APPLIED SCIENCES AND ENGINEERING

Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences

Jinchao Feng[1]*, Joshua L. Lansford[2]*, Markos A. Katsoulakis[3]†, Dionisios G. Vlachos[2,4]†

Data science has primarily focused on big data, but for many physics, chemistry, and engineering applications, data are often small, correlated and, thus, low dimensional, and sourced from both computations and experiments with various levels of noise. Typical statistics and machine learning methods do not work for these cases. Expert knowledge is essential, but a systematic framework for incorporating it into physics-based models under uncertainty is lacking. Here, we develop a mathematical and computational framework for probabilistic artificial intelligence (AI)–based predictive modeling combining data, expert knowledge, multiscale models, and information theory through uncertainty quantification and probabilistic graphical models (PGMs). We apply PGMs to chemistry specifically and develop predictive guarantees for PGMs generally. Our proposed framework, combining AI and uncertainty quantification, provides explainable results leading to correctable and, eventually, trustworthy models. The proposed framework is demonstrated on a microkinetic model of the oxygen reduction reaction.

INTRODUCTION

Models in the chemical and physical sciences have led to both new understanding and new discoveries (*1*) including new materials (*2*, *3*). Physics-based models span orders of magnitude in length and time, ranging from quantum mechanics (*4*) to chemical plants (*5*), and naturally capture physics-based constraints (*6*–*8*). Combining models across scales, known as multiscale modeling (*9*), is necessary when chemical properties are determined at the quantum level, but most experiments and relevant applications exist at the macroscale, such as in heterogeneous catalysis. At the core of model development lies the question of accuracy of a physics-based model. Going beyond sensitivity analysis (*10*, *11*), there has been growing interest in quantifying uncertainty, resulting from correlations in parameters (*12*, *13*) along with other sources of error arising in predicting new materials (*14*). In addition to ensuring trustworthiness, error quantification can enable model correctability (*15*, *16*). Still, uncertainty is an afterthought in actual physics-based model development. Currently, a model is first built deterministically without systematically accounting for the effect of both modeling errors and lack, or sparseness, of data.

Modeling uncertain data has experienced tremendous advances in data science (*17*–*20*); however, the corresponding models are empirical, can fail without guarantee, and can violate conservation laws and constraints. Current approaches for handling data based on physical laws and chemical theory are, in a sense, not truly probabilistic and require correlations and causal relationships to be known a priori. With the increasing size of chemistry datasets, it is almost impossible to apply traditional methods of model development to systems with many sources of interacting error. Global sensitivity techniques, such as Sobol indices, attribute model variance to model variables and their interaction (hereafter called "parametric uncertainty") (*21*). However, there are few methods that work beyond first-order interactions or quantify the importance of missing physics or submodels rather than parameters (hereafter called "model uncertainty") (*22*). Methods that do exist model missing physics as stochastic noise that has no structure (*23*, *24*). Therefore, there is a need to develop methods that both attribute interaction error directly to model inputs and provide predictive guarantees and, by doing this, to make models correctable and eventually trustworthy for predictions and design tasks.

Here, we address these issues by incorporating error and uncertainty directly into the design of a model. First, we introduce the use of Bayesian networks (*20*), a class of probabilistic graphical models (PGMs) (*25*), common in probabilistic artificial intelligence (AI) (*26*), to integrate simultaneously and systematically physics- and chemistry-based models, data, and expert knowledge. This framework is termed C-PGM (chemistry-PGM). Second, we derive global uncertainty indices that quantify model uncertainty stemming from different physics submodels and datasets. This framework generates predictive "worst-case" guarantees for Bayesian networks while handling correlations and causations in heterogeneous data for both parametric and model uncertainties and is based on recent work in robust methods for quantifying probabilistic model uncertainty (*27*, *28*). Our proposed framework, combining AI and uncertainty quantification (UQ), systematically apportions and quantifies uncertainty to create interpretable and correctable models; this is performed through assimilation of data and/or improvement of physical models to enable trustworthy AI for chemical sciences. We reduce the complexity of the nondeterministic polynomial time (NP)–hard problem of learning a PGM by leveraging expert knowledge of the underlying chemistry. We demonstrate this framework in the prediction of the optimal reaction rate and oxygen binding energy for the oxygen reduction reaction (ORR) using the volcano model. While UQ has been applied to deterministic volcano-based models in general (*29*), and the ORR model specifically (*30*), prior methods have been limited both by the physics model's underlying structure and, importantly, the lack in interpretability of uncertainty in predictions in terms of modeling decisions and available data in different model components. We demonstrate that about half of the model uncertainty stems from density functional theory (DFT) uncertainty, comparable error from lack of sufficient number and quality of experimental data and from correlations in

[1] Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA. [2] Department of Chemical and Biomolecular Engineering, University of Delaware,150 Academy Street, Colburn Laboratory Newark, DE 19716, USA. [3] Department of Mathematics and Statistics, University of Massachusetts at Amherst, Amherst, MA 01003, USA. [4] Catalysis Center for Energy Innovation, University of Delaware, 221 Academy Street, 250R, Newark, DE 19716, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: markos@math.umass.edu (M.A.K.); vlachos@udel.edu (D.G.V.)

parameters (~20% each), and the remaining (~10%) from the solvation model. This analysis provides a blueprint for prioritizing model components toward correctability and improved trustworthiness by underscoring the need foremost of more accurate electronic structure calculations and secondary by better experiments. We illustrate model correctability with an example.

RESULTS

Physics model for the ORR and the deterministic volcano

Hydrogen fuel cells can nearly double the efficiency of internal combustion engines and leave behind almost no emissions, especially if environmentally low footprint H_2 is available (*31*). Furthermore, the hydrogen fuel cell is a mature technology that produces electricity via the hydrogen oxidation reaction at the anode and the ORR at the cathode (Fig. 1B); polymer electrolyte membrane fuel cells for this type of reaction are commercially available (*32*). Because of the high cost of platinum (Pt) catalyst and stability problems of other materials in an acidic electrolyte, recent focus has been on developing alkaline electrolytes. This technology, while extremely promising, results in slower reaction rates (by ~2 orders of magnitude compared to a Pt/acidic electrolyte) and thus bigger devices (*33*, *34*). Overcoming slower rates with stable materials requires discovery of new, multicomponent catalysts, e.g., core-shell alloys.

The ORR depends on the formation of surface hydroperoxyl (OOH*), from molecular oxygen (O_2), and of water (H_2O), from surface hydroxide (OH*) (*35*). The complete mechanism (*7*, *36*, *37*) involves four electron steps (Fig. 1A) and is described in detail in section S1. Among these, reactions R1 and R4 are slow (*7*). Acceleration of the ORR then translates into finding materials that speed up the slower of the two reactions, R1 and R4. An approach to find new materials entails generation of an activity model (Fig. 1C) as a function of descriptor(s) that can be estimated quickly using DFT calculations (*9*). This is known as the deterministic volcano (Sabatier's principle) and has been the key model for discovery of new materials.

Next, we discuss the human workflow in constructing the volcano curve. First, we use a physics equilibrium model to compute the rate *r* from the minimum free energy of reactions R1 and R4 (*7*, *38*), such that

$$r = e^{\frac{\min\{-\Delta G_1, -\Delta G_4\}}{k_B T}} \qquad (1)$$

where k_B is the Boltzmann constant and T is the temperature. Instead of Eq. 1, one could use a more elaborate model, such as a mean-field microkinetic (detailed reaction mechanism) or a kinetic Monte Carlo model, which is a more complex multiscale model. Such models impose conservation laws (mass conservation and catalyst site balance) and are selected on the basis of expert knowledge. The Gibbs free energy ΔG_i of the ith species is calculated from the electronic energy (E_{DFT}) and includes the zero-point energy, temperature effects, and an explicit solvation energy (E_{solv}) in water, as detailed in Methods. This calculation entails, again, physics-based models (statistical mechanics here) and expert knowledge, e.g., in selecting a solvation model and statistical mechanics models. See section S1 for an explanation of the equilibrium model and resulting formula in Eq. 1.

The free energies ΔG_1 and ΔG_4 are computed as linear combinations of the free energies of species, while accounting for stoichiometry (a constraint), and are regressed versus ΔG_O (the descriptor); see data in Fig. 1C. Typically, only two to three data points for coinage metals (Ag, Au, and Cu) on the right leg of the volcano are regressed, especially if experimental data (instead of DFT data) are used (data corresponding to dotted lines are not observed in most experiments). The intersection of the two lines (Fig. 1C) determines the maximum of the volcano curve and provides optimal material properties, i.e., the ΔG_O^*; ΔG_O^* can then be matched to values of multicomponent materials to obtain materials closer to the tip of the volcano. This "human workflow" (Fig. 1D) provides a blueprint of the deterministic overall model that relies exclusively on expert knowledge in design and various physical submodels (called also components) for estimation of several key quantities.

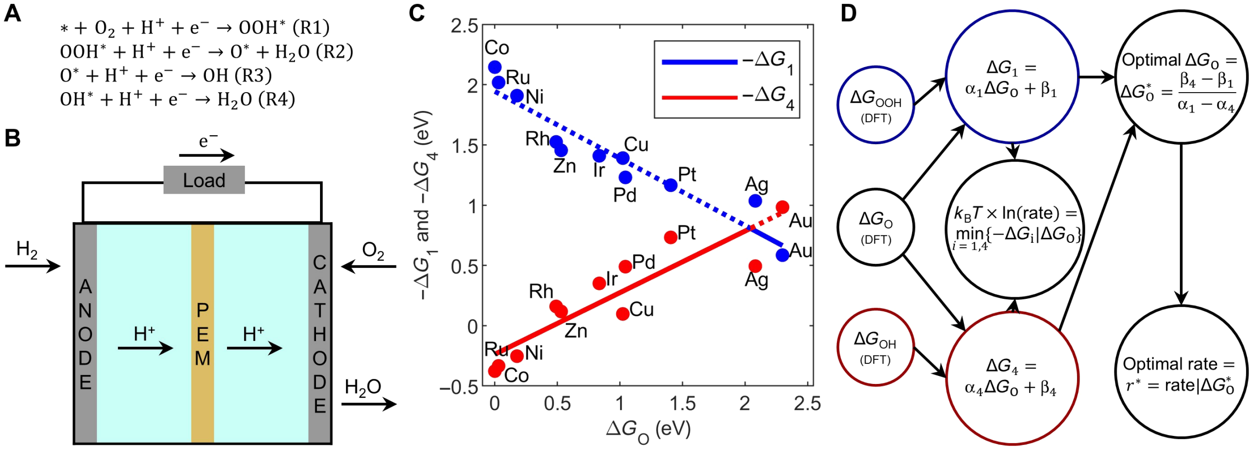

Fig. 1. Fuel cell schematic with workflow and DFT data for estimating the optimal rate and properties of best materials. (**A**) Key reaction steps (R1 to R4) in hydrogen fuel cells. R1, solvated O_2 forms adsorbed OOH*; R2, OOH* forms adsorbed surface oxygen O* and solvated H_2O; R3, O* forms adsorbed OH*; R4, H_2O forms and regenerates the free catalyst site. Asterisk (*) represents an unoccupied metal site or an adsorbed species; H^+ and e^- refer to proton and electron, respectively. (**B**) Schematic of a hydrogen fuel cell. (**C**) Negative changes in Gibbs energies ($-\Delta G_1$ and $-\Delta G_4$) for reactions R1 (blue) and R4 (red) on the close packed (111/0001) surface of face-centered (fcc) and hexagonal close-packed (hcp) metals for the most stable sites of OOH*, OH*, and O* computed (specifically for this work) via DFT (circles) and linear regressions (lines). The optimal oxygen free energy ΔG_O^* is the intersection of the two lines. The min{$-\Delta G_1, -\Delta G_4$}, indicated by the solid lines, determines the rate, estimated using Eq. 1. The optimal rate occurs at ΔG_O^*. (**D**) Deterministic "human" workflow for obtaining the optimal formation energy of surface oxygen and the rate of the ORR.

Probabilistic AI for chemistry and the probabilistic volcano

Here, we develop a probabilistic AI-based workflow that augments the human workflow (in Fig. 1D) to create a probabilistic volcano. The mathematical tool we use to formulate the probabilistic volcano is the PGM. PGMs represent a learning process in terms of random variables, depicted as vertices of the graphs, which explicitly model their interdependence in terms of a graph. This interdependence stems from (i) one variable influencing others, called causality, depicted by arrows (directed edges); and (ii) correlations among variables, depicted by simple (undirected) edges between vertices (see below). PGMs are defined as the parameterized conditional probability distribution (CPD) and for Bayesian networks are defined such that

$$P(X|\theta) = \prod_{i=1}^{n} P(X_i|\mathrm{Pa}_{X_i}, \theta_{X_i|Pa_{X_i}}) \text{ with CPD:} P(X_i|\mathrm{Pa}_{X_i}, \theta_{X_i|Pa_{X_i}}), i = 1, \ldots, n \qquad (2)$$

$Pa_{Xi} = \{X_{i1} \ldots, X_{im}\} \subset \{X_1 \ldots, X_n\}$ denotes the parents of the random variable X_i, and $\theta = \{\theta_{Xi \mid PaXi}\}^{n}_{i=1}$ are the parameters of each CPD, $P(X_i \mid Pa_{Xi}, \theta_{XiPaXi})$. Uppercase "$P$" indicates a stochastic model or submodel. A key concept in PGMs is that the random variables are conditionally independent. This concept is central to constructing complex probability models with many parameters and variables, enabling distributed probability computations by "divide and conquer" using graph-theoretic model representations. By combining the conditional probabilities in Eq. 2, we find the joint probability distribution of all random variables X. A detailed formalism for the construction of the PGM is given in section S4.

Structure learning of graphical models is, in general, an NP-hard problem (*39*, *40*) if one considers the combinatorial nature in connecting a large number of vertices. We overcome this challenge by constraining the directed acyclic graph (DAG) (*41*), representing the probabilistic ORR volcano (Fig. 2A), using domain knowledge that includes multiscale, multiphysics models discussed above, expert knowledge, and heterogeneous data (experimental and DFT) along with their statistical analysis.

First, statistical analysis of data finds hidden correlations or lack thereof between variables and is also central to building the PGM. In this example, statistical analysis of the computed formation free energy data of O*, OOH*, and OH* indicates correlations among data, i.e., connections between vertices (Fig. 2A). Specifically, ΔG_{OOH} and ΔG_{OH} are correlated with ΔG_O. The correlation coefficients of ΔG_O with $-\Delta G_1$ and $-\Delta G_4$ are −0.95 and 0.91, respectively; see section S4 for notes on statistical independent tests used. Reaction free energies ΔG_1 and ΔG_4 are linear combinations of ΔG_{OOH} and ΔG_{OH}, respectively; we use the reaction free energies as dependent vertices, as the reaction rate depends directly on ΔG_1 and ΔG_4. Subsequently, we choose ΔG_O as the independent node (descriptor) because, of all the surface intermediates, it has the fewest degrees of freedom (and therefore local minima) on any given potential energy surface for faster quantum calculations. The selection of the descriptor, which is

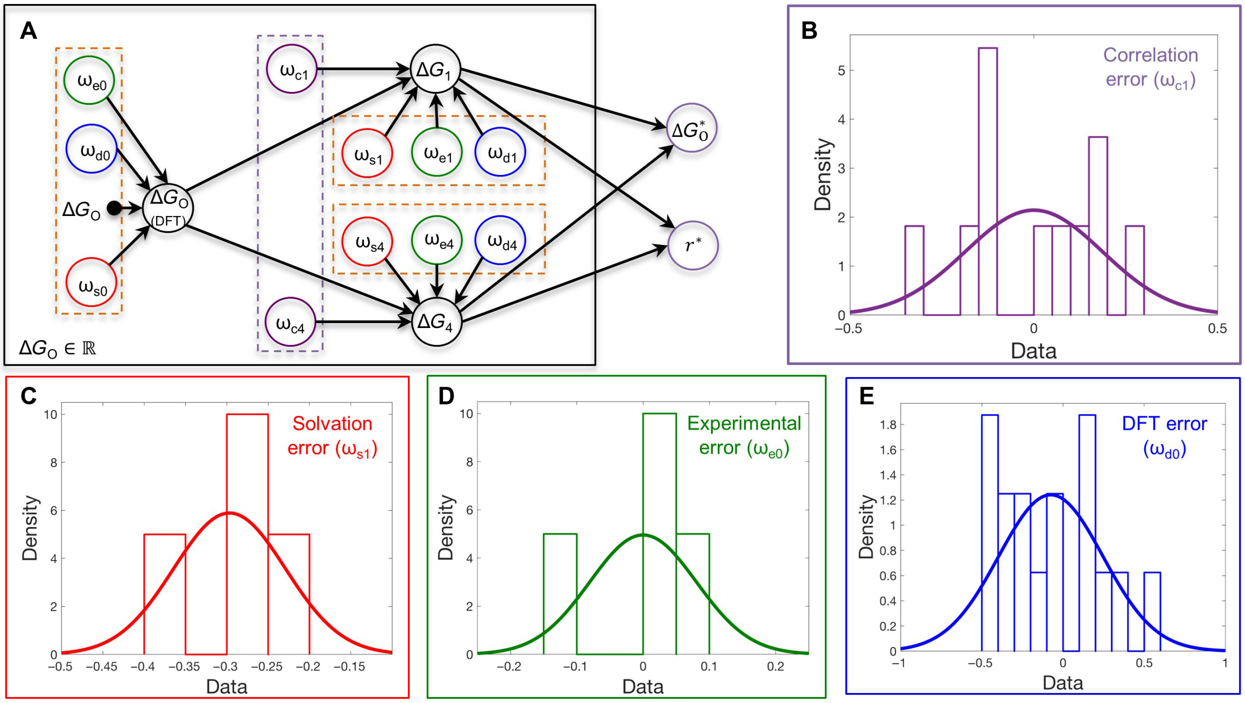

Fig. 2. Construction of the PGM. (**A**) PGM (Eq. 3) for the ORR that combines heterogeneous data, expert knowledge, and physical models; causal relationships are depicted by arrows. The PGM P is a Gaussian Bayesian network where the CPDs are selected to be Gaussians [solid lines as histogram approximations in (B) to (E)]. The PGM is built as follows: We construct $\Delta G_{O\,(DFT)}$ as a random variable from the quantum data for the oxygen binding energy; we include statistical correlations between $\Delta G_{O\,(DFT)}$ and ΔG_1/ ΔG_4 (Fig. 1C) as a random error in correlation (**B**); (**C** to **E**) we model different kinds of errors in the ΔG's, given expert knowledge; we include these random variables into the PGM and build the causal relationships (directed edges/arrows) between the corresponding random variables (ΔG's); we obtain a prediction for the optimal oxygen binding energy (ΔG^*_O) and optimal reaction rate (r^*) using physical modeling, e.g., ΔG^*_O corresponds to the value where ΔG_1 and ΔG_4 are equal in the deterministic case. This entire figure captures the (probabilistic) "AI workflow" that augments the human workflow.

another example of expert knowledge in our C-PGM, establishes causal relationship (direction of influence) represented by directed edges from ΔG_O to ΔG_1 and ΔG_4. Expert knowledge is also leveraged to assign relevant errors (ω) to vertices and directed edges. Figure 2A (colored circles) depicts the multiple uncertainties (random variables) ω modeled in each CPD of the PGM and how these (causally) influence the uncertainty of each vertex. All these causal relationships are modeled by a DAG in Fig. 2A and the Bayesian network in Eq. 3. Causality simplifies the construction and UQ analysis of the PGM. Last, the lack of an edge between ΔG_1 and ΔG_4 (Fig. 2A) is found from conditional independence tests on the DFT data. By eliminating graph edges of uncorrelated parts of the graph, the constrained DAG is profoundly simpler. A complete, step-by-step discussion of the structure learning of the ORR C-PGM is included in section S4.1.

The C-PGM structure contains information from expert knowledge, causalities, physics (physical models, conservation laws, and other constraints), correlations of data, parameters, and hierarchical priors (priors of a prior) in model learning (*13*, *25*). The physical meaning and estimation of these uncertainties are discussed below. Overall, the model for the ORR C-PGM becomes

$$\begin{gathered}\prod\nolimits_{i=\{1,4\}} p(\Delta G_i|\Delta G_{\mathrm{O(DFT)},}\,\omega_{\mathrm{ci}},\omega_{\mathrm{si}},\omega_{\mathrm{ei}},\omega_{\mathrm{di}})\prod\nolimits_{j=\{c,s,e,d\}} p(\omega_{\mathrm{ji}})\\ p(\Delta G_{\mathrm{O(DFT)}}|\omega_{s\mathrm{O}},\omega_{e\mathrm{O}},\omega_{d\mathrm{O},}\,\Delta G_{\mathrm{O}})\prod\nolimits_{k=\{s,e,d\}} p(\omega_{k\mathrm{O}})\end{gathered} \tag{3}$$

where $\Delta G_{O\ (DFT)}$ indicates a calculated value from DFT and all other ΔG values represent the "true value" given errors. Lowercase "*p*" indicates probability densities that are assumed here to be Gaussian, thus rendering the C-PGM (Eq. 3) into a Gaussian Bayesian network (*25*). Note that this PGM is used as part of an optimization scheme where $\Delta G_{O\ (DFT)}$ is formulated as a random variable given any value of the true ΔG_O and distribution of errors for ΔG_O. Leveraging the human-based (deterministic) workflow in Fig. 1D, the ORR is modeled as a stochastic optimization problem such that

$$\Delta G_{\mathrm{O}}^{*} \coloneqq \mathrm{argmax}_{\Delta G_{\mathrm{O}}}[E_P[\min\{-\Delta G_1,-\Delta G_4\}|\Delta G_{\mathrm{O}}]] \tag{4}$$

where ΔG_O^* corresponds to the optimal oxygen binding energy that maximizes the reaction rate r^*. It is convenient to compute $k_B T \ln (r^*)$,

$$k_{\mathrm{B}}\,T\ln(r^*) \coloneqq \max_{\Delta G_{\mathrm{O}}}[E_P[\min\{-\Delta G_1,-\Delta G_4\}|\Delta G_{\mathrm{O}}]] \tag{5}$$

For the rest of this paper, ΔG_O^* and $k_B T \ln (r^*)$ are considered as the QoIs (quantities of interest) that need to be optimized.

Model uncertainty, guarantees, nonparametric sensitivity, and contributions to model error for interacting variables

Model uncertainty arises from multiple sources, such as use of sparse data in Fig. 1C, hidden correlations between vertices in the graph, simplified statistics models (linear regression between free energies in Fig. 1C and Gaussian approximations of errors; Fig. 2, B to E), and uncertainty in different model components and variables. These include errors in experimental data (ω_e), DFT data (ω_d), solvation energies (ω_s), and regressions (correlations) used to determine the optimum ΔG_O^* (ω_c); correlation error is accentuated by the small data available especially on the right leg of the volcano. Experimental errors (ω_e) in ΔG_O and ΔG_{OH} can be found by repeated measurements in the same laboratory and between different laboratories. Repeated calorimetry and temperature-programmed desorption measurements for the dissociative adsorption enthalpy of O_2 in the same and different labs provide a distribution of errors for ΔG_O. The distribution of DFT errors (ω_d) is computed by comparing experimental and calculated (DFT) data across various metals. The mean value and SD of errors are provided in table S1 along with a detailed description of how errors were calculated in Methods and section S3. In Fig. 2A, the "parent vertex" is determined by the direction of the arrow such that ω_{e1} is a parent of ΔG_1 (the child). These additional uncertainties from multiple sources are shown in Fig. 2 (B to E) and are combined to build the PGM model *P*.

When building the C-PGM model *P*, "model uncertainties" arise from the sparsity and quality of the available data in different components of the model, the accuracy of the physics-based submodels, and the knowledge regarding the probability distribution of the errors (Fig. 2, B to E). Consequently, the mean value of $\min\{-\Delta G_1, -\Delta G_4\}$ with respect to *P* is itself uncertain since the probabilistic model *P* is uncertain. For this reason, we consider *P* as a baseline C-PGM model, i.e., a reasonable but inexact "compromise." Here, we take *P* to be a Gaussian Bayesian network (see Fig. 2, B to E), where the error probability distribution function for each component of the model is approximated as a normal distribution and is built using the nominal datasets and submodels discussed above. We isolate model uncertainty in each component (CPD) of the entire model (Eq. 3), in contrast to the more standard parametric (aleatoric) uncertainty already included in the stochasticity of *P* itself. We mathematically represent model uncertainty through alternative (to *P*) models *Q* that include the "true" unknown model *Q**. As examples, models *Q* can differ from *P* by (i) replacing one or more CPDs in Eq. 3 by more accurate, possibly non-Gaussian CPDs that represent better the data in Fig. 2 (B to E), (ii) more accurate multiscale physics models, and (iii) larger and more accurate datasets. Quantifying the impact of model uncertainties on predicting the QoI using *P*, instead of better alternative models *Q*, is discussed next. Overall, developing the mathematical tools to enable identification of the components of a PGM that need improvement is critical to correct the baseline model *P* with minimal resources.

Each model *Q* is associated with its own model misspecification parameter η that quantifies how far an alternative model *Q* is from the baseline model *P* via the Kullback–Leibler (KL) divergence of *Q* to *P*, $R(Q||P)$. We use the KL divergence due to its chain rule properties that allow us to isolate the impact of individual model uncertainties of CPDs in Eq. 2 on the QoIs in Fig. 2, see sections S6 and S7. To isolate and rank the impact of each individual CPD model misspecification (η_l), we consider the set of all PGMs *Q* that are identical to the entire PGM *P* except at the *l*th component CPD (for dependence on the *l*th parents) and less than η_l in KL divergence from the baseline CPD $P(X_l \mid Pa_{Xl})$ while maintaining the same parents Pa_{Xl}. We refer to this family, denoted by D^{η_l}, as the "ambiguity set" for the *l*th CPD of the PGM *P* (see section S6 for its mathematical definition). Given the set of PGM's D^{η_l}, we develop model uncertainty guarantees $\mathcal{I}_l^{\pm}(\mathrm{QoI}, P; \eta_l)$ for the QoI in Eq. 6 as the two worst-case scenarios for all possible models *Q* in D^{η_l} with respect to the baseline *P*

$$\mathcal{I}_l^{\pm}(\mathrm{QoI},P;\eta_l) = (\max/\min)_{Q\in D^{\eta_l}}[E_Q(\mathrm{QoI})-E_P(\mathrm{QoI})] \tag{6}$$

For a given η_l, the model uncertainty guarantees to describe the maximum (worst case) expected bias when only one part of the model in the PGM, $P(X_l \mid Pa_{Xl})$, is perturbed within η_l; therefore, they measure the impact of model uncertainty in any component

(CPD) of the PGM on the QoI. Since η_l are not necessarily small, the method is also nonperturbative, i.e., it is suitable for both small and large model perturbations.

Equation 6 can be also viewed as a nonparametric model sensitivity analysis for PGMs since it involves an infinite dimensional family of model perturbations D^{η_l} of the baseline model P. This family can consider the sparsity of data by addition of new or higher-quality data, e.g., higher-level DFT data, alternative densities to the Gaussians in Fig. 2 (B to E), e.g., richer parametric families or kernel-based CPDs, or more accurate submodels. All these are large, nonparametric perturbations to the baseline P model. For these reasons, Eq. 7 allows one to interpret, reevaluate, and improve the baseline model by comparing the contributions of each CPD to the overall uncertainty of the QoIs through the (model uncertainty) ranking index

$$\text{Ranking Index} = \text{Relative contribution to total model uncertainty} = \frac{\mathcal{I}_l^{\pm}(\text{QoI}, P; \eta_l)}{\Sigma_j \mathcal{I}_j^{\pm}(\text{QoI}, P; \eta_j)} \qquad (7)$$

For more details, see theorem 1 in section S6 where we show that for Gaussian Bayesian networks, the ratios in Eq. 7 are computable.

We can use two strategies regarding η_l. First, η_l can by tuned "by hand" to explore how levels of uncertainty in each component of the model, $P(X_l \mid Pa_{Xl})$, affect the QoIs. This approach is termed a stress test in analogy to finance where in the absence of sufficient data, models are subjected to various plausible or extreme scenarios. Second, instead of treating η_l as a constant, we can estimate η_l as the "distance" between available data from the unknown real model and our baseline PGM P; we refer to such η_l's as data based, in contrast to stress tests (see section S8). For example, the data can be represented by a histogram or a kernel density estimator (KDE) approximation (*42*). In this sense, the contribution to model uncertainty from any error source is both a function of its variance and how far away the error is from the baseline, e.g., the Gaussian CPDs in Fig. 2 (B to E).

Given the error distributions and their Gaussian representation in PGM model P, the expected value of $\min\{-\Delta G_1, -\Delta G_4\} \mid \Delta G_O$ (black curve) in Fig. 3A is obtained. The color bar in Fig. 3A indicates how likely a reaction rate occurs with given ΔG_O for model P (aleatoric uncertainty). The gray dashed lines in Fig. 3A correspond to the two extreme scenarios (derived in section S5) for all possible models Q by considering uncertainty in all components. We highlight in Fig. 3B the expected value (black line) and the extremes (gray dashed lines) when only the DFT error in ΔG_4 is considered. All η values here are data based and determine what models Q are considered in construction of the bounds; only PGMs that have a KL divergence that is less than or equal to η from the baseline are considered. The red, orange, and green lines indicate potential QoIs that can be computed; here, we focus on the uncertainty in the rate (y axis; difference between red lines) and the variability of optimal oxygen binding energy (x axis; difference between orange lines) as a proxy of materials selection; see section S7 for more details.

Using the model uncertainty guarantees (Eq. 6), we quantify the uncertainty and its impact on model predictions beyond the established parametric uncertainty; again, all η_l's are data based. By sourcing the impact of each submodel and/or data, Eq. 7 reveals what data, measurement, and computation should be improved. The error in the optimal reaction rate (Fig. 4A) stems from a nearly equal contribution of submodels, specifically by solvation (30%), experiment (18%), DFT (33%), and parameter correlation (18%). The uncertainty in the optimal oxygen free energy variability (Fig. 4B), i.e., the materials prediction, stems from solvation (6%), experiment (8%), DFT (48%), and parameter correlation (37%). Different QoIs are influenced to a different degree by different submodels. In both QoIs, the DFT error stands as the most influential. The correlation between O*, OOH*, and OH* is the next most important component regarding materials prediction, whereas solvation is the second ranked component regarding reaction rate. Such predictions are nonintuitive. While previous work found that parametric-based microscale uncertainties can be dampened in multiscale models (*43*), the results of this work will generalize to any models where fine-scale simulations (such as DFT) are sparse or the macroscale QoIs can be made proportional to the microscale properties. In the next section, we show that Eq. 6 and the resulting Fig. 4 can also be deployed to improve the baseline (purely Gaussian) model P.

Model correctability enabled by model UQ

Model uncertainty due to any submodel or dataset, quantified by η_l and Eq. 6, can be reduced by picking a better submodel or dataset than the original baseline model P_l. Obviously, those CPDs that exhibit larger relative predictive uncertainty in Eq. 6 should be prioritized and corrected. In our case study, reducing the DFT error requires to further develop DFT functional and methods, a long-standing pursue not addressed here. Here, we illustrate how to carry out such

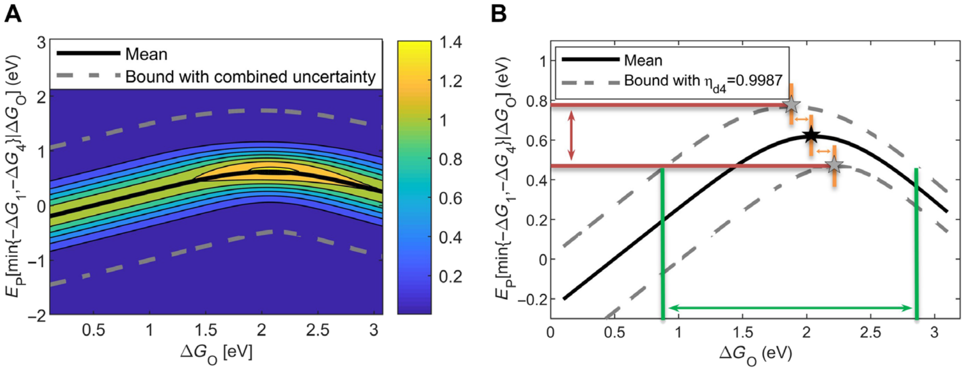

Fig. 3. Parametric and model uncertainty. (**A**) Parametric versus model uncertainty: Contour plot of the probability distribution of $\min\{-\Delta G_1, -\Delta G_4\}$ as a function of ΔG_O; the black curve is the mean (expected) value $\mathbb{E}_P[\min\{-\Delta G_1, -\Delta G_4\}|\Delta G_O]$ for the baseline ORR PGM P in Fig. 2. The gray dashed lines are the extreme bounds (guarantees) with combined model uncertainty, and the color indicates likelihood; see section S5 for more details. (**B**) Model uncertainty guarantees given by the predictive uncertainty (model sensitivity indices) (gray dashed lines) for the QoI $\min\{-\Delta G_1, -\Delta G_4\} \mid \Delta G_O$ when only the uncertainty of DFT in ΔG_4 is considered.

model correctability through an example that is feasible to do. Specifically, we consider the model consisting of the data used to construct the volcano and its statistical representation as this is the second most influential parameter in materials prediction. We performed additional DFT calculations on core-shell bimetallics to create an expanded dataset compared to that in Fig. 1C (see Fig. 5A). By doing this, we compute the model sensitivity indices $\mathcal{J}_l^{\pm}$ for the new model using theorem 1 and equation S58. More details and derivations are included in Methods.

Figure 5B shows the reduction of model uncertainty guarantees, defined as Eq. 6, which are due to the variance of error and the estimated model misspecification parameter in the correlation between DFT-calculated values of ΔG_4 and ΔG_O, when more data (bimetallics) are added. With bimetallic data included, the correlation coefficients of ΔG_O with $-\Delta G_1$ and $-\Delta G_4$ are −0.95 and 0.95, respectively. The uncertainty is reduced both due to improved correlation and reduced SE in the regression coefficients as a result of more data.

DISCUSSION

Here, we introduce PGM for chemistry to embed uncertainty into the design of a model. The approach provides a blueprint to systematically integrate desperate components of a model ranging from heterogeneous data (experimental and DFT) to expert knowledge to physics/chemistry models and constraints to correlations among data and causality between variables. Instead of a deterministic model, a probabilistic ensemble of models is created. Furthermore, the model uncertainty and sensitivity indices derived herein provide guarantees on model prediction to systematically identify the most influential model components causing predictive uncertainty and ultimately ensure trustworthiness of predictions. Overall, our proposed mathematical framework combines probabilistic AI and UQ to provide explainable results, leading to correctable and, eventually, trustworthy models. We illustrate this framework for a volcano-kinetic model for the ORR. We propagate both parametric and model uncertainty from several, small sets of input data to model

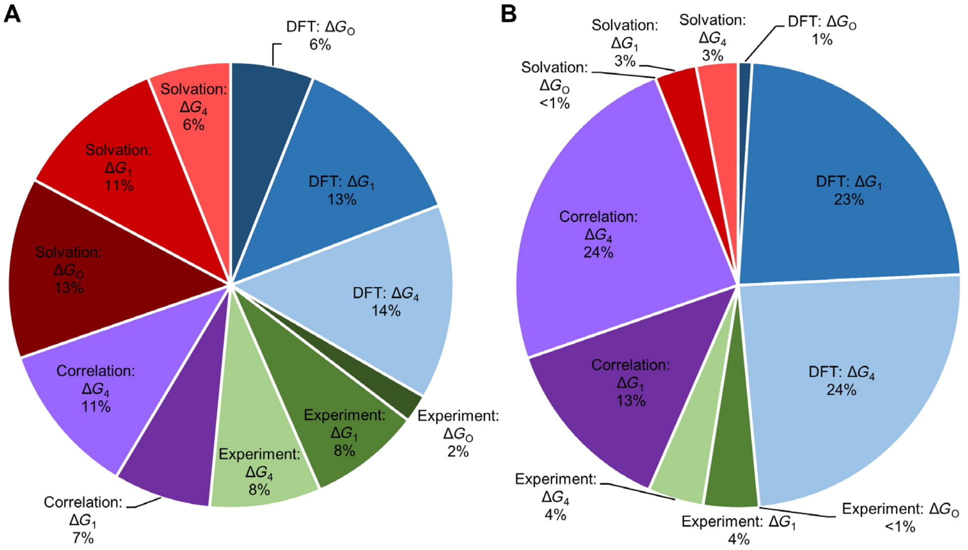

Fig. 4. Ranking indices for optimal rate and optimal oxygen binding energy in each ORR PGM submodel. Rankings for the model uncertainties in $k_\mathrm{B}T \ln(r^*)$ (**A**) and ΔG_O^* variability (**B**). See section S7 for more details.

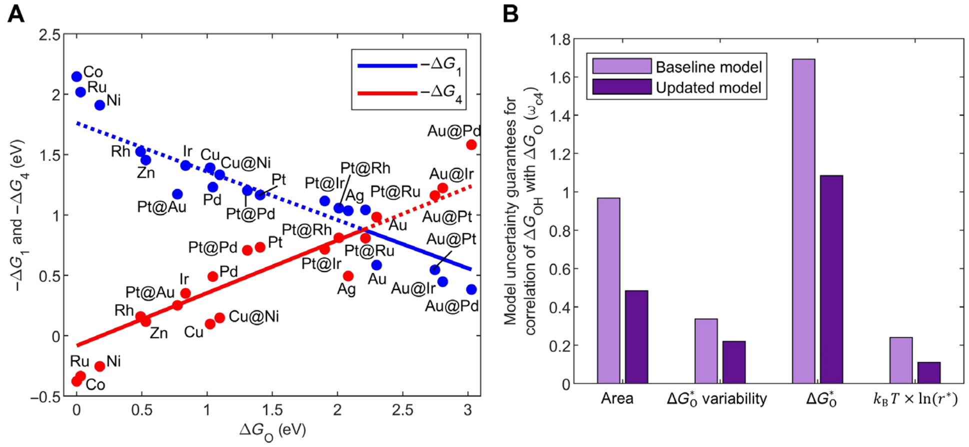

Fig. 5. Correctability of the submodel determining the volcano using DFT data. (**A**) Volcano curve with additional bimetallic data where M1@M2 indicates a shell of metal 1 on metal 2. (**B**) Uncertainty bounds when accounting for correlation error of the area of uncertainty region, ΔG_O^* variability, ΔG_O^*, and $k_\mathrm{B}T \ln(r^*)$ for both the baseline model (light purple bars) and model with bimetallic data included (dark purple bars).

predictions. We establish model error bounds on the ORR volcano to assess maximum and minimum rates given the binding energy of atomic oxygen as the primary descriptor. We assess the impact of these errors via model sensitivity indices, which quantify the percent error from uncertainty contributed by each variable to the predicted maximum ORR rate and the oxygen binding energy corresponding to that rate. The greatest contribution to errors (ordered from greatest to least) in the PGM-based ORR volcano in predicting materials arises from error in DFT calculations and correlations of OOH* and OH* binding energies with O* binding energies. Different from the materials, the reaction rates depend mainly on errors associated with DFT and solvation, yet experimental error and correlations are relatively large as well, i.e., a more equidistribution of error is observed and improved accuracy of all components is needed to size electrochemical devices. Improving the accuracy of DFT method and the quality and quantity of data can pave the way for more accurate models for finding new catalysts.

METHODS

DFT calculations

We study adsorption on the close-packed (111 and 0001) transition metal surfaces. We select the lowest energy site of O* and OH* for comparison with experiments to determine errors, which are summarized in table S1. We build the correlations for bimetallics from the lowest energy sites on the (111) and (0001) surfaces of the face-centered (fcc) and hexagonal close-packed (hcp) metals, respectively.

Vacuum phase DFT setup

We calculated binding energies and vibrational frequencies using the Vienna ab initio Simulation Package version 5.4 with the projector-augmented wave method (*44*). We use the Revised Perdew-Burke-Ernzerhof (RPBE) density functional (*45*) with D3 dispersion corrections (*46*). Simulation methods include use of spin-polarized calculations for gas-phase species and ferromagnetic metals, a 3 × 3 × 1 Monkhorst-Pack *k*-point sampling grid (*47*) for all slab calculations, and a 400-eV plane wave cutoff. Electronic energy convergence was set to 10^{-4} eV for the energy minimization step and 10^{-6} eV for frequency calculations.

For calculations of gas-phase species, the supercell size was 10 × 10 × 10 Å. A Brillouin zone was sampled at the gamma point; a 0.005 eV/Å force cutoff was used in geometry optimizations. For slab calculations, the force cutoff was set to 0.02 eV/Å with 20 Å of vacuum space. Adsorbate energies were calculated for OOH*, OH*, and O* on the most stable close-packed surface for fcc and hcp metals. The periodic cell consisted of four layers with 16 metal atoms in each layer; the bottom two layers were fixed at their bulk values, determined using a 15 × 15 × 15 *k*-point grid with the tetrahedron method and Blöchl corrections. Bulk metal lattice constants were pre-optimized with DFT using the Birch-Murnaghan equation of state (*48*). Zero-point energies are calculated for each adsorbate-surface combination and for all gas species. All input files were created using the Atomic Simulation Environment (*49*).

Solvation phase DFT setup

We emulate explicit solvation calculations from previous work (*38*) except that, here, we vary the number of water layers. Two to five layers of water were placed above a Pt(111) surface in a honeycomb pattern to simulate the aqueous phase above the surface. The double layer of water molecules was found to adequately capture water binding energies on Pt(111) and H bonds at the surface (*50*). We determined solvation energies for O* by placing it in an fcc hollow site on the water-covered surface. For OH* and OOH*, solvation energies were determined by replacing a water molecule on the surface with the respective species to determine solvation energies. Other than the choice of functional, the DFT setup was identical to that in the vacuum except that nine Pt atoms were included in each layer to accommodate the honeycomb water structure, the *k*-point sampling was increased to 4 × 4 × 1, and the plane wave cutoff was increased to 450 eV. To provide initial geometries, the Perdew-Burke-Ernzerhof (PBE) functional (*51*) was used for all solvation calculations. Solvation energy calculations on Pt(111) using the PBE functional do not cause inconsistencies with the use of the RPBE functional for vacuum phase calculations. Granda-Marulanda *et al.* (*52*) showed that on several 111 and 0001 surfaces, the average difference in OH* solvation using the PBE and RPBE functionals with dispersion corrections was 0.03 eV; the SDs using these functionals were similar at 0.08 and 0.11 eV, respectively. Rather than changes in solvation across different surfaces, we investigate the variance in solvation energy associated with the number of explicit water layers used. Because energies from PBE and RPBE are correlated, the variance in solvation energy with respect to number of water layers is expected to be similar.

Temperature effects

Temperature effects at 298 K were calculated using statistical thermodynamics in combination with the harmonic and ideal gas approximations (*53*). Both heat capacity and entropy effects were included in calculating Gibbs free energies used in the volcano curves. Entropy was removed when comparing to experimental enthalpies as discussed in section S3.

Deriving model sensitivity indices

Using robust and scalable UQ methods for general probabilistic models (*27*, *28*, *54*) as a starting point, we define "ambiguity sets" around a baseline model P and "predictive uncertainty for QoIs." Although the definitions of predictive uncertainty (section S6) and model sensitivity indices (Eq. 6) are natural and rather intuitive, it is not obvious that they are practically computable. A key mathematical finding for PGMs, demonstrated in theorem S1, is that that the guarantees $\mathcal{J}_l^{\pm}(\mathrm{QoI}, P; \eta_l)$ can be computed exactly using a variational formula for the KL divergence and the chain rule for the KL divergence; the latter point also justifies the use of the KL divergence in defining the nonparametric formulation of the model sensitivity indices. In the case where P is a Gaussian Bayesian network (G), the ranking indices in Eq. 7 are computed using Eq. 9.

Selecting new high-quality data or improved physical model for model correctability

Given a baseline model P and the sparse dataset for each submodel sampled from an unknown model Q, we can build an improved baseline model P for our ORR model following the steps below.

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

$$\eta_l = R\big(Q(X_l|Pa_{x_l}) \,\|\, P(X_l|Pa_{x_l})\big) \tag{8}$$

where Q is the surrogate model given by the KDE/histogram, using eqs. S94 and S95.

Step 2: Calculate the model uncertainty guarantees for a given QoI using eq. S58 (or eq. S60 for the general PGM)

$$\mathcal{J}_l^{\pm}(\mathrm{QoI}, P; \eta_l) \text{ for all PGM vertices } l$$

Step 3: Select the l^* component X_{l^*} of the PGM with the worst guarantees $\mathcal{J}^{+}_{l^*}(\text{QoI}, P; \eta_{l^*})$ (highest values).

Step 4: Reduce $\mathcal{J}^{+}_{l^*}(\text{QoI}, P; \eta_{l^*})$ based on eq. S58. For $\text{QoI}(X) = \min\{-\Delta G_1, -\Delta G_4\} \mid \Delta G_\text{O}$, we have that

$$\mathcal{J}^{\pm}_{l^*}(\text{QoI}, P; \eta_{l^*}) = \pm\inf_{c>0}\left[\frac{1}{c}\log\int e^{\pm c\overline{F_{l^*}}}\, P_{l^*}(d\omega_{l^*}) + \frac{\eta_{l^*}}{c}\right] \qquad (9)$$

where $\overline{F_{l^*}}(\omega_{l^*}) = \mathbb{E}_{P_{\{\omega_{l^*}\}^c}}[\text{QoI}] - \mathbb{E}_P[\text{QoI}]$. For the l^* component(s) of the PGM, we seek the most useful additional data, namely the data that tightens (reduces) the guarantees in Eq. 9. Note that the guarantees consist of two parts: the moment generating function (MGF) and the model misspecification parameter η. Therefore, adding informative data can reduce the MGF in Eq. 9 (and, thus, the uncertainty guarantees $\mathcal{J}^{\pm}_{l^*}(\text{QoI}, P; \eta_{l^*})$); since the MGF includes all moments, and, in particular, the variance (*27*), additional data can improve model P and reduce the model misspecification η (see section S8).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/42/eabc3204/DC1

REFERENCES AND NOTES

1. A. Hospital, J. R. Goñi, M. Orozco, J. L. Gelpí, Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinforma. Chem.* **8**, 37–47 (2015).
2. J. Sehested, K. E. Larsen, A. L. Kustov, A. M. Frey, T. Johannessen, T. Bligaard, M. P. Andersson, J. K. Nørskov, C. H. Christensen, Discovery of technical methanation catalysts based on computational screening. *Top. Catal.* **45**, 9–13 (2007).
3. D. A. Hansgen, D. G. Vlachos, J. G. Chen, Using first principles to predict bimetallic catalysts for the ammonia decomposition reaction. *Nat. Chem.* **2**, 484–489 (2010).
4. J. A. Pople, Nobel lecture: Quantum chemical models. *Rev. Mod. Phys.* **71**, 1267–1274 (1999).
5. S. Skogestad, I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design* (Wiley New York, 2007), vol. 2.
6. F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. R. Munter, P. G. Moses, E. Skúlason, T. Bligaard, J. K. Nørskov, Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. *Phys. Rev. Lett.* **99**, 016105 (2007).
7. F. Calle-Vallejo, J. Tymoczko, V. Colic, Q. H. Vu, M. D. Pohl, K. Morgenstern, D. Loffreda, P. Sautet, W. Schuhmann, A. S. Bandarenka, Finding optimal surface sites on heterogeneous catalysts by counting nearest neighbors. *Science* **350**, 185–189 (2015).
8. J. L. Lansford, A. V. Mironenko, D. G. Vlachos, Scaling relationships and theory for vibrational frequencies of adsorbates on transition metal surfaces. *Nat. Commun.* **8**, 1842 (2017).
9. M. Salciccioli, M. Stamatakis, S. Caratzoulas, D. G. Vlachos, A review of multiscale modeling of metal-catalyzed reactions: Mechanism development for complexity and emergent behavior. *Chem. Eng. Sci.* **66**, 4319–4355 (2011).
10. A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, Sensitivity analysis for chemical models. *Chem. Rev.* **105**, 2811–2828 (2005).
11. H. Rabitz, M. Kramer, D. Dacol, Sensitivity analysis in chemical kinetics. *Annu. Rev. Phys. Chem.* **34**, 419–461 (1983).
12. J. E. Sutton, W. Guo, M. A. Katsoulakis, D. G. Vlachos, Effects of correlated parameters and uncertainty in electronic-structure-based chemical kinetic modelling. *Nat. Chem.* **8**, 331–337 (2016).
13. J. Feng, J. Lansford, A. Mironenko, D. B. Pourkargar, D. G. Vlachos, M. A. Katsoulakis, Non-parametric correlative uncertainty quantification and sensitivity analysis: Application to a Langmuir bimolecular adsorption model. *AIP Adv.* **8**, 035021 (2018).
14. S. Gautier, S. N. Steinmann, C. Michel, P. Fleurat-Lessard, P. Sautet, Molecular adsorption at Pt(111). How accurate are DFT functionals? *Phys. Chem. Chem. Phys.* **17**, 28921–28930 (2015).
15. J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, K. W. Jacobsen, Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Phys. Rev. B* **85**, 235149 (2012).
16. J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K. Nørskov, T. Bligaard, F. Studt, C. T. Campbell, A benchmark database for adsorption bond energies to transition metal surfaces and comparison to selected DFT functionals. *Surf. Sci.* **640**, 36–44 (2015).
17. C. C. Aggarwal, in *Managing and Mining Uncertain Data*, C. C. Aggarwal, Ed. (Springer US, Boston, MA, 2009), pp. 1–36.
18. J. A. Vrugt, C. J. F. ter Braak, M. P. Clark, J. M. Hyman, B. A. Robinson, Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.* **44**, W00B09 (2008).
19. J. Freer, K. Beven, B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resour. Res.* **32**, 2161–2173 (1996).
20. N. Friedman, M. Linial, I. Nachman, D. Pe'er, Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**, 601–620 (2000).
21. I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55**, 271–280 (2001).
22. A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **181**, 259–270 (2010).
23. L. Landau, E. Lifshitz, in *Perspectives in Theoretical Physics*, L. P. Pitaevski, Ed. (Pergamon, 1992), pp. 287–297.
24. S. Taverniers, F. J. Alexander, D. M. Tartakovsky, Noise propagation in hybrid models of nonlinear systems: The Ginzburg–Landau equation. *J. Comput. Phys.* **262**, 313–324 (2014).
25. D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques* (MIT press, 2009).
26. Z. Ghahramani, Probabilistic machine learning and artificial intelligence. *Nature* **521**, 452–459 (2015).
27. P. Dupuis, M. A. Katsoulakis, Y. Pantazis, P. Plechác, Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM-ASA J. Uncertain.* **4**, 80–111 (2016).
28. M. A. Katsoulakis, L. Rey-Bellet, J. Wang, Scalable information inequalities for uncertainty quantification. *J. Comput. Phys.* **336**, 513–545 (2017).
29. J. E. Sutton, D. G. Vlachos, Effect of errors in linear scaling relations and Brønsted–Evans–Polanyi relations on activity and selectivity maps. *J. Catal.* **338**, 273–283 (2016).
30. S. Deshpande, J. R. Kitchin, V. Viswanathan, Quantifying uncertainty in activity volcano relationships for oxygen reduction reaction. *ACS Catal.* **6**, 5251–5259 (2016).
31. D. R. Palo, J. D. Holladay, R. T. Rozmiarek, C. E. Guzman-Leong, Y. Wang, J. Hu, Y.-H. Chin, R. A. Dagle, E. G. Baker, Development of a soldier-portable fuel cell power system: Part I: A bread-board methanol fuel processor. *J. Power Sources* **108**, 28–34 (2002).
32. H. A. Gasteiger, N. M. Marković, Just a dream—Or future reality? *Science* **324**, 48–49 (2009).
33. W. Sheng, H. A. Gasteiger, Y. Shao-Horn, Hydrogen oxidation and evolution reaction kinetics on platinum: Acid vs alkaline electrolytes. *J. Electrochem. Soc.* **157**, B1529 (2010).
34. J. Durst, A. Siebel, C. Simon, F. Hasché, J. Herranz, H. A. Gasteiger, New insights into the electrochemical hydrogen oxidation and evolution reaction mechanism. *Energy Environ. Sci.* **7**, 2255–2260 (2014).
35. N.-T. Suen, S.-F. Hung, Q. Quan, N. Zhang, Y.-J. Xu, H. M. Chen, Electrocatalysis for the oxygen evolution reaction: Recent development and future perspectives. *Chem. Soc. Rev.* **46**, 337–365 (2017).
36. O. Antoine, Y. Bultel, R. Durand, Oxygen reduction reaction kinetics and mechanism on platinum nanoparticles inside Nafion®. *J. Electroanal. Chem.* **499**, 85–94 (2001).
37. A. Holewinski, S. Linic, Elementary mechanisms in electrocatalysis: Revisiting the ORR tafel slope. *J. Electrochem. Soc.* **159**, H864–H870 (2012).
38. M. Núñez, J. L. Lansford, D. G. Vlachos, Optimization of the facet structure of transition-metal catalysts applied to the oxygen reduction reaction. *Nat. Chem.* **11**, 449–456 (2019).
39. D. M. Chickering, in *Learning from Data: Artificial Intelligence and Statistics V*, D. Fisher, H.-J. Lenz, Eds. (Springer New York, 1996), pp. 121–130.
40. G. F. Cooper, The computational complexity of probabilistic inference using bayesian belief networks. *Artif. Intell.* **42**, 393–405 (1990).
41. P. Spirtes et al., *Causation, Prediction, and Search* (MIT press, 2000).
42. T. Hofmann, B. Schölkopf, A. J. Smola, Kernel methods in machine learning. *Ann. Stat.* **36**, 1171–1220 (2008).
43. K. Um, E. J. Hall, M. A. Katsoulakis, D. M. Tartakovsky, Causality and Bayesian Network PDEs for multiscale representations of porous media. *J. Comput. Phys.* **394**, 658–678 (2019).
44. G. Kresse, J. Furthmüller, Efficient iterative schemes forab initiototal-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
45. B. Hammer, L. B. Hansen, J. K. Norskov, Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B* **59**, 7413–7421 (1999).
46. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
47. H. J. Monkhorst, J. D. Pack, Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
48. F. D. Murnaghan, The compressibility of media under extreme pressures. *Proc. Natl. Acad. Sci. U.S.A.* **30**, 244–247 (1944).
49. S. R. Bahn, K. W. Jacobsen, An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66 (2002).
50. S. Meng, E. G. Wang, S. Gao, Water adsorption on metal surfaces: A general picture from density functional theory studies. *Phys. Rev. B* **69**, 195404 (2004).

51. J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
52. L. P. Granda-Marulanda, S. Builes, M. T. M. Koper, F. Calle-Vallejo, Influence of Van der Waals interactions on the solvation energies of adsorbates at Pt-based electrocatalysts. *ChemPhysChem* **20**, 2968–2972 (2019).
53. D. A. McQuarrie, *Statistical Mechanics* (University Science Books, 2000).
54. K. Gourgoulias, M. A. Katsoulakis, L. Rey-Bellet, J. Wang, How biased is your model? Concentration inequalities, information and model bias. *IEEE Trans. Inf. Theory* , 1–1 (2020).
55. S. Choi, C. J. Kucharczyk, Y. Liang, X. Zhang, I. Takeuchi, H.-I. Ji, S. M. Haile, Exceptional power density and stability at intermediate temperatures in protonic ceramic fuel cells. *Nat. Energy* **3**, 202–210 (2018).
56. C. K. Dyer, Fuel cells for portable applications. *J. Power Sources* **106**, 31–34 (2002).
57. G. Sievi, D. Geburtig, T. Skeledzic, A. Bösmann, P. Preuster, O. Brummel, F. Waidhas, M. A. Montero, P. Khanipour, I. Katsounaros, J. Libuda, K. J. J. Mayrhofer, P. Wasserscheid, Towards an efficient liquid organic hydrogen carrier fuel cell concept. *Energy Environ. Sci.* **12**, 2305–2314 (2019).
58. A. Mahajan, S. Banik, D. Majumdar, S. K. Bhattacharya, Anodic oxidation of butan-1-ol on reduced graphene oxide-supported Pd–Ag nanoalloy for fuel cell application. *ACS Omega* **4**, 4658–4670 (2019).
59. V. K. Puthiyapura, D. J. L. Brett, A. E. Russell, W.-F. Lin, C. Hardacre, Biobutanol as fuel for direct alcohol fuel cells—Investigation of Sn-modified Pt catalyst for butanol electro-oxidation. *ACS Appl. Mater. Interfaces* **8**, 12859–12870 (2016).
60. X. Ren, P. Zelenay, S. Thomas, J. Davey, S. Gottesfeld, Recent advances in direct methanol fuel cells at Los Alamos National Laboratory. *J. Power Sources* **86**, 111–116 (2000).
61. B. G. Pollet, I. Staffell, J. L. Shang, Current status of hybrid, battery and fuel cell electric vehicles: From electrochemistry to market prospects. *Electrochim. Acta* **84**, 235–249 (2012).
62. F. Calle-Vallejo, M. T. M. Koper, First-principles computational electrochemistry: Achievements and challenges. *Electrochim. Acta* **84**, 3–11 (2012).
63. A. Kulkarni, S. Siahrostami, A. Patel, J. K. Nørskov, Understanding catalytic activity trends in the oxygen reduction reaction. *Chem. Rev.* **118**, 2302–2312 (2018).
64. C. H. Choi, S. H. Park, S. I. Woo, Phosphorus-nitrogen dual doped carbon as an effective catalyst for oxygen reduction reaction in acidic media: Effects of the amount of P-doping on the physical and electrochemical properties of carbon. *J. Mater. Chem.* **22**, 12107–12115 (2012).
65. W. Sheng, M. Myint, J. G. Chen, Y. Yan, Correlating the hydrogen evolution reaction activity in alkaline electrolytes with the hydrogen binding energy on monometallic surfaces. *Energy Environ. Sci.* **6**, 1509–1512 (2013).
66. A. Damjanovic, A. Dey, J. O. M. Bockris, Kinetics of oxygen evolution and dissolution on platinum electrodes. *Electrochim. Acta* **11**, 791–814 (1966).
67. R. Jinnouchi, K. Kodama, T. Hatanaka, Y. Morimoto, First principles based mean field model for oxygen reduction reaction. *Phys. Chem. Chem. Phys.* **13**, 21070–21083 (2011).
68. M. P. Hyman, J. W. Medlin, Mechanistic study of the electrochemical oxygen reduction reaction on Pt(111) using density functional theory. *J. Phys. Chem. B* **110**, 15338–15344 (2006).
69. J. M. Seminario, L. A. Agapito, L. Yan, P. B. Balbuena, Density functional theory study of adsorption of OOH on Pt-based bimetallic clusters alloyed with Cr, Co, and Ni. *Chem. Phys. Lett.* **410**, 275–281 (2005).
70. V. Tripković, E. Skúlason, S. Siahrostami, J. K. Nørskov, J. Rossmeisl, The oxygen reduction reaction mechanism on Pt(111) from density functional theory calculations. *Electrochim. Acta* **55**, 7975–7981 (2010).
71. I. Chorkendorff, J. W. Niemantsverdriet, *Concepts of Modern Catalysis and Kinetics* (John Wiley & Sons, 2017).
72. F. Calle-Vallejo, A. Krabbe, J. M. García-Lastra, How covalence breaks adsorption-energy scaling relations and solvation restores them. *Chem. Sci.* **8**, 124–130 (2017).
73. J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, H. Jónsson, Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B* **108**, 17886–17892 (2004).
74. J. Rossmeisl, Z.-W. Qu, H. Zhu, G.-J. Kroes, J. K. Nørskov, Electrolysis of water on oxide surfaces. *J. Electroanal. Chem.* **607**, 83–89 (2007).
75. Z.-D. He, S. Hanselman, Y.-X. Chen, M. T. M. Koper, F. Calle-Vallejo, Importance of solvation for the accurate prediction of oxygen reduction activities of Pt-based electrocatalysts. *J. Phys. Chem. Lett.* **8**, 2243–2246 (2017).
76. G. N. Derry, P. N. Ross, A work function change study of oxygen adsorption on Pt(111) and Pt(100). *J. Chem. Phys.* **82**, 2772–2778 (1985).
77. V. P. Ivanov, G. K. Boreskov, V. I. Savchenko, W. F. Egelhoff Jr., W. H. Weinberg, The chemisorption of oxygen on the iridium (111) surface. *Surf. Sci.* **61**, 207–220 (1976).
78. T. W. Root, L. D. Schmidt, G. B. Fisher, Adsorption and reaction of nitric oxide and oxygen on Rh(111). *Surf. Sci.* **134**, 30–45 (1983).
79. J. T. Stuckless, C. E. Wartnaby, N. Al-Sarraf, S. J. B. Dixon-Warren, M. Kovar, D. A. King, Oxygen chemisorption and oxide film growth on Ni{100}, {110}, and {111}: Sticking probabilities and microcalorimetric adsorption heats. *J. Chem. Phys.* **106**, 2012–2030 (1997).
80. C. E. Wartnaby, A. Stuck, Y. Y. Yeo, D. A. King, Microcalorimetric heats of adsorption for CO, NO, and oxygen on Pt{110}. *J. Phys. Chem.* **100**, 12483–12488 (1996).
81. Y. Y. Yeo, L. Vattuone, D. A. King, Calorimetric heats for CO and oxygen adsorption and for the catalytic CO oxidation reaction on Pt{111}. *J. Chem. Phys.* **106**, 392–401 (1997).
82. C. T. Campbell, G. Ertl, H. Kuipers, J. Segner, A molecular beam study of the adsorption and desorption of oxygen from a Pt(111) surface. *Surf. Sci.* **107**, 220–236 (1981).
83. D. H. Parker, M. E. Bartram, B. E. Koel, Study of high coverages of atomic oxygen on the Pt(111) surface. *Surf. Sci.* **217**, 489–510 (1989).
84. J. F. Weaver, J.-J. Chen, A. L. Gerrard, Oxidation of Pt(111) by gas-phase oxygen atoms. *Surf. Sci.* **592**, 83–103 (2005).
85. A. B. Anton, D. C. Cadogan, The mechanism and kinetics of water formation on Pt(111). *Surf. Sci.* **239**, L548–L560 (1990).
86. V. Climent, R. Gómez, J. M. Orts, J. M. Feliu, Thermodynamic analysis of the temperature dependence of OH adsorption on Pt(111) and Pt(100) electrodes in acidic media in the absence of specific anion adsorption. *J. Phys. Chem. B* **110**, 11344–11351 (2006).
87. E. M. Karp, C. T. Campbell, F. Studt, F. Abild-Pedersen, J. K. Nørskov, Energetics of oxygen adatoms, hydroxyl species and water dissociation on Pt(111). *J. Phys. Chem. C* **116**, 25772–25776 (2012).
88. W. Lew, M. C. Crowe, E. Karp, O. Lytken, J. A. Farmer, L. Árnadóttir, C. Schoenbaum, C. T. Campbell, The energy of adsorbed hydroxyl on Pt(111) by microcalorimetry. *J. Phys. Chem. C* **115**, 11586–11594 (2011).
89. B. deB. Darwent, *Bond Dissociation Energies in Simple Molecules* (U.S. National Bureau of Standards Reference Data, 1970), vol. 31.
90. C. T. Campbell, Atomic and molecular oxygen adsorption on Ag(111). *Surf. Sci.* **157**, 43–60 (1985).
91. E. Giamello, B. Fubini, P. Lauro, A. Bossi, A microcalorimetric method for the evaluation of copper surface area in Cu-ZnO catalyst. *J. Catal.* **87**, 443–451 (1984).
92. X. Guo, A. Hoffman, J. T. Yates Jr., Adsorption kinetics and isotopic equilibration of oxygen adsorbed on the Pd(111) surface. *J. Chem. Phys.* **90**, 5787–5792 (1989).
93. D. A. King, T. E. Madey, J. T. Yates Jr., Interaction of oxygen with polycrystalline tungsten. I. Sticking probabilities and desorption spectra. *J. Chem. Phys.* **55**, 3236–3246 (1971).
94. R. Kose, W. A. Brown, D. A. King, Calorimetric heats of dissociative adsorption for O_2 on Rh{100}. *Surf. Sci.* **402-404**, 856–860 (1998).
95. T. E. Madey, H. Albert Engelhardt, D. Menzel, Adsorption of oxygen and oxidation of CO on the ruthenium (001) surface. *Surf. Sci.* **48**, 304–328 (1975).
96. N. Saliba, D. H. Parker, B. E. Koel, Adsorption of oxygen on Au(111) by exposure to ozone. *Surf. Sci.* **410**, 270–282 (1998).
97. E. M. Stuve, S. W. Jorgensen, R. J. Madix, The adsorption of H_2O on clean and oxygen-covered Pd(100): Formation and reaction of OH groups. *Surf. Sci.* **146**, 179–198 (1984).
98. F. J. Massey, The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
99. K. Bange, T. E. Madey, J. K. Sass, E. M. Stuve, The adsorption of water and oxygen on Ag(110): A study of the interactions among water molecules, hydroxyl groups, and oxygen atoms. *Surf. Sci.* **183**, 334–362 (1987).
100. W. Zhao, S. J. Carey, Z. Mao, C. T. Campbell, Adsorbed hydroxyl and water on Ni(111): Heats of formation by calorimetry. *ACS Catal.* **8**, 1485–1489 (2018).
101. M. D. Liptak, G. C. Shields, Accurate pK_a calculations for carboxylic acids using complete basis set and gaussian-n models combined with CPCM continuum solvation methods. *J. Am. Chem. Soc.* **123**, 7314–7319 (2001).
102. D. Aronsky, P. J. Haug, Diagnosing community-acquired pneumonia with a Bayesian network. *Proc. AMIA Symp.* , 632–636 (1998).
103. R. Fung, B. Del Favero, Applying Bayesian networks to information retrieval. *Commun. ACM* **38**, 3 (1995).
104. D. Heckerman, E. J. Horvitz, B. N. Nathwani, Toward normative expert systems: Part I. The pathfinder project. *Methods Inf. Med.* **31**, 90–105 (1992).
105. E. Lazkano, B. Sierra, A. Astigarraga, J. M. Martinez-Otzeta, On the use of Bayesian networks to develop behaviours for mobile robots. *Robot. Auton. Syst.* **55**, 253–265 (2007).
106. T. S. Levitt, J. M. Agosta, T. O. Binford, in *Machine Intelligence and Pattern Recognition* (Elsevier, 1990), vol. 10, pp. 371–388.
107. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Reasoning* (Morgan Kaufmann Publishers, 1988).
108. L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference* (Springer Science & Business Media, 2013).
109. K. Zhang, J. Peters, D. Janzing, B. Schölkopf, Kernel-based conditional independence test and application in causal discovery. arXiv:1202.3775 (2012).
110. C. Heinze-Deml, J. Peters, N. Meinshausen, Invariant causal prediction for nonlinear models. *J. Causal Inference* **6**, 20170016 (2018).
111. S. Conrady, L. Jouffe, Introduction to Bayesian networks & bayesialab (2013).
112. J. Bromley, N. A. Jackson, O. J. Clymer, A. M. Giacomello, F. V. Jensen, The use of Hugin®to develop Bayesian networks as an aid to integrated water resource planning. *Environ. Model. Softw.* **20**, 231–242 (2005).

113. A. L. Madsen, M. Lang, U. B. Kjærulff, F. Jensen, in *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty* (Springer, 2003), pp. 594–605.
114. O. Woodberry, S. Mascaro, *Programming Bayesian Network Solutions with Netica* (Bayesian Intelligence, 2012).
115. D. Haughton, A. Kamis, P. Scholten, A review of three directed acyclic graphs software packages: MIM, Tetrad, and WinMine. *Am. Stat.* **60**, 272–286 (2006).
116. R. Scheines, P. Spirtes, C. Glymour, C. Meek, T. Richardson, *TETRAD 3: Tools for Causal Modeling—User's Manual* (CMU Philosophy, 1996).
117. S. Nadarajah, S. Kotz, Exact distribution of the max/min of two Gaussian random variables. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **16**, 210–212 (2008).
118. P. Dupuis, R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations* (John Wiley & Sons, 2011), vol. 902.
119. Z. Hu, L. J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization (2013), Available at Optimization Online.
120. P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **171**, 115–166 (2018).
121. E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58**, 595–612 (2010).
122. J. Goh, M. Sim, Distributionally robust optimization and its tractable approximations. *Oper. Res.* **58**, 902–917 (2010).
123. W. Wiesemann, D. Kuhn, M. Sim, Distributionally robust convex optimization. *Oper. Res.* **62**, 1358–1376 (2014).
124. R. Jiang, Y. Guan, Data-driven chance constrained stochastic program. *Math. Program.* **158**, 291–327 (2016).
125. R. Gao, A. J. Kleywegt, Distributionally robust stochastic optimization with Wasserstein distance. arXiv:1604.02199 (2016).
126. H. Lam, Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Oper. Res.* **67**, 1090–1105 (2019).
127. W. Xie, S. Ahmed, On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM J. Optim.* **28**, 1151–1182 (2018).
128. J. Blanchet, K. Murthy, Quantifying distributional model risk via optimal transport. *Math. Oper. Res.* **44**, 565–600 (2019).
129. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
130. J. Blanchet, H. Lam, Q. Tang, Z. Yuan, Robust actuarial risk analysis. *N. Am. Actuar. J.* **23**, 33–63 (2019).
131. D. Bertsimas, V. Gupta, N. Kallus, Robust sample average approximation. *Math. Program.* **171**, 217–282 (2018).
132. Z. Wang, P. W. Glynn, Y. Ye, Likelihood robust optimization for data-driven problems. *Comput. Manag. Sci.* **13**, 241–261 (2016).
133. J. Blanchet, Y. Kang, Sample out-of-sample inference based on Wasserstein distance. arXiv:1605.01340 (2016).
134. H. Lam, E. Zhou, The empirical likelihood approach to quantifying uncertainty in sample average approximation. *Oper. Res. Lett.* **45**, 301–307 (2017).
135. J. Duchi, P. Glynn, H. Namkoong, Statistics of robust optimization: A generalized empirical likelihood approach. arXiv:1610.03425 (2016).
136. J.-y. Gotoh, M. J. Kim, A. E. B. Lim, Robust empirical optimization is almost the same as mean–variance optimization. *Oper. Res. Lett.* **46**, 448–452 (2018).
137. L. Wasserman, *All of Nonparametric Statistics* (Springer, 2006).
138. X. Nguyen, M. J. Wainwright, M. I. Jordan, Estimating divergence functionals and the likelihood ratio by convex risk minimization. *IEEE Trans. Inf. Theory* **56**, 5847–5861 (2010).
139. A. Krishnamurthy, K. Kandasamy, B. Poczos, L. Wasserman, Nonparametric estimation of renyi divergence and friends, in *International Conference on Machine Learning* (ICML 2014), Beijing, China, 21 to 26 June 2014, pp. 919–927.
140. K. Kandasamy, A. Krishnamurthy, B. Poczos, L. Wasserman, Nonparametric von mises estimators for entropies, divergences and mutual informations, in *Advances in Neural Information Processing Systems* (NIPS 2015), pp. 397–405.
141. M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, D. Hjelm, Mutual Information Neural Estimation, paper presented at the Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research (PMLR 2018), Stockholmsmässan, Stockholm, Sweden, 10 to 15 July 2018.

Acknowledgements: J.L.L. and D.G.V. acknowledge valuable discussions with J. Lym, K. Alexopoulos, and G. Wittreich. **Funding:** The research of M.A.K. was partially supported by NSF TRIPODS CISE-1934846 and the Air Force Office of Scientific Research (AFOSR) under the grant FA-9550-18-1-0214. The research of J.F. was partially supported by the Defense Advanced Research Projects Agency (DARPA) EQUiPS program under the grant W911NF1520122, and part of J.F.'s work was completed during his PhD at UMass Amherst. J.L.L. and D.G.V. acknowledge support by the RAPID Manufacturing Institute, supported by the Department of Energy (DOE) Advanced Manufacturing Office (AMO), award number DE-EE0007888-9.5. RAPID projects at the University of Delaware are also made possible, in part, by funding provided by the State of Delaware. The Delaware Energy Institute acknowledges the support and partnership of the State of Delaware in furthering the essential scientific research being conducted through the RAPID projects. J.L.L.'s research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under contract no. DE-AC02-05CH11231. The 2019 to 2020 Blue Waters Graduate Fellowship to J.L.L. is also acknowledged. **Author contributions:** J.F. developed and implemented the PGM model. J.F. and M.A.K. developed UQ for PGMs. M.A.K. conceived the use of PGMs for model uncertainty, as well as the related information-theoretic tools. J.L.L. developed the physical model, and D.G.V. thought of the idea to make UQ explainable, to apply the PGM model to the ORR, and the need to apportion error to different model inputs for sparse data and missing physics. All authors contributed to writing. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All underlying DFT and experimental data are available on Zenodo. Software will be made available upon request. Additional data related to this paper may be requested from the authors.

Submitted 17 April 2020
Accepted 26 August 2020
Published 14 October 2020
10.1126/sciadv.abc3204

Citation: J. Feng, J. L. Lansford, M. A. Katsoulakis, D. G. Vlachos, Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences. *Sci. Adv.* **6**, eabc3204 (2020).

Science Advances

Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences

Jinchao Feng, Joshua L. Lansford, Markos A. Katsoulakis and Dionisios G. Vlachos

Sci Adv **6** (42), eabc3204.
DOI: 10.1126/sciadv.abc3204

ARTICLE TOOLS	http://advances.sciencemag.org/content/6/42/eabc3204
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2020/10/09/6.42.eabc3204.DC1
REFERENCES	This article cites 112 articles, 5 of which you can access for free http://advances.sciencemag.org/content/6/42/eabc3204#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.