Biomolecular Engineering*, 5:429–454, 2014.

[55] M.P.R. Haaker and P.J.T. Verheijen. Local and global sensitivity analysis for a reactor design with parameter uncertainty. *Chemical Engineering Research and Design*, 183(5):591–598, 2004.

[56] E. J. Hall and M. A. Katsoulakis. Robust information divergences for model-form uncertainty arising from sparse data in random PDE. *To appear, SIAM Journal Uncertainty Quantification*, August 2018.

[57] B. Hammer. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical Review B*, 59(11):7413–7421, 1999.

[58] Danielle A Hansgen, Dionisios G Vlachos, and Jingguang G Chen. Using first principles to predict bimetallic catalysts for the ammonia decomposition reaction. *Nature chemistry*, 2(6):484, 2010.

[59] D. Haughton, A. Kamis, and P. Scholten. A review of three directed acyclic graphs software packages: Mim, tetrad, and winmine. *The American Statistician*, 60(3):272–286, 2006.

[60] David Heckerman, Eric Horvitz, and Bharat N Nathwani. Toward normative expert systems part i. *Methods of information in medicine*, 31, 2016.

[61] Adam Holewinski and Suljo Linic. Elementary mechanisms in electrocatalysis: revisiting the orr tafel slope. *Journal of The Electrochemical Society*, 159(11):H864–H870, 2012.

[62] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 6. Springer, 2013.

[63] Eric M. Karp, Charles T. Campbell, Felix Studt, Frank Abild-Pedersen, and Jens K. Nrskov. Energetics of oxygen adatoms, hydroxyl species and water dissociation on pt(111). *The Journal of Physical Chemistry C*, 116(49):25772–25776, 2012.

[64] Markos A Katsoulakis, Luc Rey-Bellet, and Jie Wang. Scalable information inequalities for uncertainty quantification. *Journal of Computational Physics*, 336:513–545, 2017.

[65] S.J. Klippenstein, L.B. Harding, M.J. Davis, A.S. Tomlin, and R.T. Skodje. Uncertainty driven theoretical kinetics studies for CH_3OH ignition: HO_2 + CH_3OH and O_2 + CH_3OH. *Proc. Combust. Inst.*, 33(1):351–357, 2011.

[66] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[67] Sadanori Konishi and Genshiro Kitagawa. *Information criteria and statistical modeling*. Springer Science & Business Media, 2008.

[68] G. Kresse and J. Furthmller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996.

[69] G. Kresse and J. Furthmller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169–11186, 1996. PRB.

[70] G. Kresse and J. Hafner. *Ab initio* molecular dynamics for liquid metals. *Physical Review B*, 47(1):558–561, 1993. PRB.

[71] G. Kresse and J. Hafner. *Ab initio* molecular-dynamics simulation of the liquid-metal˘amorphous-semiconductor transition in germanium. *Physical Review B*, 49(20):14251–14269, 1994. PRB.

[72] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, 1999. PRB.

[73] S Kucherenko et al. A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. *Computer Physics Communications*, 181(7):1212–1217, 2010.

[74] S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. *Comput. Phys. Commun.*, 183(4):937–946, 2012.

[75] Henry Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016.

[76] Steffen L Lauritzen. Causal inference from graphical models. *Complex stochastic systems*, pages 63–107, 2001.

[77] Steffen L Lauritzen and David J Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 157–224, 1988.

[78] Elena Lazkano, Basilio Sierra, Aitzol Astigarraga, and José Maria Martinez-Otzeta. On the use of bayesian networks to develop behaviours for mobile robots. *Robotics and Autonomous Systems*, 55(3):253–265, 2007.

[79] O. Le Maitre and O.M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer, New York, NY, 2010.

[80] S.J. Lee, S. Mukerjee, E.A. Ticianelli, and J. McBreen. Electrocatalysis of CO tolerance in hydrogen oxidation reaction in PEM fuel cells. *Electrochimica Acta*, 44(19):3283–3293, 1999.

[81] Tod S Levitt, John Mark Agosta, and Thomas O Binford. Model-based influence diagrams for machine vision. In *Machine Intelligence and Pattern Recognition*, volume 10, pages 371–388. Elsevier, 1990.

[82] P. Li and Q.D. Vu. Identification of parameter correlations for parameter estimation in dynamic biological models. *BMC Systems Biology*, 7(1)(1):1–12, 2013.

[83] S. Liu, T. Ishimoto, D.S. Monder, and M. Koyama. First-principles study of oxygen transfer and hydrogen oxidation processes at the Ni-YSZ-gas triple phase boundaries in a solid oxide fuel cell anode. *The Journal of Physical Chemistry C*, 119(49):27603–27608, 2015.

[84] Laetitia Loulergue, Adrian Schilt, Renato Spahni, Valrie Masson-Delmotte, Thomas Blunier, Bndicte Lemieux, Jean-Marc Barnola, Dominique Raynaud, Thomas F. Stocker, and Jrme Chappellaz. Orbital and millennial-scale features of atmospheric ch4 over the past 800,000years. *Nature*, 453:383, 2008.

[85] Niklas Lüdtke, Stefano Panzeri, Martin Brown, David S Broomhead, Joshua Knowles, Marcelo A Montemurro, and Douglas B Kell. Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. *Journal of The Royal Society Interface*, 5(19):223–235, 2008.

[86] T.A. Mara and Tarantola S. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121, 2012.

[87] R. Masel. *Chemical kinetics and catalysis*. Wiley, New York, NY, 2001.

[88] D.A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000.

[89] A. L. Madsen, M. Lang, U. B. Kjærulff, and F. Jensen. The hugin tool for learning bayesian networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 594–605. Springer, 2003.

[90] Alexander V. Mironenko, Matthew J. Gilkey, Paraskevi Panagiotopoulou, Gregory Facas, Dionisios G. Vlachos, and Bingjun Xu. Ring activation of furanic compounds on ruthenium-based catalysts. *The Journal of Physical Chemistry C*, 119(11):6075–6085, 2015.

[91] J.D. Monkhurst, J.J.and Pack. Special points for brillouin-zone integrations. *Physical Review B, Solid State*, 13:5188–5192, 1976.

[92] J.M. Murphy, D.M.H. Sexton, D.N. Barnett, G.S. Jones, M.J. Webb, M. Collins, and D.A. Stainforth. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430:768–772, 2004.

[93] T. Nagasawa and K. Hanamura. Theoretical analysis of hydrogen oxidation reaction in solid oxide fuel cell anode based on species territory adsorption model. *Journal of Power Sources*, 290:168–182, 2015.

[94] T. Nagy and T. Turnyi. Determination of the uncertainty domain of the arrhenius parameters needed for the investigation of combustion kinetic models. *Reliability Engineering & System Safety*, 107:29–34, 2012.

[95] H.N. Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 41:135–52, 2009.

[96] Maria Navarro, Jeroen Witteveen, and Joke Blom. Polynomial chaos expansion for general multivariate distributions with correlated variables. *arXiv preprint arXiv:1406.5483*, 2014.

[97] P. Nissenson, J.L. Thomas, B.J. Finlayson-Pitts, and D. Dabdub. Sensitivity and uncertainty analysis of the mechanism of gas-phase chlorine production from nacl aerosols in the MAGIC model. *Atmospheric Environment*, 42(29):6934 – 6941, 2008.

[98] S Oladyshkin and W Nowak. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering & System Safety*, 106:179–190, 2012.

[99] Daniel R Palo, Jamie D Holladay, Robert T Rozmiarek, Consuelo E Guzman-Leong, Yong Wang, Jianli Hu, Ya-Huei Chin, Robert A Dagle, and Eddie G Baker. Development of a soldier-portable fuel cell power system: Part i: A bread-board methanol fuel processor. *Journal of Power sources*, 108(1-2):28–34, 2002.

[100] Judea Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible reasoning.* Morgan Kaufmann Publishers, Los Altos, 1988.

[101] Judea Pearl. *Causality.* Cambridge university press, 2009.

[102] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Elsevier, 2014.

[103] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: a primer.* John Wiley & Sons, 2016.

[104] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

[105] Fernando Pérez-Cruz. Kullback-leibler divergence estimation of continuous distributions. In *2008 IEEE international symposium on information theory*, pages 1666–1670. IEEE, 2008.

[106] J. Prager, H.N. Najm, and J. Zador. Uncertainty quantification in the ab initio rate-coefficient calculation for the $CH_3CH(OH)CH_3 + OH \rightarrow CH_3C^{\cdot}(OH)CH_3 + H_2O$ reaction. *Proc. Combust. Inst.*, 34(1):583–590, 2013.

[107] A.E. Raftery, D. Madigan, and J.A. Hoeting. Bayesian model averaging for regression models. *Journal of the American Statistical Association*, 92:179–191, 1997.

[108] Sharif Rahman. The f-sensitivity index. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):130–162, 2016.

[109] M.T. Reagan, H.N. Najm, R.G. Ghanem, and O.M. Knio. Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545 – 555, 2003.

[110] P.J. Roache. Quantification of uncertainty in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 29:123–160, 1997.

[111] T. W. Root, L. D. Schmidt, and Galen B. Fisher. Adsorption and reaction of nitric oxide and oxygen on rh(111). *Surface Science*, 134(1):30–45, 1983.

[112] S. J. Russell. For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning. *https://amturing.acm.org/award_winners/pearl_2658896.cfm*, (2011)

[113] Michael Salciccioli, Michail Stamatakis, S Caratzoulas, and Dion G Vlachos. A review of multiscale modeling of metal-catalyzed reactions: Mechanism development for complexity and emergent behavior. *Chemical Engineering Science*, 66(19):4319–4355, 2011.

[114] Saltelli, Andrea and Ratto, Marco and Andres, Terry and Campolongo, Francesca and Cariboni, Jessica and Gatelli, Debora and Saisana, Michaela and Tarantola, Stefano. Global sensitivity analysis: the primer *John Wiley & Sons*, 2018

[115] Brian P Setzler, Zhongbin Zhuang, Jarrid A Wittkopf, and Yushan Yan. Activity targets for nanostructured platinum-group-metal-free catalysts in hydroxide exchange membrane fuel cells. *Nature nanotechnology*, 11(12):1020, 2016.

[116] R.J. Shannon, A.S. Tomlin, S.H. Robertson, M.A. Blitz, M.J. Pilling, and P.W. Seakins. Global uncertainty propagation and sensitivity analysis in the $CH_3OCH_2 + O_2$ system: Combining experiment and theory to constrain key rate coefficients in DME combustion. *J. Phys. Chem A*, 119(28):7430–7438, 2015.

[117] Wenchao Sheng, Hubert A Gasteiger, and Yang Shao-Horn. Hydrogen oxidation and evolution reaction kinetics on platinum: acid vs alkaline electrolytes. *Journal of The Electrochemical Society*, 157(11):B1529–B1536, 2010.

[118] Kaido Sillar, Arpan Kundu, and Joachim Sauer. Ab initio adsorption isotherms for molecules with lateral interactions: Co2 in metalorganic frameworks. *The Journal of Physical Chemistry C*, 121(23):12789–12799, 2017.

[119] R.C. Smith. *Uncertainty quantification: Theory, implementation, and applications*. SIAM, Philadelphia, PA, 2014.

[120] Carolyn W. Snyder. Evolution of global temperature over the past two million years. *Nature*, 538:226, 2016.

[121] R. Scheines, P. Spirtes, C. Glymour, C. Meek, and T. Richardson. Tetrad 3: Tools for causal modeling–users manual. *CMU Philosophy*, 1996.

[122] Ralph C. Smith. Uncertainty Quantification: Theory, Implementation, and Applications. *Society for Industrial and Applied Mathematics*, 2013.

[123] Peter Spirtes, Clark N Glymour, Richard Scheines, David Heckerman, Christopher Meek, Gregory Cooper, and Thomas Richardson. *Causation, prediction, and search*. MIT press, 2000.

[124] Nian-Tzu Suen, Sung-Fu Hung, Quan Quan, Nan Zhang, Yi-Jun Xu, and Hao Ming Chen. Electrocatalysis for the oxygen evolution reaction: recent development and future perspectives. *Chemical Society Reviews*, 46(2):337–365, 2017.

[125] J.E. Sutton, W. Guo, M.A. Katsoulakis, and D.G. Vlachos. Effects of correlated parameters and uncertainty in electronic-structure-based chemical kinetic modeling. *Nature Chemistry*, 2016, DOI: 10.1038/NCHEM.2454.

[126] A.S. Tomlin, E. Agbro, V. Nevrly, J. Dlabka, and M. Vasinek. Evaluation of combustion mechanisms using global uncertainty and sensitivity analyses: A case study for low-temperature dimethyl ether oxidation. *Int. J. Chem. Kinet.*, 46(11):662–682, 2014.

[127] Aradhna K. Tripati, Christopher D. Roberts, and Robert A. Eagle. Coupling of co¡sub¿2¡/sub¿ and ice sheet stability over major climate transitions of the last 20 million years. *Science*, 326(5958):1394–1397, 2009.

[128] N. Vu-Bac, R. Rafiee, X. Zhuang, T. Lahmer, and T. Rabczuk. Uncertainty quantification for multiscale modeling of polymer nanocomposites with correlated parameters. *Composites Part B: Engineering*, 68:446 – 464, 2015.

[129] Haiqin Wang. Using sensitivity analysis to validate bayesian networks for airplane subsystem diagnosis. In *Aerospace Conference, 2006 IEEE*, pages 10–pp. IEEE, 2006.

[130] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, NY, 2004.

[131] L. Wasserman. *All of Nonparametric Statistics*. Springer, New York, NY, 2006.

[132] Alfons Weber and Eugene A. McGinnis. The raman spectrum of gaseous oxygen. *Journal of Molecular Spectroscopy*, 4(1):195–200, 1960.

[133] O. Woodberry and S. Mascaro. *Programming Bayesian network solutions with netica*. Bayesian Intelligence, 2012.

[134] Heng Xiao, J-L Wu, J-X Wang, Rui Sun, and CJ Roy. Quantifying and reducing model-form uncertainties in reynolds-averaged navier–stokes simulations: A data-driven, physics-informed bayesian approach. *Journal of Computational Physics*, 324:115–136, 2016.

[135] Dongbin Xiu. *Numerical methods for stochastic computations: a spectral method approach*. Princeton university press, 2010.

[136] Dongbin Xiu and George Em Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

[137] J. Zador, I.G. Zsely, and T. Turanyi. Local and global uncertainty analysis of complex chemical kinetic systems. *Reliability Engineering & System Safety*, 91(10-11):1232–1240, 2006.

[138] Xiaopin Zhong and Qiu Li. Component importance and sensitivity analysis in bayesian networks. In *Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), 2013 International Conference on*, pages 320–325. IEEE, 2013.

[139] T. Ziehn, K.J. Hughes, J.F. Griffiths, R. Porter, and A.S. Tomlin. A global sensitivity study of cyclohexane oxidation under low temperature fuel-rich conditions using HDMR methods. *Combustion Theory and Modelling*, 13(4):589–605, 2009.

[140] T. Ziehn and A.S. Tomlin. A global sensitivity study of sulfur chemistry in a premixed methane flame model using HDMR. *Int. J. Chem. Kinet.*, 40(11):742–753, 2008.

[141] T. Ziehn and A.S. Tomlin. Gui-HDMR - a software tool for global sensitivity analysis of complex models. *Environmental Modelling & Software*, 24(7):775–, 2009.

[142] I.G. Zsely, J. Zador, and T. Turanyi. Uncertainty analysis of NO production during methane combustion. *Int. J. Chem. Kinet.*, 40(11):754–768, 2008.