Predicting binding motifs of complex adsorbates using machine learning with a physics-inspired graph representation

Wenbin Xu[1,2], Karsten Reuter[2] and Mie Andersen[3,4*]

[1]Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Lichtenbergstr. 4, Garching, D-85748, Germany.
[2]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, Berlin, D-14195, Germany.
[3*]Aarhus Institute of Advanced Studies, Aarhus University, Høegh-Guldbergs Gade 6B, Aarhus C, 8000, Denmark.
[4*]Department of Physics and Astronomy - Center for Interstellar Catalysis, Aarhus University, Ny Munkegade 120, Aarhus C, 8000, Denmark.

*Corresponding author(s). E-mail(s): mie@phys.au.dk;

Abstract

Computational screening in heterogeneous catalysis relies increasingly on machine learning models for predicting key input parameters due to the high cost of computing these directly using first-principles methods. This becomes especially relevant when considering complex materials spaces, e.g. alloys, or complex reaction mechanisms with adsorbates that may exhibit bi- or higher-dentate adsorption motifs. Here we present a data-efficient approach to the prediction of binding motifs and associated adsorption enthalpies of complex adsorbates at transition metals (TMs) and their alloys based on a customized Wasserstein Weisfeiler-Lehman graph kernel and Gaussian Process Regression. The model shows good predictive performance, not only for the elemental TMs on which it was trained, but also for an alloy based on these TMs. Furthermore, incorporation of minimal new training data allows for predicting an out-of-domain TM. We believe the model may be useful in active learning approaches, for which we present an ensemble uncertainty estimation approach.

Many surface catalytic reactions of vital importance to our society such as Fischer-Tropsch, methanol, or higher oxygenate synthesis have complex reaction mechanisms with numerous intermediates ranging from atoms and simple molecules to (possibly oxygenated) C1, C2 or larger fragments. It is well-known that modeling of these latter complex species at TM catalysts must account for their ability to exhibit a wide range of adsorption motifs, including mono-, bi- and higher-dentate adsorption modes.[1–3] Density-functional theory (DFT) with van der Waals corrections can in principle provide the energetics of such adsorption motifs at moderate cost and satisfactory accuracy. Nevertheless, already the identification of the most stable adsorption motifs of adsorbates involved in ethanol synthesis at a simple monometallic catalyst such as Rh(111) is a formidable task,[4–7] and the investigation of broader classes of materials such as TM alloys is generally out of reach due to the combinatorial

1

explosion of possible active sites and adsorption motifs.

Machine learning (ML) models have already shown their potential for replacing expensive DFT calculations in order to tackle the screening of large materials spaces for accelerated catalyst discovery.[8–12] However, most works so far have been limited in scope to the consideration of atoms or small molecules with mono-dentate adsorption motifs. For these simple species, models now routinely achieve the prediction of adsorption enthalpies with a root-mean-square-error (RMSE) around 0.1–0.2 eV, which is then comparable to the intrinsic DFT accuracy. Unfortunately, most of these methods cannot easily be extended to complex adsorbates with bi- or higher-dentate adsorption motifs. One notable attempt to treat complex adsorbates is provided in the Open Catalyst Project where the direct prediction of relaxed adsorption enthalpies is achieved by incorporating a graph representation of the initial structure into a graph convolutional neural network.[13] However, approaches that rely purely on connectivity and geometry-based features have revealed poor data efficiency and thus cannot be used without excessively large training databases.[13–15] Moreover, the predictive performance for complex adsorbates in the Open Catalyst Database is still below practical usefulness with a mean absolute error (MAE) for in-domain prediction around 0.6 eV.[13]

In this work we develop and test a data-efficient, physics-inspired ML model applicable for both simple and complex adsorbates based on graph representation, the Wasserstein Weisfeiler-Lehman (WWL) graph kernel [16], and Gaussian Process Regression (GPR). We abbreviate the model WWL-GPR. For comparison, we show also results for predictions of simple and complex adsorbates using two popular, fundamentally different ML approaches that employ input in vector form instead of graph representation, namely the Sure Independence Screening and Sparsifying Operator (SISSO) approach [17, 18] and GPR with a radial basis function kernel (RBF-GPR). We train our ML models for complex adsorbates on a relatively small database (around 1700 data points) of DFT adsorption enthalpies relevant to ethanol synthesis calculated at the fcc(211) and fcc(111) facets of four TMs; Cu, Rh, Pd, and Co. Compared to the Open Catalyst Dataset, our dataset is smaller by about a factor of 300, covers less diverse surfaces and adsorbates, but exhibits a much denser sampling of diverse adsorption motifs for each catalyst / adsorbate combination considered.

More importantly, we do not rely on graph representation alone, but augment it with node attributes representing physically motivated properties, e.g. d-band moments (surfaces), HOMO/LUMO energy levels (adsorbate molecules) and features of the local geometry, all derived from either the clean surfaces or the adsorbates in the gas phase. The model achieves an in-domain prediction of adsorption enthalpies with a RMSE of about 0.2 eV and also shows good extrapolative performance for two test cases; the CuCo bimetallic alloy and an out-of-domain element (Pt), the latter however only after incorporation of adsorption enthalpies of atomic species on Pt into the training database. Finally, we show that data points with large prediction errors can be quite reliably captured from an ensemble uncertainty estimation approach.

Results

WWL-GPR model

The ML task in our work is to directly predict the relaxed adsorption enthalpies corresponding to a range of plausible initial guesses of the adsorption motif based on graph representation. Thereby, for a given surface/adsorbate combination of interest, both the most stable and all meta-stable adsorption motifs as well as their associated adsorption enthalpies can be predicted. Our task is thus quite similar to the task denoted as IS2RE (initial state to relaxed energy) in the Open Catalyst Project, however, we do not directly use the initial state geometry, but only its graph representation. We note that an entirely different approach to this task is to train a ML interatomic potential [19, 20] to relax the initial structure and thereby predict both the relaxed structure and adsorption enthalpy – such approaches are however not a topic of this work.

Fig. 1(a) depicts a schematic of our physics-inspired WWL-GPR model. We rely on graph representation, which is a versatile method for representing isolated molecules [21, 22], crystal structures,[23] or the combined surface/adsorbate

system,[24–26] in which every atom in the structure is a node with edges representing chemical bonds to neighboring atoms. Graph representation can be used in connection with neural networks,[13–15] which however generally requires very large training databases. Since we are here interested in developing a data-efficient method, we focus on a kernel-based method (GPR) in connection with a customized version of the recently developed WWL graph kernel.[16] Fig. 1(b) illustrates the node embedding scheme, the calculation of the Wasserstein distance (distribution relationship) between the graphs, and the subsequent WWL graph kernel calculation. The WWL graph kernel allows for continuous node attributes, for which we use physically motivated electronic and geometric features calculated from the clean surface and isolated adsorbate. Finally, we incorporate into the WWL kernel some surface adsorption motivated hyperparameters to learn better representations, cf. Fig. 1(c); edge weights, which differentiate chemical bonds in the three classes adsorbate-adsorbate, surface-surface and adsorbate-surface, as well as inner and outer cutoffs and weights. The latter are used during the computation of the Wasserstein distance to emphasize the importance of various atomic shells around the active site for the adsorption energy prediction. We note that attention algorithms widely used in neural networks serve a similar purpose.[15]

Prediction of simple adsorbates

We begin by evaluating the performance of the three considered ML models, SISSO, RBF-GPR and WWL-GPR, for predicting a database of simple adsorbates with mono-dentate adsorption motifs (cf. Methods section). We carry out 5-fold cross validation, that is, the database is shuffled and partitioned into five equal-sized subsamples stratified by adsorbates. The training is then carried out based on four of the subsamples while retaining the fifth subsample for validation. This is repeated five times until all data points have been used once for validation. Fig. 2 shows the resulting parity plot of DFT-calculated against ML-predicted adsorption enthalpies as well as violin plots of the absolute error distributions. It should be noted here that the SISSO results are obtained

using similar hyperparameters as in our previous work [10, 27] (eight-dimensional rung three descriptor). In principle, we would expect a better performance than the here presented RMSE of 0.24 eV for even more complex models, cf. Fig. S5 in the Supporting Information (SI). However, the identification of more complex models is computationally intractable with the SISSO method. Rather than raw performance, the merit of the SISSO approach is that the identified descriptors are (somewhat simple) analytical functions of the features, which are thus easier to interpret than black-box ML models. We also note that the reason for the different performance of the descriptors identified in the present work compared to our previous work is that here we train a single model on the entire database (single-task learning) in order to be able to make a direct comparison to the GPR models, whereas in our previous work separate fitting coefficients were used for each adsorbate (multi-task learning). More information about the identified SISSO descriptors is provided in Section S3.2 in the SI.

For the GPR models the model complexity can be more easily tailored and after optimization of the relevant hyperparameters (cf. Section S3.5 and Table S9 in the SI) we obtain a RMSE of 0.13 eV independently of whether we use vector input (RBF-GPR) or graph representation (WWL-GPR). Also the maximum absolute error (maxAE) decreases from 1.11 eV (SISSO) to around 0.60 eV in the GPR models. Based on the similar performance of the two GPR models, we can conclude that there is no added value from employing graph representation for the simple adsorbates. The reactivity is apparently already captured by the averaged surface atom features and the adsorbate-specific features used in the RBF-GPR model.

Prediction of complex adsorbates

We next turn to a database of complex adsorbates with 41 different adsorbates in mono-, bi-, and higher-dentate adsorption motifs (cf. Methods section). Since we already concluded in the preceding section on simple adsorbates that single-task SISSO is not competitive in terms of performance, we focus here only on the GPR models. The 5-fold cross validation results presented in Fig. 3 show that for this more challenging database
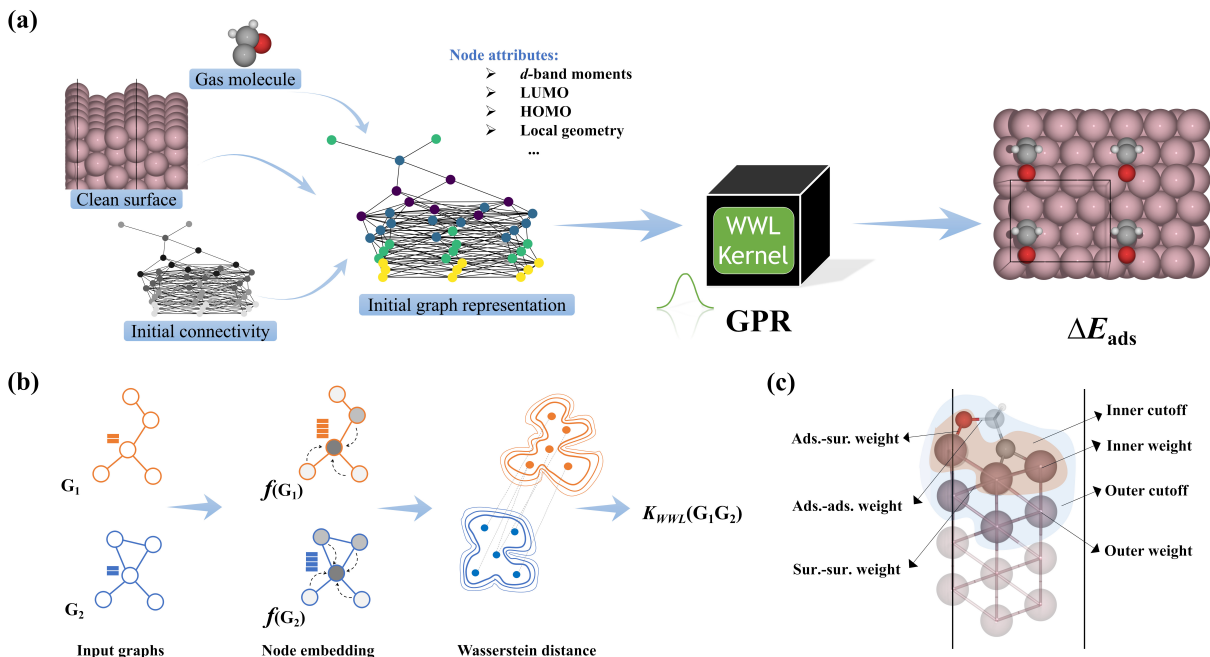
Fig. 1 Schematic illustration of the WWL-GPR model. (a) The adsorption enthalpy of the relaxed structure is predicted from a graph representation of the initial structure with node attributes computed from the gas-phase molecule and clean surface via the WWL graph kernel and GPR. (b) Illustration of the WWL graph kernel. (c) Surface adsorption motivated hyperparameters incorporated into the WWL kernel.
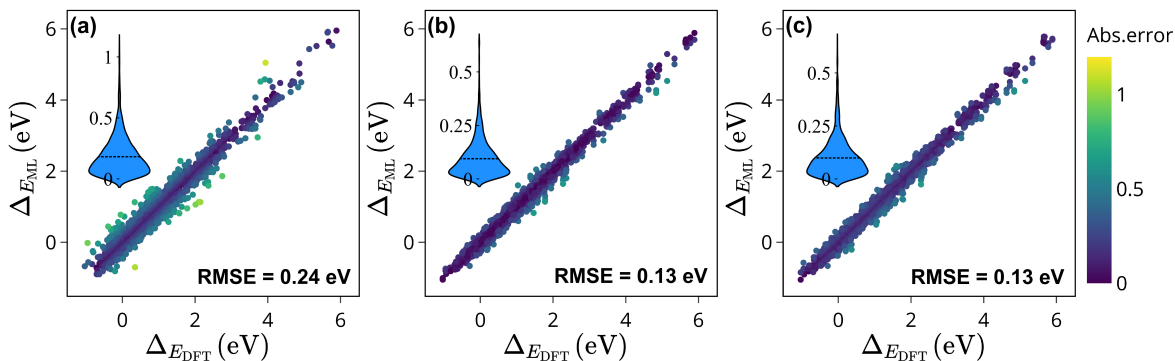

Fig. 2 Parity plot of DFT-calculated vs. ML-predicted adsorption enthalpies using (a) SISSO, (b) RBF-GPR, (c) WWL-GPR with 5-fold cross validation for the simple adsorbates database. The violin plots in the insets illustrate the absolute error distributions (in eV). The internal dashed line marks the mean.

the graph-based WWL-GPR model has a superior performance (RMSE of 0.18 eV) compared to RBF-GPR (RMSE of 0.47 eV). Also the maxAE decreases from 2.23 eV (RBF-GPR) to 0.91 eV in the WWL-GPR model. A learning curve for the WWL-GPR model is presented in Fig. S6 in the SI, which shows that an RMSE of 0.3 eV can be achieved by only training on 30% of the database

($\sim$ 500 data points) and a RMSE of 0.2 eV is achieved at 70% of the database ($\sim$ 1200 data points).

In order to visualize what trends the WWL-GPR model has identified in the complex adsorbates database, we present in Fig. 4 a kernel principal component analysis (KPCA), which is a non-linear dimensionality reduction technique.
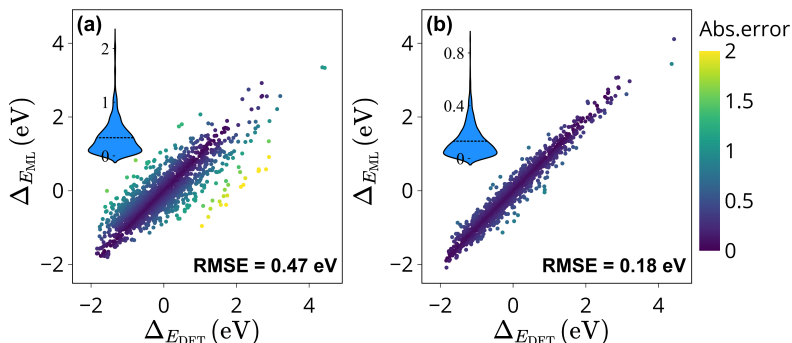
Fig. 3 Parity plot of DFT-calculated vs. ML-predicted adsorption enthalpies using (a) RBF-GPR and (b) WWL-GPR with 5-fold cross validation for the complex adsorbates database. The violin plots in the insets illustrate the absolute error distributions (in eV). The internal dashed line marks the mean.

Specifically, we here present the two dimensions that explain the largest fraction of the variance. Points that are close together in this space are similar in the feature space. The analysis of the entire complex adsorbate database in Fig. 4(a) shows that the different metals are distinguished as parallel clusters, where for each cluster there is a similar distribution of sub-clusters containing the individual adsorbates. In Fig. 4(b) the same analysis is presented for the Rh surfaces only. Again, the different adsorbates form clusters, where each point in a cluster corresponds to a separate adsorption motif of the adsorbate. A similar clustering cannot be observed in KPCA plots for the RBF-GPR model, cf. Fig. S7 in the SI, which is probably related to the fact that this model does not have any structural information about the different adsorbates and their associated adsorption motifs due to the lack of graph representation.

Having established the excellent interpolation performance of the WWL-GPR model, we next assess the predictive performance of the model for extrapolation tasks concerning data that are dissimilar to those in the training database, i.e. out-of-domain prediction. This is highly important for the practical application of the model to catalyst screening. The two tasks we consider are a) predictions for a bimetallic CuCo catalyst, i.e. an alloy of elemental metals present in our database, and b) predictions for a new element (Pt) when merely incorporating adsorption enthalpies of atomic species (C, H, and O) at Pt into the database. For these tasks we selected 8 adsorbates spanning both atomic species and

larger molecules, and including some with bidentate adsorption motifs (cf. Table S4 in the SI). Since it has previously been emphasized in the literature that in extrapolative, data-poor regimes, a careful choice of regularization can substantially improve the robustness of a model [28, 29], we re-optimized the hyperparameters for the extrapolation tasks. Specifically, they were optimized by minimizing the loss function ($\text{RMSE}_{\text{interpolation}}$ + $2*\text{RMSE}_{\text{extrapolation}}$), where $\text{RMSE}_{\text{extrapolation}}$ is the RMSE of the combined dataset for the two extrapolation tasks. Since these latter tasks are more challenging, they were given a higher weight (two) in the loss function than the weight of the interpolation task (one). Comparing the hyperparameters obtained previously for the complex adsorbates database (base case in Table S10 in the SI) with the new hyperparameters optimized specifically for the extrapolation tasks (base case in Table S11 in the SI), we see that indeed both the length scale and the regularization term increase for the extrapolation tasks, resulting in a smoother ML model, which is consistent with the previous literature observations. The obtained RMSEs for the new hyperparameters are 0.25 eV for interpolation within the complex adsorbate database, 0.23 eV for task a) (CuCo alloy) and 0.30 eV for task b) (Pt). The higher RMSE obtained for task b) reflects that an unknown element is much harder to predict than an alloy of known elements, even when incorporating some minimal information about the unknown element into the training database through the atomic adsorption enthalpies. We would expect the performance for
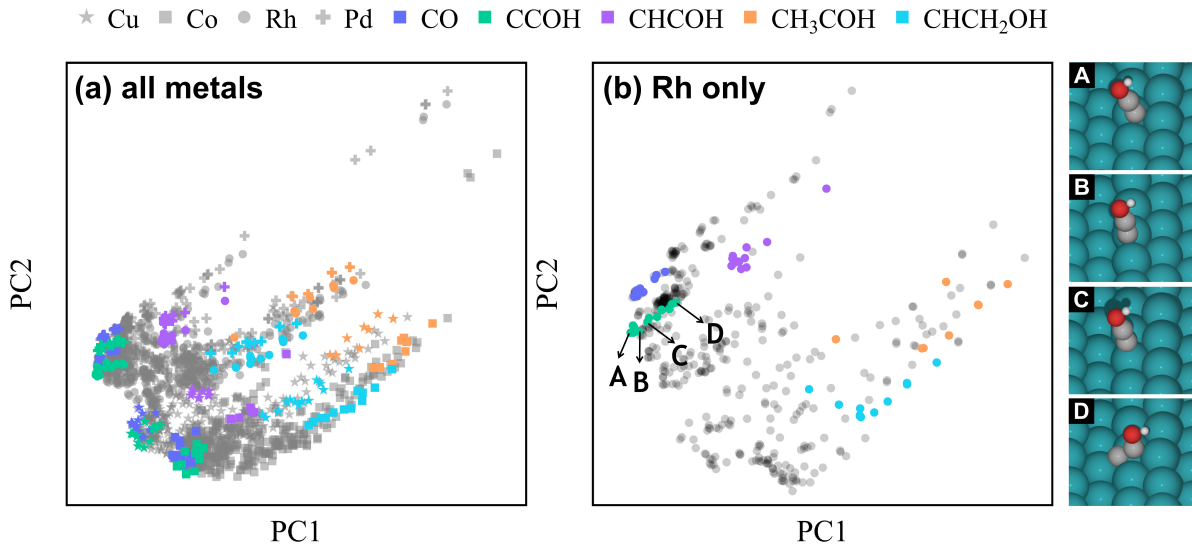
Fig. 4 Two-dimensional kernel principal component (PC) analysis plots for the WWL-GPR model with hyperparameters optimized for interpolation for (a) all metals and (b) Rh surfaces only. The locations and structures of selected adsorbates are highlighted.

Pt to improve if more adsorbates were added to the training database.

Uncertainty quantification

Up till now we have demonstrated that our WWL-GPR model can be quite confidently applied with RMSE around 0.2–0.3 eV to flat and stepped metal and bimetallic catalysts, as long as some (at least minimal) training data involving the considered elemental metals are provided. However, apart from the average RMSE to expect, it is also useful to be able to directly assess the expected uncertainty on a single predicted data point. For example, uncertainty quantification (UQ) combined with sensitivity analysis of microkinetic models[30–32] can be used to assess error propagation and the extent to which conclusions drawn from a model are robust to input parameter uncertainty.[33, 34] Furthermore, UQ is used in active learning approaches, where the training database is iteratively updated through selected DFT calculations, e.g. of data points with a high estimated uncertainty.[35, 36]

In view of these applications, we are here primarily interested in the extent to which a high estimated uncertainty correlates with a high actual error of the model predictions. To assess this point, we compare the intrinsic UQ provided in a single GPR model through the standard deviation (SD) of the posterior distribution to the UQ provided by the SD of an ensemble (100 in total) of GPR models with fixed hyperparameters optimized for interpolation. The latter are constructed through bootstrapping of the training data, i.e. data points are drawn randomly with replacement. Specifically, we here use a random 80%/20% training/test split of the complex adsorbates database stratified by adsorbate. Note that the added computational cost of establishing the ensemble model is insignificant since we use a fixed training/test split, and since the kernel between the training and test set only needs to be computed once.

As expected, the prediction accuracy obtained from the single and the ensemble model is almost identical (RMSE of 0.17 eV versus 0.18 eV, respectively). A plot of estimated uncertainties versus absolute prediction errors of the two models is presented in Fig. 5(a) and (b). For comparison, we show also in Fig. 5(c-f) some distribution-based measures of the quality of a UQ that have recently been discussed in the literature, i.e. calibration, sharpness, and dispersion.[37] A useful UQ method should have a small miscalibration

area (a good match between the expected and observed cumulative error distribution), a small sharpness value (small error estimates) and a large dispersion value (disperse error estimates). For these latter quantities the performances of the single and ensemble models are quite similar with the single model having a slightly better calibration and the ensemble model having a slightly better sharpness and dispersion. However, with our primary interest being active learning, it is much more intriguing to see that the ensemble model does a better job than the single model at assigning a high uncertainty to data points with a high actual prediction error. In particular, the group of points with an estimated uncertainty higher than 0.2 eV in the ensemble model includes the largest prediction errors, whereas this is not the case for the single model, which actually assigns a quite low uncertainty to some of the largest prediction errors. We therefore conclude that the ensemble model is best suited for active learning approaches. We note here that we do not expect any quantitative match between the absolute error and the uncertainty in Fig. 5, partly because these are not directly comparable quantities (one is a SD and the other an absolute error), and partly because it has been shown that specific calibration measures [38, 39] are required for quantitatively accurate UQ in both single and ensemble GPR models.

Discussion

We begin by discussing the origin of the superior performance of the WWL-GPR model over the RBF-GPR model. First of all, we note that it is not surprising that for complex adsorbates, simply accounting for the surface and adsorbate in terms of features averaged over the atoms directly involved in the bonding as well as global features of the adsorbate (e.g. HOMO/LUMO levels) and clean surface (e.g. work function) as done in the RBF-GPR model is insufficient. In contrast, the graph representation provides direct access to structural information about the system, i.e. the number and types of atoms in the adsorbate and how these atoms connect to each other and to the surface, possibly in complex bi- or higher-dentate adsorption motifs. Atom-specific features related to the local electronic or geometric structure can be directly used as node attributes, e.g.

through SOAP descriptors, and we can introduce surface adsorption motivated hyperparameters as discussed above and illustrated in Fig. 1(c). In this connection, it is interesting to note that while the WWL-GPR model finds that the optimal cutoff values are one node distance for both inner and outer cutoff for the simple adsorbates database (i.e. mostly the atoms directly involved in surface/adsorbate bonding are judged important), the optimal inner and outer cutoffs (weights) are one (0.60) and two (0.06) node distances, respectively, for the complex adsorbates database, cf. Table S9 in the SI (i.e. also atoms neighboring the immediately bonding atoms are judged important, although with smaller weights). The effect of more distant atoms is not taken into account in the RBF-GPR model, which then possibly relates to its decreased performance for complex adsorbates. Note also that during the node embedding scheme of the WWL graph kernel, the node attribute of every atom is updated with information about the node attributes of the neighboring atoms, cf. Section S3.4.2 in the SI. That is, even if weights beyond the outer cutoff are zero, the atoms there can still have a non-negligible influence on the kernel value.

With its impressive performance for an alloy made up of elements present in our database (CuCo, RMSE of 0.23 eV) as well as its reasonable performance for a new metal based on only minimal training data (Pt, RMSE of 0.30 eV), we trust that our WWL-GPR model could be useful for catalyst screening purposes, e.g. for exploring reactions with complex adsorbates on alloy surfaces. Here the complexity encountered from the many possible adsorption motifs of each adsorbate on each type of alloy surface makes direct DFT investigations computationally intractable, while reliable ML force fields or density-functional tight-binding methods for the simultaneous treatment of many different adsorbates and/or alloy surfaces are still difficult to obtain.[2, 13]

We envision that it could be particularly rewarding to apply our model in the context of an active learning strategy, where the training database is iteratively expanded towards catalytically interesting and/or previously poorly explored regions of the catalyst space. Since it is important for active learning to have reliable UQ, we compared the performance of the intrinsic UQ in a single GPR model with that from
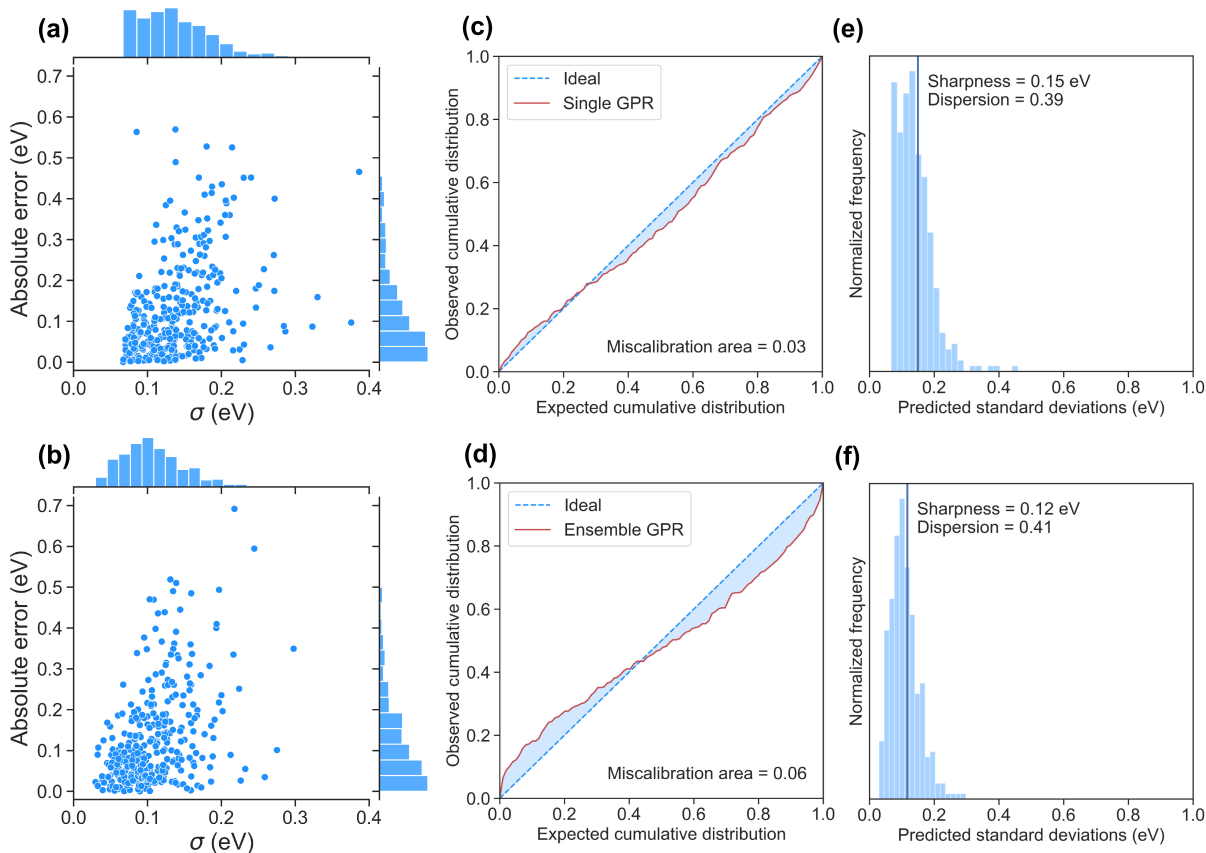
Fig. 5 The uncertainty vs. absolute error, calibration curve, and distribution plot of the predicted standard deviations for (a), (c) and (e) the single GPR model and (b), (d) and (f) the ensemble GPR model, respectively.

an ensemble of GPR models, and find that the ensemble model is the most reliable for capturing large prediction errors by flagging them with a high estimated uncertainty. Furthermore, it is important to have a low model training cost, if the model is to be iteratively improved by the incorporation of new data (e.g. an interesting new metal or alloy surface) followed by model retraining. This is a key advantage of our data-efficient GPR model compared to e.g. deep neural networks. For active learning purposes, we also recommend to use the model with different hyperparameter settings depending on the exploitative or explorative nature of the task at hand. Specifically, we can confirm previous literature reports that hyperparameters characterized, among others, by larger length scale and regularization terms are beneficial for accurate predictions in data-poor regions of the catalyst space.

Methods

DFT databases

The ML models are trained and tested on two different databases termed 'simple adsorbates' and 'complex adsorbates'. The database of simple adsorbates is taken from Refs. [10, 40]. After a post-processing step the database contains 1422 data points and includes adsorption enthalpies of eight simple adsorbates with mono-dentate coordination; C, H, O, CO, OH, CH, CH_2 and CH_3. The considered surfaces include the fcc(100), fcc(110), fcc(111) and fcc(211) facets of pure Ni, Cu, Ru, Rh, Pd, Ag, Ir, Pt and Au, the bcc(210) facet of Fe as well as the stepped hcp(0001) facet of Co. For alloy catalysts, the database contains the adsorbates on the four single-atom alloys Ag@Cu, Pt@Rh, Pd@Ir, and Au@Ni (i.e. the single atom Ag, Pt, Pd or Au dispersed in the surface of

another host metal) and the four AB bimetallic alloys AgPd, IrRu, PtRh, and AgAu.

The complex adsorbates database contains 1679 data points and includes 41 different small and large adsorbates involved in ethanol synthesis on fcc(111) and fcc(211) facets of Cu, Rh, Pd, and Co. Examples of complex adsorbates are CHCO, CCHOH, CH_2CH_2O and CH_3CH_2OH, and the full list of adsorbates is provided in Table S3 in the SI. Furthermore, selected adsorbates are calculated at the CuCo(111) surface and the Pt(111) and Pt(211) surfaces for model testing purposes. The adsorbates contain up to nine atoms and cover mono-, bi-, and higher-dentate adsorption modes. The database is constructed using an automated workflow and DFT settings that are compatible with the simple adsorbates database (Quantum Espresso code [41], BEEF-vdW functional [42]). Further computational details and overviews of both databases are given below and in Section S1 of the SI.

Database construction and workflow

The initial geometries of the surface/adsorbate systems are generated using the CatKit software.[25] CatKit employs a graph representation of the surface atoms to enumerate mono- and bi-dentate adsorption sites, where the latter are defined by a neighboring node-edge pair of the graph. For each adsorbate, a manual tagging of the bonding atoms for mono- and bi-dentate adsorption motifs is required (see Table S3 in the SI). CatKit then adds the adsorbates at the enumerated adsorption sites by employing some simple geometric procedures to produce good guesses for the angles and bond lengths in the system. We note that CatKit obviously does not generate all possible adsorption motifs (which would be computationally intractable), but only those that are judged most plausible. This adds a human bias into the generation of the database. Furthermore, not all initial geometries generated are actually stable, but could transform into other structures during the DFT relaxation.

In order to overcome some of these limitations, we added the following steps to our computational workflow. During the DFT relaxation, we monitor the graph representation of the system and assign it to the following four cases. (i) if the graph representation is unchanged, the data point is simply added to our database (32.3 % of cases). (ii) if the structure transforms into another graph which is already covered in the CatKit-enumerated structures (28.4 % of cases), only the calculation with the most favorable adsorption enthalpy is added to the database to avoid duplicates. (iii) if the structure transforms into a non-valid graph, i.e. a graph that is incompatible with our direct graph-based ML model (e.g. adsorbate dissociation, surface reconstruction) the calculation is discarded (23.4 % of cases). (iv) if the structure transforms into a valid graph that was not enumerated by CatKit (15.9 % of cases), the data point is added to the database with updated initial graph representation and the new adsorption motif is tested also for the other surfaces of interest. The latter case (iv) as well as large adsorbates whose initial adsorption motifs cannot be well controlled by CatKit are the source of all higher-dentate adsorption motifs in our database (see examples in Fig. S3 in the SI). Our workflow is implemented with AIIDA,[43] which is a scalable computational infrastructure providing advanced automation to allow interfacing with external simulation software. In our case this entails customized python scripts interfacing with CatKit, the Atomic Simulation Environment (ASE) software [44] and the Quantum Espresso DFT code.

DFT computational details

The simple adsorbates database used here is taken from Ref. [40] and [10]. For the DFT calculations of the complex adsorbates database, the following settings were used in full compliance with the simple adsorbates database. We used the Quantum ESPRESSO code [41] employing the Bayesian error estimation functional with van der Waals correlation (BEEF-vdW) [42] and ultrasoft pseudopotentials. Pseudopotentials for Cu, Rh, Pd, and Pt were generated using the "atomic" code by A. Dal Corso (v.5.0.2 svn rev. 9415) [45] and for Co using the Vanderbilt code version 7.0.0 [46]. To relieve the interaction between the adsorbates, we modelled the fcc(211) slab in a (3×1) cell and the fcc(111) slab in a (3×3) cell. In both cases this corresponds to 9 atoms per atomic layer. The CuCo(111) alloy surface is modelled in a (4×2) cell and contains 16 atoms per layer. We used a (4×4) k-point grid for the pure metal slabs and a

(3×3) grid for the CuCo alloy slab. All slabs contained four atomic layers, where the bottom two layers were kept fixed in their bulk-truncated positions, while the top layers and the adsorbates were relaxed until the maximum force on each atom fell below 0.05 eV/Å (cf. Fig. S1 in the SI). All DFT calculations were carried out as periodic slab calculations employing a vacuum region of 20 Å perpendicular to the surface and a dipole correction. Spin polarization was taken into account for the calculations involving Co. The cutoff energy was set to 500 eV and 5000 eV for the orbitals, and the charge density, respectively, and a Fermi level smearing of 0.1 eV was used. The resulting adsorption enthalpies are formation energies referenced to gaseous CH_3OH, CO, and H_2O.

The features that require DFT calculations were obtained as follows. For the clean surfaces involved in both the simple and complex adsorbate databases, we first carried out a geometry relaxation as outlined above. The projected density of states (PDOS) was calculated using the smearing-free tetrahedron method and an energy spacing of 0.01 eV. We used a (14×14) k-point grid for the pure metal fcc and bcc slabs, a (7×21) grid for the SG225 fcc alloys, a (14×21) for the SG221 fcc alloys, a (7×14) grid for the Co hcp slab, and a (7×42) for the hcp alloy structures, except for the CuCo(111) surface used in the complex adsorbate database, where we used a (12×12) grid.

For the calculation of band moments, we integrated empty bands up to the energy above the Fermi level where the PDOS had fallen below a value of 0.01 $Å^{-3}eV^{-1}$. The features involving the density of states at the Fermi level were calculated using a smearing of 0.1 eV in the PDOS calculation, and the PDOS was averaged over the interval ± 0.1 eV around the Fermi level. For the calculation of adsorbate-specific features, we carried out a structural optimization of the isolated adsorbate positioned in a cubic supercell with a side length of 15.0 Å. We used a Fermi-level smearing of 0.01 eV and the Brillouin zone integration was performed using the Gamma point only.

Further details on ML models

The WWL-GPR model is compared to two other ML models (SISSO and RBF-GPR) that do not use graph representation, but input in vector form with features of the clean surface and of the isolated adsorbates. The features used in SISSO and RBF-GPR are specific to the surface, site or adsorbate considered, where site-specific features are calculated by averaging over the metal atoms to which the adsorbate coordinates (clean surface features) or the bonding atoms of the adsorbate (isolated adsorbate features). The WWL-GPR model also uses atom-specific features as node attributes, for example, electronic properties of individual surface atoms or features of the local geometry of the clean surface and isolated adsorbate through Smooth Overlap of Atomic Positions (SOAP) descriptors.[47] All details about the features used in the three compared ML models are provided in Section S2 in the SI and Section S3 provides more information about each of the models, including a more in-depth discussion of hyperparameters.

It should be emphasized that the WWL-GPR model leverages only features from the initial guess geometry, specifically the graph connectivity, and electronic and geometric features calculated from the clean surface and isolated adsorbate. From a computational screening point of view this is essential for keeping the computational cost of model predictions low. The computationally most intensive part of the model prediction is the DFT calculation of the clean surface to obtain the node attributes (e.g. d-band moments) for the surface atoms. However, given that we target 41 different adsorbates in various possible adsorption motifs for each surface, this is still a low cost per ML prediction.

For SISSO, we previously used an approach to target simple adsorbates where the free parameters of the identified models were fitted to each adsorbate separately.[10, 27]

A similar approach has been taken in most other works targeting simple adsorbates.[11, 14, 48, 49]

In the present work we instead fit a single model to all adsorbates, and the different adsorbates are then instead distinguished from each other via adsorbate-specific features such as HOMO/LUMO energy levels.

Supplementary Information. Additional details on DFT databases, primary features, machine learning models, and kernel principle component analysis are provided.

Author contributions. W.X. carried out the DFT calculations, workflow and ML methods development. K.R. and M.A. conceived and supervised the project. All authors contributed to analyzing the data and writing the manuscript.

Competing interests. The authors declare no competing interests.

References

[1] Cao, A., Schumann, J., Wang, T., Zhang, L., Xiao, J., Bothra, P., Liu, Y., Abild-Pedersen, F., Nørskov, J.K.: Mechanistic insights into the synthesis of higher alcohols from syngas on CuCo alloys. ACS Catal. **8**(11), 10148–10155 (2018)

[2] Chang, C., Medford, A.J.: Application of density functional tight binding and machine learning to evaluate the stability of biomass intermediates on the Rh(111) surface. J. Phys. Chem. C **125**(33), 18210–18216 (2021)

[3] Wang, Z., Li, Y., Boes, J., Wang, Y., Sargent, E.: CO_2 electrocatalyst design using graph theory (2020). Preprint at https://doi.org/10.21203/rs.3.rs-66715/v1

[4] Choi, Y., Liu, P.: Mechanism of ethanol synthesis from syngas on Rh (111). J. Am. Chem. Soc. **131**(36), 13054–13061 (2009)

[5] Michel, C., Auneau, F., Delbecq, F., Sautet, P.: C–H versus O–H bond dissociation for alcohols on a Rh (111) surface: A strong assistance from hydrogen bonded neighbors. ACS Catal. **1**(10), 1430–1440 (2011)

[6] Filot, I.A.W., Broos, R.J.P., van Rijn, J.P.M., van Heugten, G.J.H.A., van Santen, R.A., Hensen, E.J.M.: First-principles-based microkinetics simulations of synthesis gas conversion on a stepped rhodium surface. ACS Catal. **5**(9), 5453–5467 (2015)

[7] Gu, T., Wang, B., Chen, S., Yang, B.: Automated generation and analysis of the complex catalytic reaction network of ethanol synthesis from syngas on Rh(111). ACS Catal. **10**(11), 6346–6355 (2020)

[8] Tran, K., Ulissi, Z.W.: Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. Nat. Catal. **1**(9), 696–703 (2018)

[9] Noh, J., Back, S., Kim, J., Jung, Y.: Active learning with non-*ab initio* input features toward efficient CO_2 reduction catalysts. Chem. Sci. **9**(23), 5152–5159 (2018)

[10] Andersen, M., Levchenko, S.V., Scheffler, M., Reuter, K.: Beyond scaling relations for the description of catalytic materials. ACS Catal. **9**(4), 2752–2759 (2019)

[11] Wang, S.-H., Pillai, H.S., Wang, S., Achenie, L.E., Xin, H.: Infusing theory into deep learning for interpretable reactivity prediction. Nat. Commun. **12**(1), 1–9 (2021)

[12] Fung, V., Hu, G., Ganesh, P., Sumpter, B.G.: Machine learned features from density of states for accurate adsorption energy prediction. Nat. Commun. **12**(1), 88 (2021)

[13] Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., Palizhati, A., Sriram, A., Wood, B., Yoon, J., Parikh, D., Zitnick, C.L., Ulissi, Z.: Open Catalyst 2020 (OC20) dataset and community challenges. ACS Catal. **11**(10), 6059–6072 (2021)

[14] Back, S., Yoon, J., Tian, N., Zhong, W., Tran,

K., Ulissi, Z.W.: Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. J. Phys. Chem. Lett. **10**(15), 4401–4408 (2019)

[15] Gu, G.H., Noh, J., Kim, S., Back, S., Ulissi, Z., Jung, Y.: Practical deep-learning representation for fast heterogeneous catalyst screening. J. Phys. Chem. Lett. **11**(9), 3185–3191 (2020)

[16] Togninalli, M., Ghisu, E., Llinares-López, F., Rieck, B., Borgwardt, K.: Wasserstein Weisfeiler-Lehman graph kernels (2019). Preprint at https://arxiv.org/abs/1906.01277

[17] Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M., Ghiringhelli, L.M.: SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. Phys. Rev. Mater. **2**(8), 083802 (2018)

[18] Ouyang, R., Ahmetcik, E., Carbogno, C., Scheffler, M., Ghiringhelli, L.M.: Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. J. Phys.: Mater. **2**(2), 024002 (2019)

[19] Deringer, V.L., Caro, M.A., Csányi, G.: Machine learning interatomic potentials as emerging tools for materials science. Adv. Mater. **31**(46), 1902765 (2019)

[20] Klicpera, J., Becker, F., Günnemann, S.: Gemnet: Universal directional graph neural networks for molecules (2021). Preprint at https://arxiv.org/abs/2106.08903

[21] Wen, M., Blau, S.M., Spotte-Smith, E.W.C., Dwaraknath, S., Persson, K.A.: BonDNet: A graph neural network for the prediction of bond dissociation energies for charged molecules. Chem. Sci. **12**(5), 1858–1868 (2021)

[22] Tang, Y.-H., de Jong, W.A.: Prediction of atomization energy using graph kernel and active learning. J. Chem. Phys. **150**(4), 044107 (2019)

[23] Xie, T., Grossman, J.C.: Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys. Rev. Lett. **120**(14), 145301 (2018)

[24] Montoya, J.H., Persson, K.A.: A high-throughput framework for determining adsorption energies on solid surfaces. Npj Comput. Mater. **3**(1), 14 (2017)

[25] Boes, J.R., Mamun, O., Winther, K., Bligaard, T.: Graph theory approach to high-throughput surface adsorption structure generation. J. Phys. Chem. A **123**(11), 2281–2285 (2019)

[26] Deshpande, S., Maxson, T., Greeley, J.: Graph theory approach to determine configurations of multidentate and high coverage adsorbates for heterogeneous catalysis. Npj Comput. Mater. **6**(1), 79 (2020)

[27] Xu, W., Andersen, M., Reuter, K.: Data-driven descriptor engineering and refined scaling relations for predicting transition metal oxide reactivity. ACS Catal. **11**(2), 734–742 (2020)

[28] Rupp, M.: Machine learning for quantum mechanics in a nutshell. Int. J. Quantum Chem. **115**(16), 1058–1073 (2015)

[29] Deringer, V.L., Bartók, A.P., Bernstein, N., Wilkins, D.M., Ceriotti, M., Csányi, G.: Gaussian process regression for materials and molecules. Chem. Rev. **121**(16), 10073–10141 (2021)

[30] Bruix, A., Margraf, J.T., Andersen, M., Reuter, K.: First-principles-based multiscale modelling of heterogeneous catalysis. Nat. Catal. **2**(8), 659–670 (2019)

[31] Meskine, H., Matera, S., Scheffler, M., Reuter, K., Metiu, H.: Examination of the concept of degree of rate control by first-principles kinetic monte carlo simulations. Surf. Sci. **603**(10-12), 1724–1730 (2009)

[32] Medford, A.J., Wellendorff, J., Vojvodic, A., Studt, F., Abild-Pedersen, F., Jacobsen,

K.W., Bligaard, T., Nørskov, J.K.: Assessing the reliability of calculated catalytic ammonia synthesis rates. Science **345**(6193), 197–200 (2014)

[33] Sutton, J.E., Guo, W., Katsoulakis, M.A., Vlachos, D.G.: Effects of correlated parameters and uncertainty in electronic-structure-based chemical kinetic modelling. Nat. Chem. **8**(4), 331 (2016)

[34] Döpking, S., Matera, S.: Error propagation in first-principles kinetic monte carlo simulation. Chem. Phys. Lett. **674**, 28–32 (2017)

[35] Flores, R.A., Paolucci, C., Winther, K.T., Jain, A., Torres, J.A.G., Aykol, M., Montoya, J., Nørskov, J.K., Bajdich, M., Bligaard, T.: Active learning accelerated discovery of stable iridium oxide polymorphs for the oxygen evolution reaction. Chem. Mater. **32**(13), 5854–5863 (2020)

[36] Kunkel, C., Margraf, J.T., Chen, K., Oberhofer, H., Reuter, K.: Active discovery of organic semiconductors. Nat. Commun. **12**(1), 1–11 (2021)

[37] Tran, K., Neiswanger, W., Yoon, J., Zhang, Q., Xing, E., Ulissi, Z.W.: Methods for comparing uncertainty quantifications for material property predictions. Mach. Learn.: Sci. Technol. **1**(2), 025006 (2020)

[38] Palmer, G., Du, S., Politowicz, A., Emory, J.P., Yang, X., Gautam, A., Gupta, G., Li, Z., Jacobs, R., Morgan, D.: Calibrated bootstrap for uncertainty quantification in regression models (2021). Preprint at https://arxiv.org/abs/2105.13303

[39] Kuleshov, V., Fenner, N., Ermon, S.: Accurate uncertainties for deep learning using calibrated regression. In: International Conference on Machine Learning, pp. 2796–2804 (2018). PMLR

[40] Deimel, M., Reuter, K., Andersen, M.: Active site representation in first-principles microkinetic models: Data-enhanced computational screening for improved methanation catalysts. ACS Catal. **10**(22), 13729–13736

(2020)

[41] Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G.L., Cococcioni, M., Dabo, I., Corso, A.D., de Gironcoli, S., Fabris, S., Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N., Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A., Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G., Seitsonen, A.P., Smogunov, A., Umari, P., Wentzcovitch, R.M.: QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J. Phys.: Condens. Matter **21**(39), 395502 (2009)

[42] Wellendorff, J., Lundgaard, K.T., Møgelhøj, A., Petzold, V., Landis, D.D., Nørskov, J.K., Bligaard, T., Jacobsen, K.W.: Density functionals for surface science: Exchange-correlation model development with bayesian error estimation. Phys. Rev. B **85**(23), 235149 (2012)

[43] Huber, S.P., Zoupanos, S., Uhrin, M., Talirz, L., Kahle, L., Häuselmann, R., Gresch, D., Müller, T., Yakutovich, A.V., Andersen, C.W., Ramirez, F.F., Adorf, C.S., Gargiulo, F., Kumbhar, S., Passaro, E., Johnston, C., Merkys, A., Cepellotti, A., Mounet, N., Marzari, N., Kozinsky, B., Pizzi, G.: AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. Sci. Data **7**(1), 300 (2020)

[44] Larsen, A.H., Mortensen, J.J., Blomqvist, J., Castelli, I.E., Christensen, R., Dułak, M., Friis, J., Groves, M.N., Hammer, B., Hargus, C., Hermes, E.D., Jennings, P.C., Jensen, P.B., Kermode, J., Kitchin, J.R., Kolsbjerg, E.L., Kubal, J., Kaasbjerg, K., Lysgaard, S., Maronsson, J.B., Maxson, T., Olsen, T., Pastewka, L., Peterson, A., Rostgaard, C., Schiøtz, J., Schütt, O., Strange, M., Thygesen, K.S., Vegge, T., Vilhelmsen, L., Walter, M., Zeng, Z., Jacobsen, K.W.: The atomic simulation environment—a python library for working with atoms. J. Phys.: Condens. Matter **29**(27), 273002 (2017)

[45] Dal Corso, A.: Pseudopotentials periodic table: From H to Pu. Comput. Mater. Sci. **95**, 337–350 (2014)

[46] Garrity, K.F., Bennett, J.W., Rabe, K.M., Vanderbilt, D.: Pseudopotentials for high-throughput DFT calculations. Comput. Mater. Sci. **81**, 446–452 (2014)

[47] Bartók, A.P., Kondor, R., Csányi, G.: On representing chemical environments. Phys. Rev. B. **87**(18), 184115 (2013)

[48] Esterhuizen, J.A., Goldsmith, B.R., Linic, S.: Theory-guided machine learning finds geometric structure-property relationships for chemisorption on subsurface alloys. Chem **6**(11), 3100–3117 (2020)

[49] Andersen, M., Reuter, K.: Adsorption enthalpies for catalysis modeling through machine-learned descriptors. Acc. Chem. Res. **54**(12), 2741–2749 (2021)