

# Chapter 3

## Machine Learning Predictions of Factors Affecting the Activity of Heterogeneous Metal Catalysts

Ichigaku Takigawa, Ken-ichi Shimizu, Koji Tsuda  
and Satoru Takakusagi

**Abstract** The ultimate goal in heterogeneous catalytic science is to accurately predict trends in catalytic activity based on the electronic and geometric structures of active metal surfaces. Such predictions would allow the rational design of materials having specific catalytic functions without extensive trial-and-error experiments. The d-band center values of metals are well known to be an important parameter affecting the catalytic activity of these materials, and activity trends in metal surface catalyzed reactions can be explained based on the linear Brønsted–Evans–Polanyi relationship and the Hammer–Nørskov d-band model. The present work demonstrates the possibility of employing state-of-the-art machine learning methods to predict the d-band centers of metals and bimetals while using negligible CPU time compared to the more common first-principles approach.

**Keywords** Heterogeneous catalysis • d-band center • Machine learning

---

I. Takigawa  
Graduate School of Information Science and Technology, Hokkaido University,  
Sapporo 060-0814, Japan  
e-mail: takigawa@ist.hokudai.ac.jp

K. Shimizu · S. Takakusagi (✉)  
Institute for Catalysis, Hokkaido University, Sapporo 001-0021, Japan  
e-mail: takakusa@cat.hokudai.ac.jp

K. Shimizu  
e-mail: kshimizu@cat.hokudai.ac.jp

K. Tsuda  
Graduate School of Frontier Sciences, University of Tokyo, Kashiwa 277-8561, Japan  
e-mail: tsuda@k.u-tokyo.ac.jp

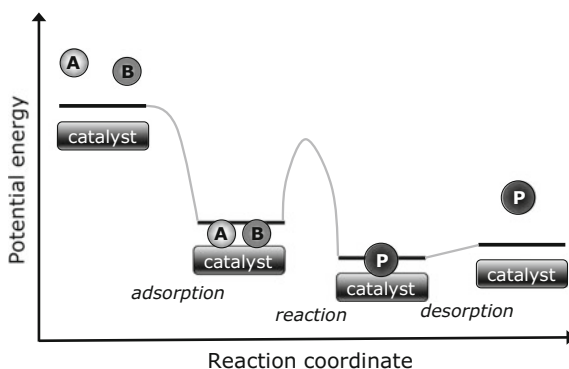
### 3.1 Introduction

Heterogeneous catalysis plays a key role in the industrial production of various chemicals. Over 80% of catalytic processes use heterogeneous catalysts to achieve high conversion and/or selectivity through lowering the activation barriers leading to the desired products [1, 2]. The majority of these materials consist of active transition metal or alloy nanoparticles dispersed on oxide supports, such as  $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_2$ ,  $\text{MgO}$ ,  $\text{TiO}_2$ , and  $\text{CeO}_2$ .

Heterogeneous catalysis is a surface phenomenon that involves a sequence of elementary steps, including adsorption, surface diffusion, chemical rearrangement of the adsorbed intermediates (the actual reaction), and desorption of the products, as shown in Fig. 3.1. Thus, detailed experimental and theoretical characterizations of the surface electronic/geometric structures during these steps are indispensable in order to understand the reaction mechanisms and to be able to enhance the activity and selectivity of the catalysts. However, even though surface characterization techniques have improved dramatically in recent years, new catalytic materials are still primarily developed through trial-and-error experiments. This is because catalytic reactions are actually more complicated than the process illustrated in Fig. 3.1 due to the complexity of catalyst surface structures and the effects of a large number of parameters (such as temperature, pressure, metal particle size/shape, and metal–support interactions). Unfortunately, the empirical development of catalytic materials is typically time-consuming and expensive with no guarantee of success.

For these reasons, the theory-based rational prediction of activity trends in catalysis is one of the ultimate goals in catalytic science. Such predictions would allow the design of surfaces with specific catalytic properties without extensive experimentation. To this end, it is important to elucidate the factors that control activity, also known as descriptors. To date, the bond energies derived from bulk oxide properties or the adsorption energies of reactants have been used as descriptors to predict the activity of metal/metal oxide surfaces [1]. Activity–descriptor plots typically exhibit a so-called volcano shape due to several effects. First, the strong

**Fig. 3.1** Potential energy diagram of a heterogeneous catalytic reaction ( $A + B \rightarrow P$ ) with gaseous reactants (A, B), product (P), and solid metal catalyst



binding of an intermediate can result in surface poisoning, whereas weak binding leads to low coverage of the surface; in both cases, the catalytic rates are less than optimal. Consequently, moderate interactions produce the highest reactivity (representing Sabatier's principle). In addition, a linear relationship between activation energy and adsorbate–surface interaction energy, known as the Brønsted–Evans–Polanyi relation, has been demonstrated by several groups based on theoretical calculations [3–8]. The above effects allow a semiquantitative understanding of the activity trends in heterogeneous catalytic systems by simply considering the bond energies derived from bulk oxide properties and/or the adsorption energies of a reactant to a first approximation.

Recently, a simple but powerful approach based on machine learning (ML) techniques combined with density functional theory (DFT) calculations has attracted much attention as a novel tool for the rapid screening of metal catalyst reactivity. This method makes it possible to predict various catalyst properties typically calculated using DFT, such as reactant gas adsorption energies on various metal or alloy surfaces. This is done by constructing an appropriate regression model and using explanatory variables (often termed descriptors) that correlate with intrinsic properties of the constituent metals and/or reactant gases. Once the regression model is successfully constructed, it permits the rapid identification of the optimal catalyst for a target reaction by interpolation without calculating results for all the other candidates. Ras and Rothenberg et al. presented a simple and efficient model based on genetic algorithm variable selection and Partial Least Squares (PLS) regression for predicting the adsorption of molecules (heats of adsorption) on metal surfaces [9]. Their model used six descriptors for each metal (number of d-electrons, surface energy, first ionization potential, as well as atomic radius, volume, and mass) and three for each adsorptive species (HOMO–LUMO energy gap, molecular volume, and mass). This method was found to accurately predict the chemisorption of a range of adsorptive compounds ( $\text{H}_2$ , HO,  $\text{N}_2$ , CO, NO,  $\text{O}_2$ ,  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ ,  $\text{NH}_3$ , and  $\text{CH}_4$ ) on a variety of metals (Fe, Co, Ni, Cu, Mo, Ru, Rh, Pd, Ag, W, Ir, Pt, and Au) as calculated using DFT or reported in the literature. This group also acquired experimental adsorption data for CO,  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{H}_2$ ,  $\text{N}_2$ , and  $\text{O}_2$  on Ni, Pt, and Rh supported on  $\text{TiO}_2$ , and confirmed that their model, using the same descriptors, generated results in good agreement with the data. Ma and Xin et al. systematically calculated CO adsorption energies on 250–300 {100}-terminated multimetallic alloy surfaces and presented an ML-augmented chemisorption model for  $\text{CO}_2$  electroreduction catalyst screening [10, 11]. They demonstrated that artificial neural networks are able to reproduce the complex, nonlinear interactions of CO adsorbed on multimetallic alloy surfaces with an error of approximately 0.1 eV. The associated results identified multimetallic alloys that show promise with regard to improving the efficiency and selectivity during the electrochemical reduction of  $\text{CO}_2$  to  $\text{C}_2$  species. Okamoto developed a method based on a combination of DFT calculations and data mining to find the optimum composition for PtRu alloys to minimize the CO adsorption energy, since CO poisoning of the alloy catalysts tends to deactivate the catalytic function in proton exchange membrane fuel cells (PEMFCs) [12]. He first calculated the CO

adsorption energies on 44 PtRu(111) bimetallic slabs having various compositions and subsequently employed multiple regression analysis for the data mining. This work determined that the resulting model accurately predicted CO adsorption energies on PtRu surfaces. This regression model also identified the optimum composition associated with a minimum CO adsorption energy, which was later confirmed by DFT calculations using the same alloy composition.

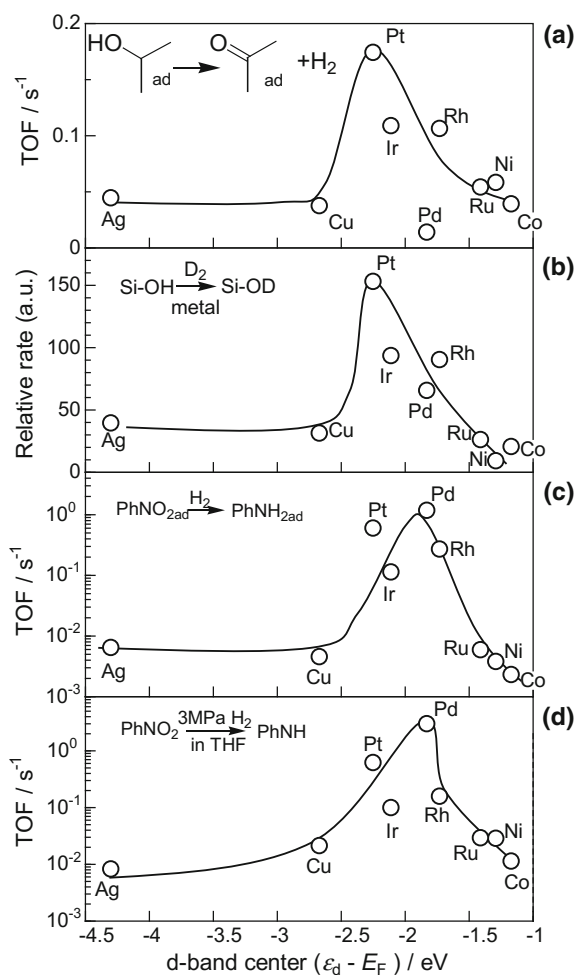
The above examples demonstrate that ML techniques can effectively predict the interaction energy between a specific adsorbate and a given metal surface for a particular reaction, and can sometimes assist in finding the optimal catalytic material. However, the interaction energy may not always be used as a universal descriptor for predicting activity trends in different catalytic reactions by various transition metal catalysts. For this reason, the present work focused on the so-called d-band center, which is one of the most important activity-controlling factors and can be used to explain activity trends in various types of catalytic reactions. In this study, we employed state-of-the-art ML techniques to predict the DFT-calculated d-band centers for metals and bimetals.

### 3.2 The d-Band Center: A Widely Accepted Indicator Explaining Activity Trends in Metal Catalysts

Nørskov et al. performed a series of systematic DFT calculations and proposed the semiempirical concept of the d-band model [3–5]. The model assumes that the d-electrons of transition metals play the most important role in chemisorption. This approach involves linear scaling between the energy of the d-band center ( $\epsilon_d$ ) relative to the Fermi level ( $E_F$ ) and the adsorption energy for a given adsorbate. The higher the d-states are in energy relative to the Fermi level, the emptier the anti-bonding states and the larger the adsorption energy of an adsorbed species on a surface. A calorimetric study by Lu et al. [13] subsequently provided experimental evidence to support the d-band model. This work showed moderate linear correlations between the experimental heats of adsorption of CO, H<sub>2</sub>, O<sub>2</sub>, and C<sub>2</sub>H<sub>4</sub> on various metal surfaces and the positions of the d-band centers as calculated by Hammer and Nørskov [3]. The d-band model also predicts that adsorbate binding energies should correlate with one another [5]. Since the transition-state structures on different metals tend to be rather similar, the activation energy for an elementary reaction should exhibit a linear relationship with the energy change for the elementary reaction. Thus, the kinetic parameter for a catalytic reaction involving a metal can be written as  $\epsilon_d - E_F$ , equivalent to the position of the d-band center relative to  $E_F$ . Recent experimental studies have demonstrated the validity of the d-band model when describing trends in catalytic activity [14–18]. As an example, Furukawa et al. found a relationship between the d-band centers of Ni and Ni<sub>3</sub>M (M = Ge, Nb, Sn, Ta, or Ti) intermetallics and their activation energies with regard to the H<sub>2</sub>-D<sub>2</sub> equilibration [16].

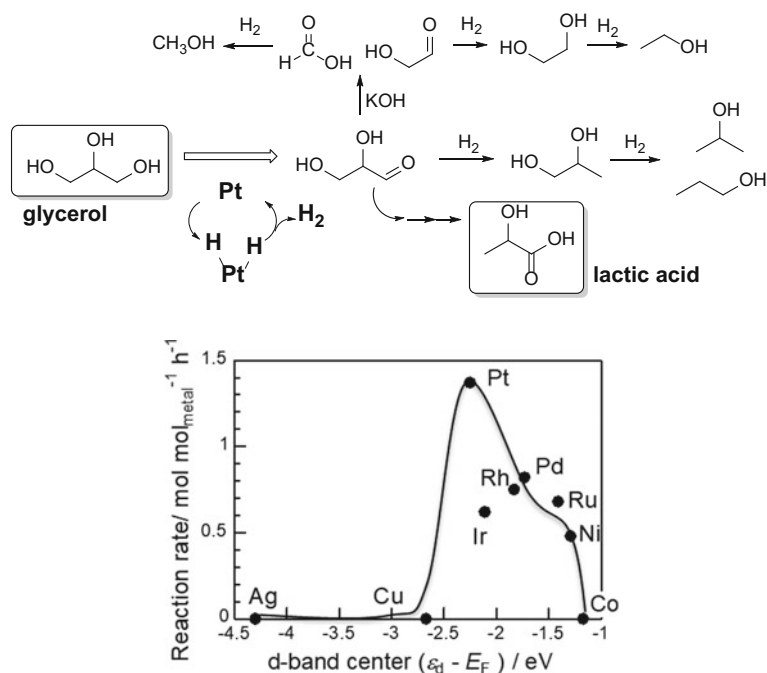
To confirm whether the activity trends in multistep catalytic reactions can be understood in terms of the d-band model in combination with linear energy relations, Tamura et al. studied correlations between the reaction rates of dehydrogenation and hydrogenation reactions and the associated  $\epsilon_d - E_F$  values [17]. The activities per surface metal atom, or turnover frequency (TOF), for various metal-loaded  $\text{SiO}_2$  samples with similar particle size ranges (8.9–11.7 nm) were plotted against the d-band center values (Fig. 3.2). In the cases of the dehydrogenation of 2-propanol adspecies (2- $\text{PrOH}_{\text{ad}}$ ) on the surface (Fig. 3.2a), the hydrogenation of  $\text{PhNO}_{2\text{ad}}$  (Fig. 3.2b), the OH/OD exchange of surface SiOH groups under  $\text{D}_2$  (Fig. 3.2c), and the liquid phase hydrogenation of  $\text{PhNO}_2$  by  $\text{M}/\text{SiO}_2$  (Fig. 3.2d), the activities generally show volcano-type variations with the d-band center values, except for the Pd catalyst in Fig. 3.2a. A common trend is

**Fig. 3.2** The activities of metal catalysts for various reactions versus d-band center values



evident in which, as the d-band center moves further from  $E_F$ , the metal–hydrogen (M–H) and metal–oxygen (M–O) bond energies become weaker [4]. Each of the reactions in Fig. 3.2 includes the formation and dissociation of M–H bonds as a common elementary step. Dehydrogenation and hydrogenation reactions include the formation and dissociation of M–O bonds. Hence, these results suggest that moderate M–H and/or M–O bond strengths favor the reactions. The observation of similar volcano-type trends for different reactions demonstrates that the d-band center can serve as a general activity–descriptor for the catalytic systems shown in Fig. 3.2a. This outcome can possibly be explained by considering that the strong binding of surface intermediates via M–H and M–O bonds leads to surface poisoning, whereas weak binding limits the availability of the intermediates. In both cases, the catalytic rates are less than optimal. Consequently, Pt-group metal catalysts with moderate bond strengths give the highest activities.

If the rate-limiting step and/or relatively slow steps involve the formation and decomposition of the same bond (for example, a metal–hydrogen bond), the d-band center can serve as a descriptor for a complicated multistep reaction. The conversion of glycerol to lactic acid (and byproducts) [18] is a typical example of a complex multistep reaction involving the formation and decomposition of a metal–hydrogen bond (Fig. 3.3). In this case, there is a good volcano-type correlation



**Fig. 3.3** The activities of metal catalysts during the dehydrogenation of glycerol to lactic acid as a function of the d-band center value

between the d-band center value and the catalytic activity. Pt, having an intermediate d-band center level, shows higher catalytic activity than the other metals, because the interaction between surface intermediates and the metal surface is moderately strong, which tends to favor metal-catalyzed dehydrogenation.

The above examples demonstrate that the d-band model (in combination with linear energy relations) can be used to understand the activity trends in transition metal-catalyzed multistep reactions. We can conclude that this model is an important concept with regard to assessing or predicting reactivity trends in the heterogeneous catalysis of transition metals during multistep organic reactions. Thus, as a first approximation, the reactivity of a metal catalyst for an organic reaction can be described by a single parameter: the d-band center.

### 3.3 Prediction of the d-Band Center Values for Mono- and Bimetallic Systems by Machine Learning

#### 3.3.1 *Data-Driven Prediction of d-Band Center Values by Machine Learning Methods*

Herein, we present our most recent results regarding the ML-based predictions of d-band centers for metallic and bimetallic compounds [19]. Using DFT calculations, Nørskov's group [3, 20] determined the d-band centers for 11 different metals (Fe, Co, Ni, Cu, Ru, Rh, Pd, Ag, Ir, Pt, and Au) and for the associated 110 bimetallic pairs having two different structures (surface impurities and overlayers on clean metal surfaces) [21]. In this case, the d-band centers were independently calculated using first principles for each metal or bimetal under typical conditions. In contrast, our own study involved a quantitative investigation of a fully data-driven approach based on ML that infers the d-band center of a metal or a bimetal from those of other metals and bimetals. As an example, it would be of significant interest to know whether or not the d-band center of the Cu–Co pair can be somehow inferred from those of Cu, Au, Cu–Fe, Ni–Ru, Pd–Co, and Rh–Pd from the materials informatics perspective. Our result shows sufficient predictability of d-band centers by ML methods using a small set of readily available properties of metals as descriptors. Given the rapid increase in data in recent years, this outcome would suggest that ML methods may possibly substitute for or complement first-principles calculations.

#### 3.3.2 *Datasets and Descriptors*

To assess the accuracy of ML predictions, we employed  $\epsilon_d - E_F$  data for 11 metals (Fe, Co, Ni, Cu, Ru, Rh, Pd, Ag, Ir, Pt, and Au) and all the associated pairwise

**Table 3.1** The “impurities” dataset: DFT-calculated d-band centers (eV) of metals (bold) and 1% guest metals ( $M_g$ ) doped into the surfaces of host metals ( $M_h$ ) as reported by Nørskov’s group [3, 20]. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

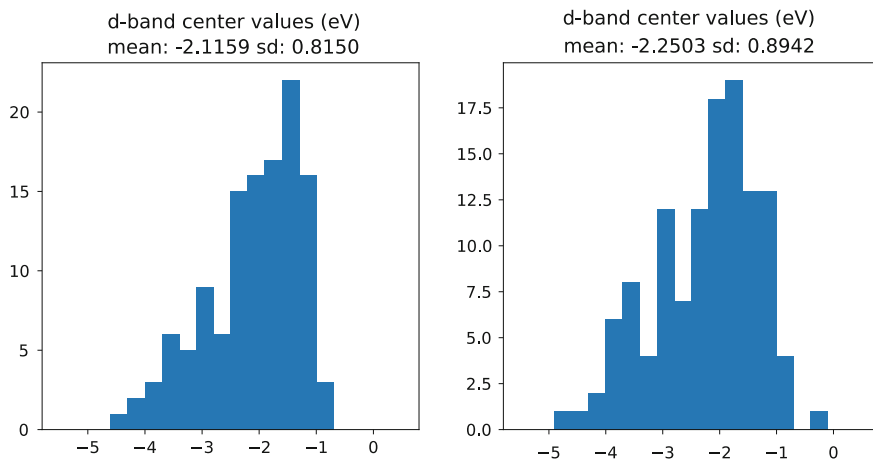
$M_g$ $M_h$	Fe	Co	Ni	Cu	Ru	Rh	Pd	Ag	Ir	Pt	Au
Fe	<b>-0.92</b>	-0.87	-1.12	-1.05	-1.21	-1.46	-2.16	-1.75	-1.28	-2.01	-2.34
Co	-1.16	<b>-1.17</b>	-1.45	-1.33	-1.41	-1.75	-2.54	-2.08	-1.53	-2.36	-2.73
Ni	-1.20	-1.10	<b>-1.29</b>	-1.10	-1.43	-1.60	-2.26	-1.82	-1.43	-2.09	-2.42
Cu	-2.11	-2.07	-2.40	<b>-2.67</b>	-2.09	-2.35	-3.31	-3.37	-2.09	-3.00	-3.76
Ru	-1.20	-1.15	-1.40	-1.29	<b>-1.41</b>	-1.58	-2.23	-1.68	-1.39	-2.03	-2.25
Rh	-1.49	-1.39	-1.57	-1.29	-1.69	<b>-1.73</b>	-2.27	-1.66	-1.56	-2.08	-2.22
Pd	-1.46	-1.29	-1.33	-0.89	-1.59	-1.47	<b>-1.83</b>	-1.24	-1.30	-1.64	-1.66
Ag	-3.58	-3.46	-3.63	-3.83	-3.46	-3.44	-4.16	<b>-4.30</b>	-3.16	-3.80	-4.45
Ir	-1.90	-1.84	-2.06	-1.90	-2.02	-2.26	-2.84	-2.24	<b>-2.11</b>	-2.67	-2.85
Pt	-1.92	-1.77	-1.85	-1.53	-2.11	-2.02	-2.42	-1.81	-1.87	<b>-2.25</b>	-2.30
Au	-2.93	-2.79	-2.93	-3.01	-2.86	-2.81	-3.39	-3.35	-2.58	-3.10	<b>-3.56</b>

bimetallic alloys (110 pairs of a host metal,  $M_h$ , and a guest metal,  $M_g$ ). These values were obtained from a DFT study by Refs. [3, 20] for two different structures: those having surface impurities (Table 3.1) and those with overlayers (Table 3.2). In the original datasets, the d-band centers for bimetals are given as shifts relative to the clean metal values and so have been converted to values relative to the Fermi level. In Table 3.1, the surfaces considered are the most closely packed, and 1% guest metals are doped into the topmost surfaces of the host metals. In Table 3.2, the overlayer structures are pseudomorphic and guest metal monolayers are formed on the surface of the host metals. The histograms of d-band centers for each

**Table 3.2** The “overlayers” dataset: DFT-calculated d-band centers (eV) of metals (bold) and the guest metal ( $M_g$ ) monolayers on the surfaces of host metals ( $M_h$ ) as reported by Nørskov’s group [3, 20]. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

$M_g$ $M_h$	Fe	Co	Ni	Cu	Ru	Rh	Pd	Ag	Ir	Pt	Au
Fe	<b>-0.92</b>	-0.78	-0.96	-0.97	-1.65	-1.64	-2.24	-2.17	-1.87	-2.40	-3.11
Co	-1.18	<b>-1.17</b>	-1.37	-1.23	-1.87	-2.12	-2.82	-2.53	-2.26	-3.06	-3.56
Ni	-0.33	-1.18	<b>-1.29</b>	-1.17	-1.92	-2.03	-2.61	-2.43	-2.15	-2.82	-3.39
Cu	-2.42	-2.29	-2.49	<b>-2.67</b>	-2.89	-2.94	-3.71	-3.88	-2.99	-3.82	-4.63
Ru	-1.11	-1.04	-1.12	-1.11	<b>-1.41</b>	-1.53	-1.88	-1.81	-1.54	-2.02	-2.27
Rh	-1.42	-1.32	-1.39	-1.51	-1.70	<b>-1.73</b>	-2.12	-1.81	-1.70	-2.18	-2.30
Pd	-1.47	-1.29	-1.29	-1.03	-1.94	-1.58	<b>-1.83</b>	-1.68	-1.52	-1.79	-1.97
Ag	-3.75	-3.56	-3.62	-3.68	-3.80	-3.63	-4.03	<b>-4.30</b>	-3.50	-3.93	-4.51
Ir	-1.78	-1.71	-1.78	-1.55	-2.12	-2.14	-2.53	-2.20	<b>-2.11</b>	-2.60	-2.70
Pt	-1.90	-1.72	-1.71	-1.47	-2.13	-2.01	-2.23	-2.06	-1.96	<b>-2.25</b>	-2.33
Au	-3.03	-2.82	-2.85	-2.86	-3.09	-2.89	-3.21	-3.44	-2.77	-3.13	<b>-3.56</b>





**Fig. 3.4** Histogram of d-band centers for the “impurities” (left) and “overlayers” (right) datasets

structure are provided in Fig. 3.4. Although the two structures are physically very different, the Pearson’s correlation coefficient between Tables 3.1 and 3.2 is 0.948 ( $p < 0.001$ ) and thus the d-band centers exhibit significant correlation. Therefore, in order to differentiate these structure-specific values, any data-driven prediction requires a highly adaptive mechanism that can capture the subtle differences.

Regarding the choice of descriptors for metals, we pretested several candidates and chose nine physical properties (Table 3.3) that are readily available from the periodic table and a standard reference source [22]. From a practical point of view, it is important to choose readily accessible but characteristic values as descriptors in order to effectively bypass time-consuming DFT calculations while maintaining sufficient prediction accuracy. Each metal can thus be represented as a nine-dimensional vector of the descriptor values. Accordingly, an 18-dimensional concatenated vector of  $M_h$  and  $M_g$  values was used for predictions of the d-band centers of bimetals. In the case of monometallic surfaces, we employed an 18-dimensional vector by concatenating two vectors for the same metal. We also searched for smaller subsets of descriptors yielding simpler models among the 18 descriptors by assessing the relevance or redundancy of each descriptor. Table 3.4 shows the correlation matrix between descriptors and demonstrates highly correlated descriptor variables. This result prompted us to investigate variable selection with the aim of identifying a smaller nonredundant subset of the 18 descriptors. Table 3.5 indicates the correlation coefficients between each descriptor and the d-band center values. It can be seen that no single descriptor exhibits direct correlation with the d-band centers.

**Table 3.3** Input features (descriptors) used for the prediction of d-band centers from Ref. [22]. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

Metal	G	R/Å	AN	AM/ g mol <sup>-1</sup>	P	EN	IE/ eV	$\Delta_{\text{fus}}H$ / J g <sup>-1</sup>	$\rho$ / g cm <sup>-3</sup>
Fe	8	2.66	26	55.85	4	1.83	7.90	247.3	7.87
Co	9	2.62	27	58.93	4	1.88	7.88	272.5	8.86
Ni	10	2.60	28	58.69	4	1.91	7.64	290.3	8.90
Cu	11	2.67	29	63.55	4	1.90	7.73	203.5	8.96
Ru	8	2.79	44	101.07	5	2.20	7.36	381.8	12.10
Rh	9	2.81	45	102.91	5	2.28	7.46	258.4	12.40
Pd	10	2.87	46	106.42	5	2.20	8.34	157.3	12.00
Ag	11	3.01	47	107.87	5	1.93	7.58	104.6	10.50
Ir	9	2.84	77	192.22	6	2.20	8.97	213.9	22.50
Pt	10	2.90	78	195.08	6	2.20	8.96	113.6	21.50
Au	11	3.00	79	196.97	6	2.40	9.23	64.6	19.30

Group ( $G$ )

Bulk Wigner–Seitz radius ( $R$ ) in Å

Atomic number ( $AN$ )

Atomic mass ( $AM$ ) in g mol<sup>-1</sup>

Period ( $P$ )

Electronegativity ( $EN$ )

Ionization energy ( $IE$ ) in eV

Enthalpy of fusion ( $\Delta_{\text{fus}}H$ ) in J g<sup>-1</sup>

Density at 25 °C ( $\rho$ ) in g cm<sup>-3</sup>

**Table 3.4** Correlation matrix of the nine descriptors for the 11 metals in Table 3.3. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

	G	R/Å	AN	AM	P	EN	IE	$\Delta_{\text{fus}}H$	$\rho$
G	1.00								
R/Å	0.43	1.00							
AN	0.24	0.77	1.00						
AM	0.24	0.75	1.00	1.00					
P	0.18	0.82	0.98	0.98	1.00				
EN	0.07	0.66	0.78	0.77	0.84	1.00			
IE	0.28	0.48	0.82	0.84	0.74	0.53	1.00		
$\Delta_{\text{fus}}H$	-0.75	-0.72	-0.58	-0.58	-0.53	-0.28	-0.66	1.00	
$\rho$	0.13	0.61	0.97	0.98	0.94	0.73	0.84	-0.46	1.00

### 3.3.3 Monte Carlo Cross-Validation for Assessing the Prediction Accuracies of ML Models

Our primary intent was to assess the data-driven prediction of the d-band center of a given metal (or bimetal) from the d-band centers of other metals and bimetals. To

**Table 3.5** Correlation coefficients between each of the 18 descriptors and the d-band centers. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

(For host metal)									
	G	R/Å	AN	AM	P	EN	IE	$\Delta_{\text{fus}}\text{H}$	$\rho$
Impurities	-0.63	-0.53	-0.29	-0.29	-0.26	-0.02	-0.15	0.56	-0.17
Overlayers	-0.63	-0.34	-0.11	-0.11	-0.06	0.13	-0.06	0.49	0.00
(For guest metal)									
	G	R/Å	AN	AM	P	EN	IE	$\Delta_{\text{fus}}\text{H}$	$\rho$
Impurities	-0.24	-0.35	-0.26	-0.26	-0.27	-0.28	-0.24	0.33	-0.20
Overlayers	-0.22	-0.47	-0.41	-0.40	-0.42	-0.39	-0.31	0.36	-0.33

do so, we first randomly separated 121 targets (11 metals and 110 bimetals) into two disjoint sets: a “test set” of size  $n$  and a “training set” of size  $121-n$ . The subsequent challenge was to evaluate the accuracy with which the d-band centers of the test set could be predicted using those of the training set. As a first step, an ML model was constructed using the training set. Following this, the model was employed to predict the d-band centers of the test set and the root-mean-square errors (RMSEs) between the predicted and true values (the ground truth) were calculated for the purposes of predictability evaluation. A single-shot random trial of this procedure could provide estimates of RMSE values, whereas those estimates vary depending on the split between the training and test sets (with a certain level of variance). For quantitative evaluations, we reduced this estimation variance by repeating the single-shot trials over 100 random test/training splits (that is, for 100 random leave- $n$ -out trials) and used the mean of 100 RMSE estimates to assess the prediction accuracy of the ML model. The test set in each trial was never used to build the corresponding ML model in that trial, and hence simulated yet unseen targets to be predicted. Another benefit of this approach was that it was also possible to control the size,  $n$ , of the test set, and so to determine the training set size required for accurate predictions. It should be noted that this general approach is well established in statistics and is referred to as Monte Carlo cross-validation [23] or as leave- $n$ -out [24], random permutation cross-validation (shuffle and split) [25], or random subsampling cross-validation [26]. This method was determined to be a better match for our scenario than more typical choices such as k-fold cross-validation or bootstrapping.

### 3.3.4 Machine Learning Methods and Hyperparameter Selection

We initially selected 11 ML regression models: five linear models (linear regression (OLS), PLS regression (PLS), L1-penalized linear regression (LASSO), L2-penalized linear regression (RIDGE), and robust linear regression (RANSAC))

**Table 3.6** List of the ML regression methods and tuning parameters (the hyperparameters to be tuned)

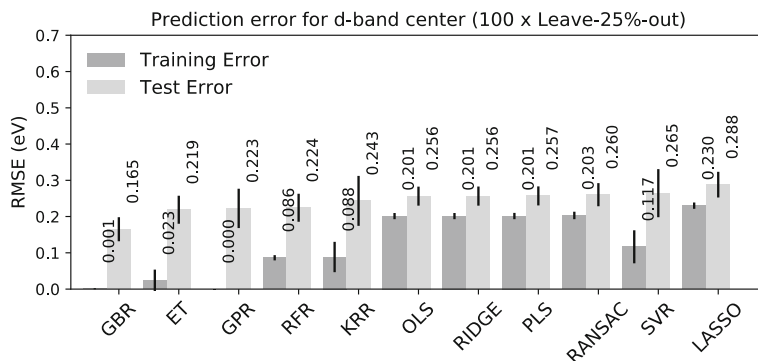
Abbreviation	Tuning parameters [tested range]
OLS, RANSAC	(No tuning parameters) (min_samples = 50 for RANSAC)
PLS	n_components $\in$ [1, 2, ..., # of vars]
LASSO, RIDGE	alphas $\in$ [1.0, $10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$ , $10^{-5}$ ]
GPR	theta0 $\in$ [1.0, $10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$ , $10^{-5}$ ]
KRR	alpha, gamma $\in$ [1.0, $10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$ , $10^{-5}$ ]
SVR	C $\in$ [1.0, 10, $10^2$ , $10^3$ , $10^4$ ] gamma $\in$ [1.0, $10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$ , $10^{-5}$ ]
RFR, ET	max_depth $\in$ [4, 6, 8, 10] n_estimators $\in$ [100, 250, 300]
GBR	learning_rate $\in$ [1.0, $10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$ , $10^{-5}$ ] max_depth $\in$ [4, 6, 8, 10] n_estimators $\in$ [100, 250, 300]

and six nonlinear models (Gaussian-process regression (GPR), kernel ridge regression (KRR), support vector regression (SVR), random forest regression (RFR), extra-trees regression (ET), and gradient boosting regression (GBR)). Each of these approaches is based on a popular, easy-to-use, off-the-shelf ML package: scikit-learn (<http://scikit-learn.org>) [24]. The models selected herein are those most commonly used in the ML field, and details concerning the individual methods can be found in standard ML references [27, 28].

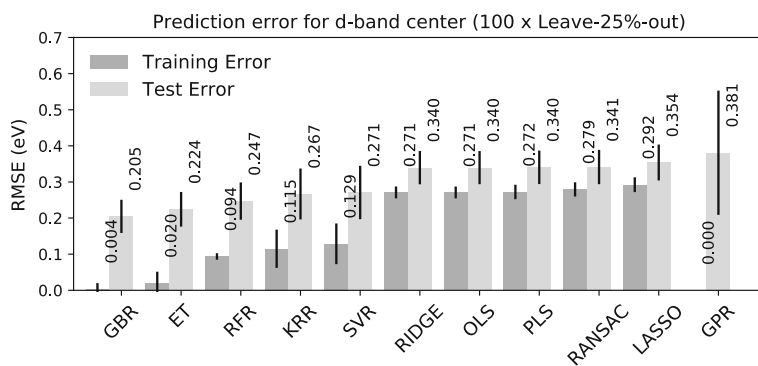
In practice, some models include tuning parameters called *hyperparameters* in addition to the target parameters to be estimated from the training set, and these hyperparameters must be determined prior to training. In such cases, the appropriate setting of these parameters is the key to successful predictions. These hyperparameters were determined by assessing a reasonable range of candidate values in an exhaustive manner (via a grid search), as shown in Table 3.6, and choosing the best parameters by threefold cross-validation with the training set. It should be noted that this selection process was performed for each training/test split independently, and the test data in each split were never used to select a hyperparameter.

### 3.3.5 Screening and Evaluation of Predictive ML Methods

We evaluated the prediction performance of 11 ML models using Monte Carlo cross-validation with 100 random leave-25%-out splits, in conjunction with internal threefold cross-validation with the training set to ensure selection of the optimum model (Table 3.6). That is, the following random trials were each performed 100 times. Assuming that 25% of Tables 3.1 or 3.2 has not yet been obtained, the ML method statistically infers those values using the other available 75% of the values. Following this, the RMSE of the difference between the predicted values and the ground truth is calculated for each trial, and these RMSEs are then averaged to



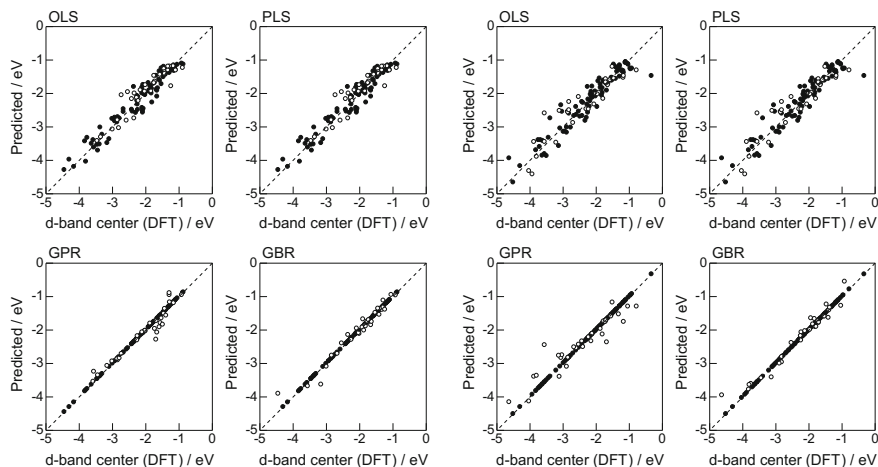
**Fig. 3.5** Prediction performances of the 11 ML models for the “impurities” dataset



**Fig. 3.6** Prediction performances of the 11 ML models for the “overlayers” dataset

obtain the mean RMSE and its standard deviation. The results of d-band center prediction evaluations are shown in Figs. 3.5 and 3.6 for the surface impurity and surface overlayer trials, respectively. Among the various methods examined, the GBR approach exhibited the best prediction performance. This was not unexpected, since GBR [29] is widely used and has performed well in top-level data prediction contests such as the Kaggle competition in recent years [30, 31]. Technically, it is an ensemble model composed of boosted regression trees, which often give accurate and stable predictions.

Figure 3.7 illustrates the predictive performances of four typical ML methods (OLS, PLS, GPR, and GBR) in a single-shot cross-validation with a 75% training set (●) and a 25% test set (○). These trials used all 18 descriptors: nine for the host and nine for the guest metal. The x-axis in these plots represents the DFT-calculated d-band center values (the ground truth values), while the y-axis gives the predictions from the ML methods. Deviations from the  $x = y$  line indicate prediction errors. Clearly, the predictions by linear models (OLS and PLS) exhibit larger



**Fig. 3.7** DFT-calculated local d-band centers for the “impurities” (left) and “overlayers” (right) datasets. Legend: (●) training set = 75%, (○) test set = 25%. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

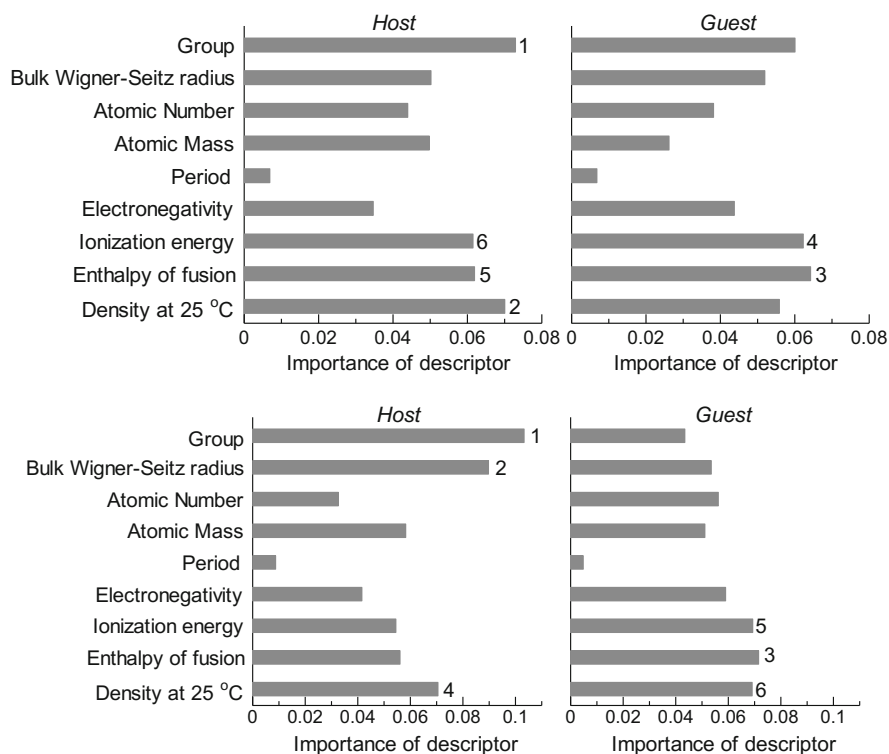
deviations for test sets than those by the nonlinear models (GPR and GBR), while the GBR model exhibits the least deviation from the line. It should be noted that the PLS method performs best at the hyperparameter setting of  $n_{\text{components}} =$  the number of descriptor variables, implying that linear dimensional reduction does not work for this problem because the PLS is identical to OLS in this setting.

The mean RMSE values of the linear models were larger than those of the nonlinear models, suggesting that the nonlinear models were more accurate. From these results, we concluded that GBR would be the best choice for the prediction of the d-band centers. It is known that the GBR model is more flexible than the linear regression models and exhibits greater stability than the GPR model, which is more sensitive to the hyperparameter settings. It should be noted that the linear models show lower standard deviations than the nonlinear models, which are more flexible but also more dependent on hyperparameter settings. Thus, the GPR approach worked well in the case of the “impurities” dataset but gave poor results with the “overlayers” values (due to overfitting only of the given training set), even though GPR achieved zero training errors in both cases. These results suggest that the incorrect choice of hyperparameters for GPR can significantly affect the performance of this method, while controlling the method simply by cross-validation could be difficult in real-world scenarios. Conversely, we observed that the tree-ensemble-based methods such as GBR, ET, and RFR were not greatly affected by the hyperparameter choices, so it is relatively simple to understand what is controlled by each hyperparameter. Therefore, the use of highly adaptive methods such as GPR, SVR, and KRR requires careful tuning and could be difficult in practice given that the d-band centers of the “impurities” and “overlayers” datasets are highly correlated (with a correlation coefficient of 0.948), as discussed in

Sect. 3.3.2. The ensemble approaches used in the GBR, ET, and RFR methods could mitigate possible large deviances in predictions by stabilizing the prediction variances.

### 3.3.6 The Importance of Descriptors to GBR Predictions

We subsequently investigated the relevance or redundancy of each of the 18 descriptors for the host and guest metals in Table 3.3 that had been used in the GBR model. GBR is based on an ML ensemble technique referred to as “boosting”. This technique adaptively combines large numbers of relatively simple regression-tree models that recursively partition the data using a single selected descriptor. Thus, it provides a feature importance score for each descriptor: a weighted average of the number of times (or the extent of contribution to the entire prediction) the descriptor is selected for partitioning. This score can be used to assess the relative importance

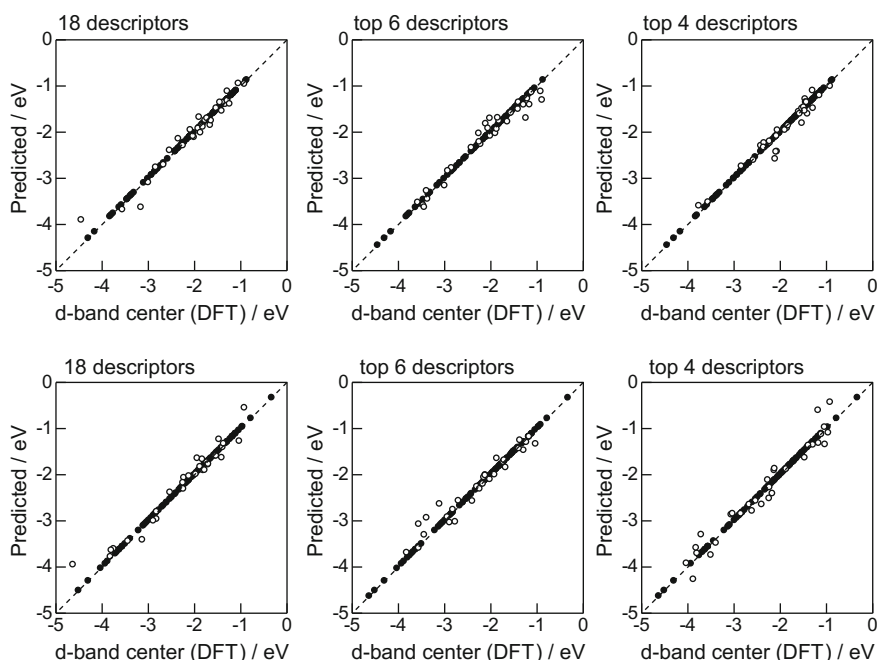


**Fig. 3.8** Feature importance scores of the descriptors for the GBR prediction of the d-band centers using the “impurities” (upper) and “overlayers” (lower) datasets. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

of that descriptor with respect to the predictability of the d-band center values. Note that these values are only employed with GBR, and the statistical importance of descriptors varies with the ML method used.

Figure 3.8 shows the feature importance scores of all 18 descriptors with regard to predicting the “impurities” and “overlayers” datasets. The six most important descriptors are highlighted, with a rank next to their bars. In the case of the “impurities,” these were (1) the group in the periodic table in which the host metal is found, (2) the density of the host metal at 25 °C, (3) the guest metal enthalpy of fusion, (4) the guest metal ionization energy, (5) the host metal enthalpy of fusion, and (6) the host metal ionization energy. In contrast, the most important factors for the “overlayers” dataset were (1) the host metal group in the periodic table, (2) the bulk Wigner–Seitz radius of the host metal, (3) the guest metal enthalpy of fusion, (4) the host metal density at 25 °C, (5) the guest metal ionization energy, and (6) the guest metal density at 25 °C.

To evaluate the effect of the number of descriptors on the predictive performance of GBR, the prediction results with 18 (that is, all), the top six and the top four descriptors are compared in Fig. 3.9. For quantitative evaluations, we also repeated the tests 100 times with random splits and calculated the mean RMSEs for these



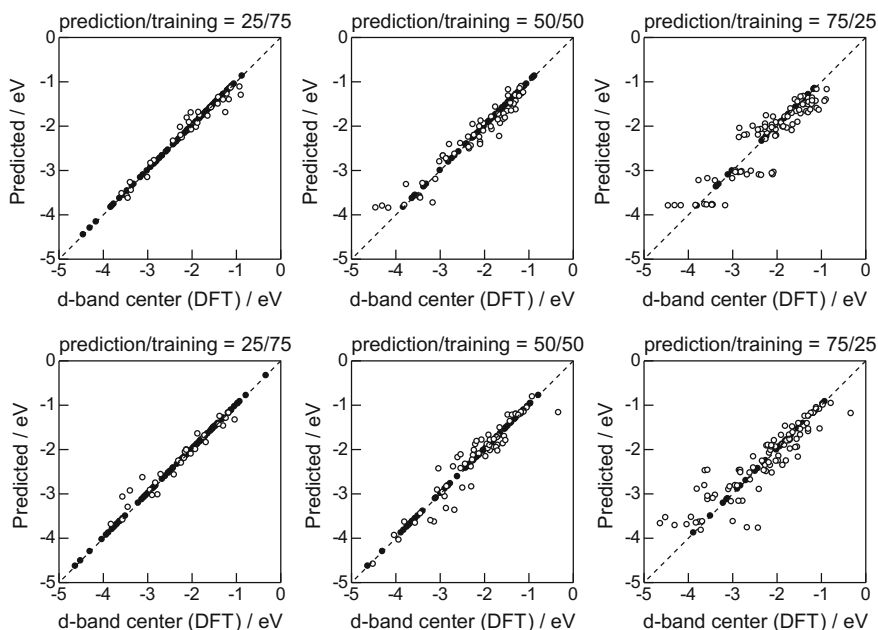
**Fig. 3.9** DFT-calculated local d-band center values for the “impurities” (upper) and “overlayers” (lower) datasets correlated with the values predicted by GBR with 18 (all), the top six and the top four descriptors. Legend: (●) training set = 75%, (○) test set = 25%. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry



predictions. The resultant values for the “impurities” dataset were  $0.17 \pm 0.04$ ,  $0.18 \pm 0.04$ , and  $0.16 \pm 0.04$  eV for 18, 6 and, 4 descriptors, respectively. For the “overlayers” dataset, the respective values were  $0.19 \pm 0.04$ ,  $0.19 \pm 0.04$ , and  $0.23 \pm 0.05$  eV. These data demonstrate that the ML prediction performance remained moderately good even when employing only four descriptors. Furthermore, it was found that the prediction accuracy obtained with six descriptors was superior to that with four descriptors when the test set proportion was increased above 25%. Based on these results, we used the GBR model with the top six descriptors for the subsequent analysis.

### 3.3.7 Model Estimations Using Different Test/Training Splits

Finally, we attempted to determine the size of training set required for ML to achieve sufficient prediction performance. Figure 3.10 shows the predictive performances using GBR with the top six descriptors when employing various test/training set ratios (25%/75%, 50%/50%, and 75%/25%). That is, we withheld 25%,



**Fig. 3.10** DFT-calculated local d-band center values for the “impurities” dataset (upper) and the “overlayers” dataset (lower) and the values predicted by GBR with the top six descriptors. Legend: (●) training set = 75%, (○) test set = 25%. Reproduced from Ref. [19] with permission from the Royal Society of Chemistry

50%, or 75% of the “impurities” and “overlayers” datasets as the test sets, and predicted these using GBR based on the remaining values. In Table 3.1 (“impurities”) and Table 3.2 (“overlayers”), we have 121 values in total, and 25%/75% corresponds to sets with size 30/91, 50%/50% to 61/60, and 25%/75% to 90/31. For quantitative evaluation, we also calculated the mean RMSEs for the 100 random splits for each setting. The resultant values for the “impurities” dataset were  $0.18 \pm 0.04$ ,  $0.23 \pm 0.05$ , and  $0.38 \pm 0.07$  eV for the 25%/75%, 50%/50%, and 75%/25% tests. In the case of the “overlayers” dataset, these values were  $0.19 \pm 0.04$ ,  $0.27 \pm 0.05$ , and  $0.41 \pm 0.08$  eV. These results quantitatively exhibit a general trend of ML such that a greater quantity of data generates better results and also show that d-band center values can be predicted with a moderate level of accuracy (RMSE =  $0.38 \pm 0.07$  eV for the “impurities” set, RMSE =  $0.41 \pm 0.08$  eV for “overlayers”), even when only 25% of the data are available and 75% are missing. This result provides a useful guideline for the trade-off between the predictive performance and data availability.

### 3.4 Conclusion and Future Prospects

The d-band center is one of the most important activity-controlling factors in heterogeneous metal catalysts. The work reported herein demonstrates that the values for monometallic (Fe, Co, Ni, Cu, Ru, Rh, Pd, Ag, Ir, Pt, and Au) and bimetallic surfaces having two different structures (surface impurities and overlayers on clean metal surfaces) can be predicted reasonably well using an ML method (the GBR method) in conjunction with six readily available descriptors. This ML-based prediction of the d-band centers requires a minimal amount of CPU time compared to first-principles DFT calculations. Our results demonstrate the potential to use ML methods in the design of catalysts and the possibility of catalyst development without extensive trial-and-error experimental testing.

Predictions of DFT-calculated, activity-controlling factors such as d-band centers, and reactant gas adsorption energy values by data-driven ML techniques have the potential to support the rapid discovery of specific catalytic materials in the near future. In addition, calculating the “activity” of a material (that is, the reaction rate whose dominant term is the activation energy) directly in place of activity-controlling factors will make the identification of optimal catalysts much easier and faster. Unfortunately, the description of the transition states of multistep reaction processes in heterogeneous catalysis is still at the leading edge of work in the field of computational quantum chemistry, due to the difficulties arising from modeling large collections of atoms involving numerous degrees of freedom and many electrons. However, we are hopeful that this challenge will be overcome in the future, thus allowing the direct and rapid prediction of activity trends with the aid of ML techniques.

**Acknowledgements** This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas “Nano Informatics” (Grant No. 25106010) from the Japan Society for the Promotion of Science (JSPS).

## References

1. G. Ertl, H. Knözinger, F. Schüth, J. Weitkamp, *Handbook of Heterogeneous Catalysis*, vol. 1, 2nd edn. (Wiley–VCH, Weinheim, 2008)
2. G.A. Somorjai, Y. Li, *Introduction to Surface Chemistry and Catalysis*, 2nd edn. (Wiley, 2010)
3. B. Hammer, J.K. Nørskov, *Adv. Catal.* **45**, 71 (2000)
4. V. Pallassana, M. Neurock, L.B. Hansen, B. Hammer, J.K. Nørskov, *Phys. Rev. B* **60**, 6146 (1999)
5. T. Bligaard, J.K. Nørskov, S. Dahl, J. Matthiesen, C.H. Christensen, J. Sehested, *J. Catal.* **224**, 206 (2004)
6. J.K. Nørskov, T. Bligaard, J. Rossmeisl, C.H. Christensen, *Nat. Chem.* **1**, 37 (2009)
7. Z.W. Seh, J. Kibsgaard, C.F. Dickens, I. Chorkendorff, J.K. Nørskov, T.F. Jaramillo, *Science* **355**, 146 (2017)
8. F. Calle-Vallejo, D. Loffreda, M.T.M. Koper, P. Sautet, *Nat. Chem.* **7**, 403 (2015)
9. E.-J. Ras, M.J. Louwerse, M.C. Mittelmeijer-Hazeleger, G. Rothenberg, *Phys. Chem. Chem. Phys.* **15**, 4436 (2013)
10. X. Ma, Z. Li, L.E.K. Achenie, H. Xin, *J. Phys. Chem. C* **6**, 3528 (2015)
11. Z. Li, X. Ma, H. Xin, *Catal. Today* **280**, 232 (2017)
12. Y. Okamoto, *Chem. Phys. Lett.* **395**, 279 (2004)
13. C. Lu, I.C. Lee, R.I. Masel, A. Wieckowski, C. Rice, *J. Phys. Chem. A* **106**, 3084 (2002)
14. E. Toyoda, R. Jinnouchi, T. Hatanaka, Y. Morimoto, K. Mitsuhara, A. Visikovskiy, Y. Kido, *J. Phys. Chem. C* **115**, 21236 (2011)
15. T. Anniyev, S. Kaya, S. Rajasekaran, H. Ogasawara, D. Nordlund, A. Nilsson, *Angew. Chem. Int. Ed.* **51**, 7724 (2012)
16. S. Furukawa, K. Ehara, K. Ozawa, T. Komatsu, *Phys. Chem. Chem. Phys.* **16**, 19828 (2014)
17. M. Tamura, K. Kon, A. Satsuma, K.I. Shimizu, *ACS Catal.* **2**, 1904 (2012)
18. S.M.A.H. Siddiki, A.S. Touchy, K. Kon, T. Toyao, K. Shimizu, *ChemCatChem* **2017**(9), 2816 (2017)
19. I. Takigawa, K. Shimizu, K. Tsuda, S. Takakusagi, *RSC Adv.* **6**, 52587 (2016)
20. A. Ruban, B. Hammer, P. Stoltze, H.L. Skriver, J.K. Nørskov, *J. Mol. Catal. A* **115**, 421 (1997)
21. A. Vojvodic, J.K. Nørskov, F. Abild-Pedersen, *Top. Catal.* **57**, 25 (2014)
22. *CRC Handbook of Chemistry and Physics*, 83rd edn., edited by D.R. Lide (CRC Press, London, 2002)
23. R.R. Picard, R.D. Cook, *J. Am. Stat. Assoc.* **79**(387), 575 (1984)
24. J. Shao, *J. Am. Stat. Assoc.* **88**(422), 486 (1993)
25. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **12**, 2825 (2011)
26. N. Japkowicz, M. Shah, *Evaluating Learning Algorithms: A Classification Perspective* (Cambridge University Press, 2011)
27. K. Murphy, *Machine Learning: A Probabilistic Perspective* (The MIT Press, 2012)
28. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2013)

29. J. Friedman, *Ann. Statist.* **29**(5), 1189 (2001)
30. T. Chen, T. He, *JMLR. Workshop Conf. Proc.* **42**, 69 (2015)
31. T. Chen, C. Guestrin (2016), [arXiv:1603.02754](https://arxiv.org/abs/1603.02754)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

