Chapter 8
Optimal Dopant Selection for Water Splitting with Cerium Oxides: Mining and Screening First Principles Data

V. Botu, A.B. Mhadeshwar, S.L. Suib and R. Ramprasad

Abstract We propose a powerful screening procedure, based on first principles computations and data analysis, to systematically identify suitable dopants in an oxide for the thermochemical water splitting process. The screening criteria are inspired by Sabatier's principle, and are based on requirements placed on the thermodynamics of the elementary steps. Ceria was chosen as the parent oxide. Among the 33 dopants across the periodic table considered, Sc, Cr, Y, Zr, Pd and La are identified to be the most promising ones. Experimental evidence exists for the enhanced activity of ceria for water splitting when doped with Sc, Cr and Zr. The surface oxygen vacancy formation energy is revealed as the primary descriptor correlating with enhanced water splitting performance, while the dopant oxidation state in turn primarily governs the surface oxygen vacancy formation energy. The proposed screening strategy can be readily extended for dopant selection in other oxides for different chemical conversion processes (e.g., CO_2 splitting, chemical looping, etc.).

V. Botu
Department of Chemical and Biomolecular Engineering, University of Connecticut, Storrs, CT 06269, USA
e-mail: venkatesh.botu@uconn.edu

A.B. Mhadeshwar
Center for Clean Energy and Engineering, University of Connecticut, Storrs, CT 06269, USA

A.B. Mhadeshwar
Present Address: ExxonMobil Research and Engineering, Annandale, NJ 08801, USA

S.L. Suib
Department of Chemistry, University of Connecticut, Storrs, CT 06269, USA

S.L. Suib · R. Ramprasad (✉)
Institute of Materials Science, University of Connecticut, Storrs, CT 06269, USA
e-mail: rampi@ims.uconn.edu

R. Ramprasad
Department of Materials Science and Engineering, University of Connecticut, Storrs, CT 06269, USA
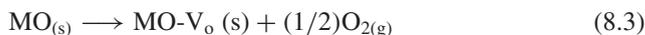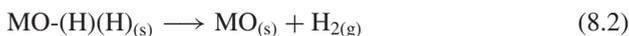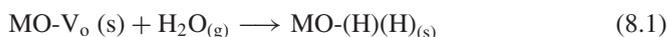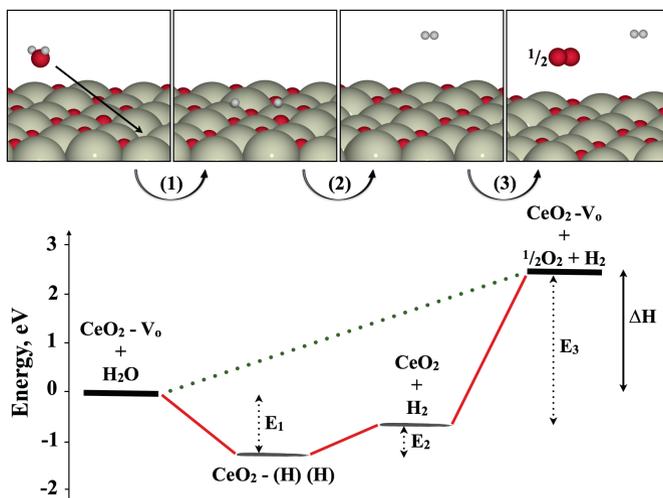
8.1 Introduction

Utilizing dopants to optimize, enhance, or fundamentally change the behavior of a parent material has been exploited in many situations ranging from material strengthening to electronics to electrochemistry. The search and identification of suitable dopant candidates has been laborious though, and dominated either by lengthy trial-and-error strategies (guided by intuition) or plain serendipity. We are entering an era where such Edisonian approaches are gradually being augmented (and sometimes, replaced) by rational strategies based on advanced computational screening [1]. Often these strategies rely on first principles methods, that provide a reasonably accurate description of the underlying chemistry [2–4]. More recently, it has been shown that supplementing first principles investigations with data-driven approaches can help identify meaningful correlations within the data [5–13]. In the present contribution, we offer such a prescription for the selection of suitable dopants within cerium oxides in order to enhance the thermochemical splitting of water.

Complete gas phase thermolysis of water is highly endothermic ($\Delta H = +2.53\,\text{eV}$) requiring temperatures in excess of $4000\,\text{K}$ to be thermodynamically favorable, making such reactions unviable for H_2 synthesis [14, 15]. On the other hand, partial thermolysis via a multistep process in the presence of MO catalysts provides an attractive practical alternative [15, 16]. The latter approach is performed at two distinct temperatures (both well below $4000\,\text{K}$): a high-temperature ($\approx 2200\,\text{K}$) reduction step that involves creation of O vacancies in the MO (and the consequent evolution of O_2 gas), and lower-temperature ($\approx 900\,\text{K}$) oxidation steps in the presence of steam, which lead to the filling up of O vacancy centers (resulting in the evolution of H_2 gas). Owing to this multistep procedure, an additional step to separate the H_2 and O_2 products is eliminated entirely. Equations (8.1)–(8.3) below represent a reordered version (for ease of subsequent discussion) of the multiple steps involved in this process.

$$\text{MO-V}_o \text{ (s)} + H_2O_{(g)} \longrightarrow \text{MO-(H)(H)}_{(s)} \tag{8.1}$$

$$\text{MO-(H)(H)}_{(s)} \longrightarrow \text{MO}_{(s)} + H_{2(g)} \tag{8.2}$$

$$\text{MO}_{(s)} \longrightarrow \text{MO-V}_o \text{ (s)} + (1/2)O_{2(g)} \tag{8.3}$$

The (s) and (g) subscripts represent solid and gas phases, respectively. Equations (8.1) and (8.2) are the low-temperature steps, with MO-V_o and MO-(H)(H) representing, respectively, the oxide containing an O vacancy and the oxide in which the O vacancy is filled up by a H_2O molecule (with '(H)(H)' indicating that the H atoms of H_2O are adsorbed on the oxide surface). Equation (8.3) is the high-temperature activation step that leads to the creation of MO-V_o.

Unfortunately, several MOs require temperatures in excess of $2700\,\text{K}$ (leading to poor H_2 production efficiencies), leaving only a subset of oxides based on Zn, Fe and Ce to be the most promising [17, 18]. Oxides of Zn and Fe are prone to sintering, phase transformation or volatility due to the proximity of the high temperature step

Fig. 8.1 Reaction pathway and energetics (*red solid line*) for the dissociation of H_2O on an undoped ceria surface. CeO_2-V_o is an oxide with a vacancy, CeO_2-(H)(H) is an oxide with vacancy filled by a H_2O molecule and CeO_2 is a stoichiometric surface. The *green dotted line* shows the minimum energy pathway for dissociation. Ce, O and H are represented by beige, *red* and *white* colors respectively

to their melting points [19]. CeO_2, on the other hand, displays high stability and high melting temperature ($\approx 2600\,K$), and is thus overwhelmingly favored [17].

Still, the efficiency of H_2 production with CeO_2 is quite low ($<1\,\%$) [18]. This low efficiency is rooted in the high temperatures ($>1900\,K$) required for the reduction step (8.3), related directly to the large O vacancy formation energy of CeO_2, along with other operational difficulties [18, 20]. Figure 8.1 shows the energies E_1, E_2 and E_3 of (8.1), (8.2) and (8.3), respectively, computed here using density functional theory (DFT) (details below), and helps identify the causes of the low efficiency. The dotted line indicates the uphill nature of the water splitting process. The ideal system should display E_1 and E_2 close to zero (for facile H_2 evolution at low temperatures), and small E_3 values (to alleviate the burden on the reduction step). In the case of CeO_2, E_1 is too negative and E_3 is too positive.

A pathway to circumvent these hurdles is to control the energetics of (8.1)–(8.3) individually by the introduction of dopants (although, of course, the overall energetics of H_2O splitting cannot be altered). For instance, this strategy may be used to destabilize O in CeO_2 (and thus reduce the O vacancy formation energy) [17, 21–27]. Doping CeO_2 with a plethora of elements has been explored in the recent past [28–40], and many dopants (e.g., Zr, Cr, Sc) have been shown to help significantly increase the efficiency of H_2 production by reducing the temperatures required to accomplish (8.3) [32, 34, 35]. *Nevertheless, a clear rationale for why a given dopant is desirable, and a framework for the systematic (non-Edisonian) selection of dopants is currently unavailable.* This work attempts to fill that gap. First, we propose a framework to systematically screen for dopants, based on guidelines inspired by Sabatier's

principle, then we identify the best candidates using first principles methods, and finally we use data analysis methods, specifically feature selection, to identify the primary factors that make these dopants attractive.

8.2 Screening Framework

In the present first principles/data-driven based work, we consider a host of dopants in CeO$_2$, including 33 elements spanning the 4th, 5th and 6th period of the Periodic Table (specifically the *alkali*, *alkaline earth* and *d* series elements). Assuming that the energetics of (8.1)–(8.3) determine whether a dopant is favorable or not, we define the following screening criteria to be used in a successive manner:

- *Criterion 1:* $0 \leq E_3^D \leq E_3$
- *Criterion 2:* $0 \leq E_1^D \leq \delta$
- *Criterion 3:* $0 \leq E_1^D + E_2^D \leq \delta$

The superscripts D merely indicate that these are the energetics of doped ceria.

The rationale underlying this specific choice and sequence of screening criteria stems from insights derived from Sabatier's principle, and may be understood as follows (cf. Fig. 8.1). *Criterion 1* merely states that the O vacancy formation energy (which is what E_3^D represents) should not be too small to prevent further water dissociation nor too large (certainly not larger than that of undoped ceria (E_3)) to mandate higher activation temperatures. This criterion is listed first because E_3^D appears to most strongly control the temperature requirement of the costly high-temperature step, and also because E_3^D is the easiest quantity to compute (as it does not involve the H$_2$O species at all). *Criterion 2* states that E_1^D should also be bracketed, but by a smaller range. Noting that overall dissociation of water for undoped ceria is too negative (see Fig. 8.1), thus potentially adding an energy penalty to subsequent steps, we generously allow δ to be 1.5 eV, which is a reasonable choice considering energy uncertainties within DFT and the neglection of entropy. *Criterion 3* is specific to thermochemical water splitting and bounds the overall oxidation process within δ, ensuring that E_1^D or E_2^D occur at a lower temperature compared to E_3^D. In the case where this no longer holds, the process fails to fall within the realm of thermochemical water splitting.

8.3 First Principles Studies

8.3.1 Methods and Models

To measure the thermodynamic quantity, E_i^D, where i is 8.1, 8.2 or 8.3, DFT calculations were performed using the VASP code with the semi-local Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional and a cutoff energy of 400 eV to

accurately treat the valence O 2s, 2p and Ce 5s, 5p, 4f, 5d, 6s states [41–43]. The electron-core interactions were captured by projector-augmented (PAW) potentials, and all calculations were spin polarized to ensure the true electronic state of O and reduced Ce was captured [44]. The computed lattice parameter of bulk CeO_2 (5.47 Å) is in good agreement with the corresponding experimental value (5.41 Å) [38]. A 96-atom bulk $2 \times 2 \times 2$ supercell model and a 60-atom (2×2) surface model (5 O-Ce-O trilayers) cleaved along the (111) plane were used in all calculations. The bottom 3 trilayers of the slab were fixed to recover the bulk nature of the material, and a vacuum of 15 Å along the c axis ensured minimal spurious interactions between periodic images. A Γ-centered k-point mesh of $3 \times 3 \times 3$ and $3 \times 3 \times 1$ were used for the bulk and surface calculations, respectively. The Hubbard (U) correction was not applied as no universal U value captures the true electronic state of all elements. Also, given that we consider a dilute vacancy limit, the effect of electron localization is insignificant as shown previously [45, 46].
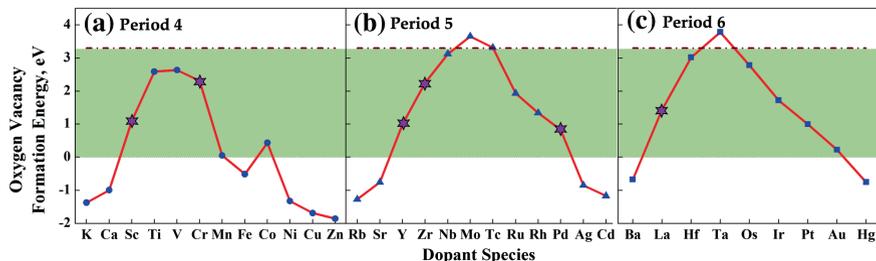
8.3.2 Enforcing the 3-Step Criteria

Dopants were introduced by replacing a single Ce atom at the center of the bulk model and at the 1st trilayer of the surface model. Our analysis indicated that the majority of the dopants favored the surface site to the bulk by ≈ 0.3 eV. Upon exploring the local coordination environment, a surface dopant was found to be 6-fold coordinated whereas a bulk dopant was 8-fold coordinated. Given the preference of a surface site, all dopants are assumed to occupy the surface unless specified otherwise.

The primary effect of introducing dopants is to induce a local perturbation to disrupt bonding between the metallic and O atoms, thereby altering its ability to form surface O vacancies, as measured by E_3^D (cf. Fig. 8.1), computed here as

$$E_3^D = E_{CeO_2\text{-}V_o}^D - E_{CeO_2}^D + \frac{1}{2}\mu_{O_2} \qquad (8.4)$$

where $E_{CeO_2\text{-}V_o}^D$ and $E_{CeO_2}^D$ are, respectively, the DFT energies of a doped surface with and without an O vacancy, and μ_{O_2} is the chemical potential of O, taken here to be the DFT energy of an isolated O_2 molecule. In all cases, the O vacancy is created adjacent to the dopant. Figure 8.2 shows E_3^D for various choices of the dopants, with the dot-dashed horizontal line indicating the corresponding value for the undoped case. Dopants adopting a low valence state compared to Ce (e.g., alkali, alkaline earth and late transition series metals) display low O vacancy formation energy, consistent with the observed high O_2 yield by ceria doped with Mn, Fe, Ni and Cu [47]. Conversely, dopants adopting a similar or higher valence state than Ce lead to high E_3^D values (e.g., Mo, Tc, and Ta). These trends are not entirely surprising, and have been noted before in CeO_2 as well as $BaTiO_3$ [48–50].

Fig. 8.2 Oxygen vacancy formation energy (E_3^D) of doped ceria with elements from the (**a**) 4th, (**b**) 5th and (**c**) 6th period of the Periodic Table. *Dot-dashed maroon line* indicates E_3^D for undoped ceria. *Light green region* indicates dopants that survived *Criterion 1*, while ★ identifies dopants that survived the 3 screening criteria
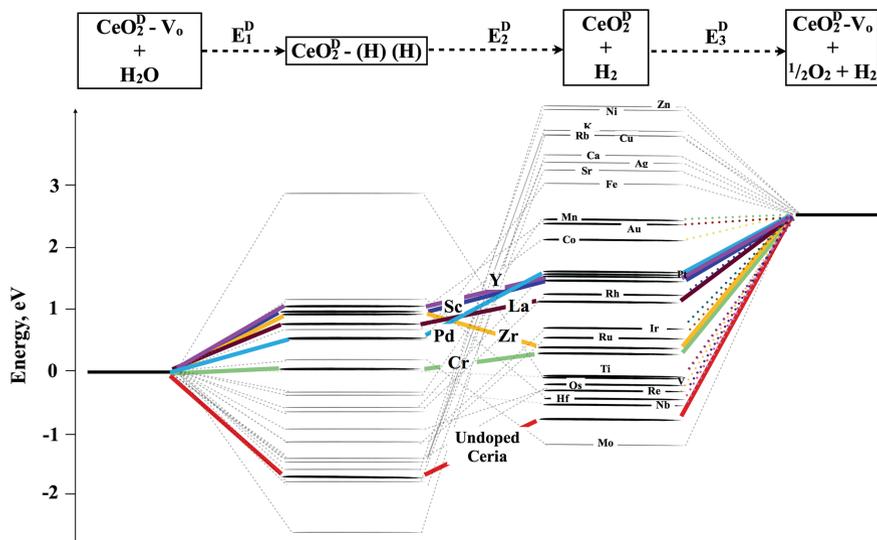
E_1^D helps assess the impact of dopants on the dissociative adsorption of water on the doped surface, and is computed as

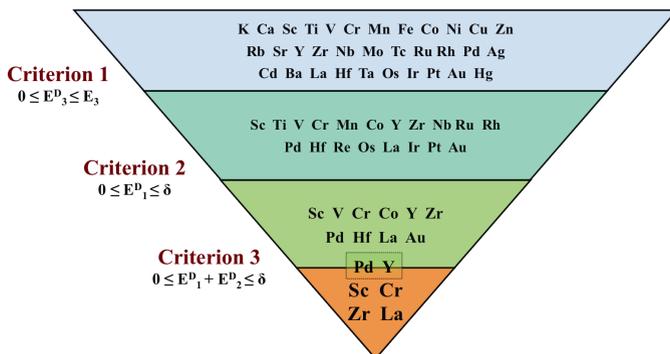$$E_1^D = E_{CeO_2\text{-}(H)(H)}^D - E_{CeO_2\text{-}V_o}^D - \mu_{H_2O} \tag{8.5}$$

where $E_{CeO_2\text{-}(H)(H)}^D$ is the DFT energy of a doped surface upon the dissociative adsorption of water at the vacancy site. Upon dissociation, OH fills the vacancy site, while H has two possible adsorption sites; atop an adjacent O or a dopant atom. Interestingly, dopants exhibiting spontaneous vacancy formation ($E_3^D < 0\,eV$) fail to accommodate a H atop a dopant, while those dopants that do facilitate H atop a dopant have an alternative lower energy pathway for dissociation. μ_{H_2O} is the chemical potential of water, taken here to be the DFT energy of an isolated H_2O molecule.

With E_1^D and E_3^D at hand (and E_2^D given by $\Delta H - E_1^D - E_3^D$), a plot that is equivalent to Fig. 8.1 but for the case of doped ceria surfaces is shown in Fig. 8.3. We now enforce *Criterion 1*, namely, $0 \le E_3^D \le E_3$, with $E_3 = 3.3\,eV$ (this value is consistent with past work [45]). Of the 33 dopants originally considered, 19 dopants (Sc, Ti, V, Cr, Mn, Co, Y, Zr, Nb, Ru, Rh, Pd, La, Hf, Re, Os, Ir, Pt and Au) satisfy this criterion (given by the dopants within the shaded region in Fig. 8.2). *Criterion 1* picks out those dopants that alter the surface reducibility in just the appropriate manner.

Next, we enforce *Criterion 2*, namely, $0 \le E_1^D \le \delta$, with $\delta = 1.5\,eV$, on the 19 dopants that pass *Criterion 1*, resulting in the selection of Sc, V, Cr, Co, Y, Zr, Pd, La, Hf and Au. Lastly, enforcing *Criterion 3* on the 10 dopants results in the down selection of 4 promising candidates (Sc, Cr, Zr and La). Inspection of Fig. 8.3 shows that Pd and Y, although they do not pass *Criterion 3*, can be viewed as 'near misses'. These are hence included in our final list of favored candidates. Figure 8.4 summarizes the list of dopants that passed each stage of the screening process. The 6 dopants identified, namely, Sc, Cr, Zr, La, Pd and Y, lead to desired energetic profiles, with E_1^D and E_2^D low enough to allow for reasonable H_2O dissociation yields at moderate temperatures, and with E_3^D significantly smaller than undoped ceria allowing for low reduction temperatures (c.f., Fig. 8.3). Dopants such as Mn, Fe, Ni, Cu, Sr, Ag, and

Fig. 8.3 Reaction pathway and energetics for the multistep thermochemical splitting of H_2O on a doped ceria surface. CeO_2^D-V_o is a doped surface with vacancy, CeO_2^D-(H)(H) is a doped surface with vacancy filled by a H_2O molecule and CeO_2^D is a doped stoichiometric surface. Color *solid lines* identify the 4 promising dopants and undoped CeO_2. *Grey dashed lines* identifies the non feasible dopants, while partly colored and *greyed dashed lines* identifies dopants that pass Criterion 1

Fig. 8.4 A hierarchical chart showing the list of dopants before and after each stage of the screening process. Sc, Cr, Zr and La were identified as the promising dopant elements, whilst Pd and Y can be viewed as the near miss cases

Ca, which display small or negative E_3^D, do not pass our tests. Although low E_3^D values imply facile surface reduction (this is in fact what is observed experimentally for Mn and Fe) [47], such a tendency would not be appropriate for the multistep thermochemical water splitting process targeted here (lower yields were observed for Ni, Cu and Fe doped CeO_2 compared to undoped CeO_2) [28]. *Criterion 1*, as mentioned above, is imposed precisely to eliminate such candidates. However, dopants that lead

to small or negative E_3^D may be appropriate for photocatalytic water splitting which require surface reduction to occur low temperatures (≈ 300 K) [51].

Of the 6 promising dopants identified, experimental evidence exists for the enhanced performance of ceria when doped with Sc, Cr and Zr for the thermo-chemical water splitting process. Cr doped CeO_2 is known to lower the reduction and oxidation temperature to 750 and 350 K, respectively [35]. Zr and Sc dopants increase the H_2 yield 4-fold and almost 2-fold, respectively, with respect to the undoped situation [28, 29, 38]. Lastly, although not conclusive, La doping appears to improve H_2 yield [39, 52]. The observed performances are strong functions of the synthesis, processing and measurement details. The present work ignores such complexities, and probes only the dominant and primary chemical factors that may control performance.

Irrespective of these difficulties, such a guided screening strategy has led us to some promising candidates, shown as stars in Fig. 8.2. Clearly, the best candidates display an O vacancy formation energy in the 1–2.5 eV range, i.e., neither too high nor too low, thereby respecting Sabatier's principle. It thus appears that the O vacancy formation energy may be used as a 'descriptor' of the activity of doped ceria. This conclusion is consistent with an earlier similar proposal which was based on phase boundaries in surface phase diagrams of ceria exposed to an oxygen reservoir [45].

Thus far, by relying on first principles methods we are able to recognize whether a dopant increases or decreases the O vacancy formation energy, with respect to the undoped material, followed by its corresponding impact on the dissociation of water. However, an understanding of the complex dependence of the chemical attributes of a dopant and the O vacancy formation energy is absent. In the next section, with the help of data analysis methods we attempt to understand the results of the first principles computations for the spectrum of dopants considered.

8.4 Data Analysis

The mining and extraction of information forms the core of the field of data analysis, which lies under a broader umbrella of methods known as machine learning (ML) [53]. Within data analysis a subset of methods, known as feature selection, allows us to unearth correlations between variables [10, 13, 53–56]. In the context of this work, the variables are the chemical factors characterizing a dopant and the corresponding O vacancy formation energy of doped ceria. Given the strong correlation between the O vacancy formation energy and the activity, as discussed above, by identifying the key dopant factors that contribute to the O vacancy formation energy, a more educated guess on its impact on the corresponding thermodynamic activity can be made.

In order to discover such patterns, firstly, each dopant element needs to be represented numerically by a vector of numbers (also referred to as features or fingerprint in the ML community) that uniquely identifies the dopant element. Our choice of features stems from fundamental chemical factors, that are often used to describe
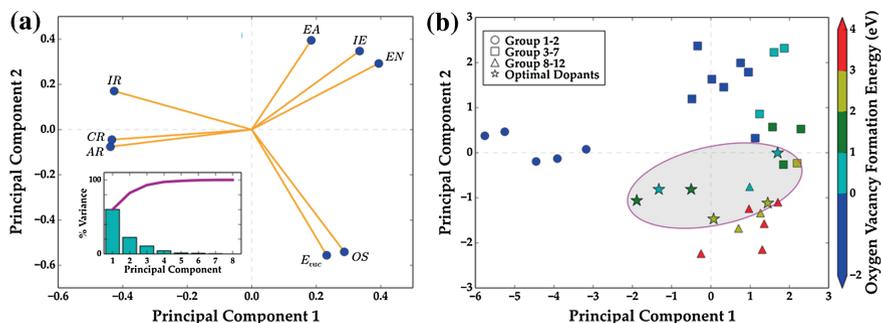
elements in the periodic table. The 7 factors considered in this work are; atomic radius (AR), ionic radius (IR), covalent radius (CR), ionization energy (IE), electronegativity (EN), electron affinity (EA) and oxidation state (OS). To eliminate any bias induced by the spread of the feature values, the dataset was normalized to a mean of 0 and variance of 1. On these set of chemical factors we use two feature selection methods: (i) principal component analysis and (ii) random forests, to narrow down the dominant factors that govern the descriptor (O vacancy formation energy). In the sections to follow we provide a brief overview of these methods and discuss the insights gained. We refer the readers to [53, 57–61] for a more exhaustive description. The data analysis routines used were implemented within the MATLAB statistical toolkit and Scikit-learn python module [62, 63].

8.4.1 Principal Component Analysis

Principal component analysis (PCA) is a common dimensionality reduction technique, often used to identify the dominant subset of features from a larger pool. By transforming the original features into uncorrelated and orthogonal pseudo variables, that are a linear combination of the original features (as done in this work, although non-linear combinations have been recently developed), it allows us to pin point the dominant contributions [10, 55–58]. The new transformed variables are referred to as *principal components* (PCs), which are solutions to the eigen-transformation of the covariance matrix. As with any eigen-transformation problem, the eigenvalues and eigenvectors play a critical role. The eigenvalue of a PC indicates the % of variance captured within the original dataset, whilst the eigenvector provides the coefficients that dictate the linear transformation. We shall make use of this information to down select the dominant chemical factors of a dopant.

First, we plot the transformation coefficient values of the 7 features for the first and second PCs in Fig. 8.5a. Such a plot is referred to as the loadings plot, in which correlated features cluster together. Only the first and second PCs are used as it captures ≈80 % of the variance within the original dataset (c.f., inset of Fig. 8.5a). Clearly, the dopant's OS is strongly correlated with the O vacancy formation energy. The CR, AR, IE and EN are close to orthogonal to the O vacancy formation energy, suggesting a negligible contribution to the descriptor. On the other hand, the IR and EA are not truly orthogonal, thus their contribution towards the descriptor cannot be ignored. Another interesting phenomena is the congregation of subsets of the 7 features. This isn't entirely surprising, as one would recognize that the AR, CR are similar quantities, and their grouping in the loadings plot further validates this notion. Similarly, the IE and EN group together and appear negatively correlated to the AR and CR, given their ≈180° separation. By looking at the relative position of all the features in Fig. 8.5a, we can conclude that of the original 7 features considered only 3 are important; OS, IR and EA, in governing the O vacancy formation energy.

Next, we use the linear transformation coefficients of the PCs to transform the original dopant dataset (also referred to as the scores plot) and plot the first and
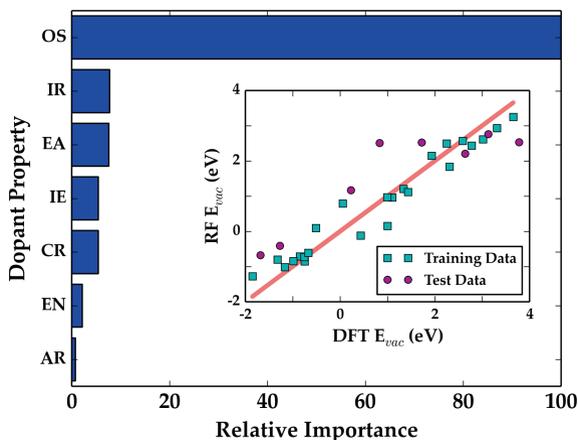
Fig. 8.5 **a** PCA loadings plot showing the correlated dopant features. The features are; atomic radius (AR), ionic radius (IR), covalent radius (CR), ionization energy (IE), electronegativity (EN), electron affinity (EA) and oxidation state (OS). E_{vac} is the O vacancy formation energy. The inset shows the % contribution of each PC to the variance in the dataset. The oxidation state (OS) is the dominant feature governing the O vacancy formation energy. **b** PCA scores plot for the first and second principal components. The dopant elements group together based on their features and the O vacancy formation energy. ⋆ represents the final 6 dopants after the 3 step screening processes. The 6 dopants occupy a sub-space of the scores plot as highlighted by the *grey region*

second PCs in Fig. 8.5b. Each dopant element in Fig. 8.5b has further been classified according to its relative location in the periodic table (as indicated by the different marker type) and the corresponding O vacancy formation energy (marker fill color). Firstly, dopants of similar type, groups 1–2, 3–7 and 8–12 can be seen aggregating together. In particular, dopants that adopt a low valence state lie predominantly in the top/left quadrants, whilst the high valence dopants lie in the bottom/right quadrants, giving rise to an increasing O vacancy formation energy in the direction of the bottom right quadrant. Not surprisingly, amongst the low valence dopants, the alkali and alkaline earth metals further segregate from the late transition series metals, based on their differences of atomic size, amongst others. Now, upon highlighting the location of the 6 promising candidates (Sc, Cr, Y, Zr, Pd and La), as indicated by the stars, they can be seen to occupy only a small subspace of the plot (highlighted by the grey region of Fig. 8.5b). This suggests that in the high dimensional transformation these elements have similar traits, and equivalentaly a similar thermodynamic activity. Therefore, if one could identify other possible dopants that populate the grey region in Fig. 8.5b, we can further extend the chemical space to achieve improved water dissociation.

8.4.2 Random Forest

Another important class of feature selection algorithms are random forests (RF). Unlike PCA, random forests work by constructing a regression (or classification) model first, in this case between the 7 features and the O vacancy formation energy,

Fig. 8.6 Relative feature importance arranged in descending order for the developed RF model. The features are; atomic radius (AR), ionic radius (IR), covalent radius (CR), ionization energy (IE), electronegativity (EN), electron affinity (EA) and oxidation state (OS). E_{vac} is the O vacancy formation energy. The inset shows a parity plot, comparing the density functional theory (DFT) and RF predicted O vacancy formation energy (E_{vac}). The regression model has an R^2 value of 0.94. The oxidation state (OS) is the dominant feature governing the O vacancy formation energy

following which the important features are then extracted as a by-product. The framework is built upon an ensemble of individual regression models, also known as decision trees [53, 59–61]. The prediction of each individual tree is then averaged across the ensemble, resulting in the final or true predicted value. Given our limited dataset size (based on 33 dopant elements), we selected a 75 % split for training, with the remaining kept aside as validation/testing. Each decision tree in the model is then trained on a subset of the original training dataset, a procedure known as bootstrapping. The combination of bootstrapping and ensemble averaging makes RF models robust and devoid of overfitting, a common issue in ML. We generate a forest of 250 trees, based on the 7 dopant features described earlier and the O vacancy formation energy. The final regression model we obtained has an R^2 value of 0.95 (c.f., inset Fig. 8.6), suggesting a good fit. Then by using mean decrease impurity metric, we estimate the relative importance of each feature in the regression model [61].

In Fig. 8.6, we plot the relative importance of the 7 features in descending order. Clearly, the role of a dopant's OS supersedes all others. This observation is consistent with the PCA analysis above. Also, it can be seen that IR and EA rank 2nd and 3rd in feature importance in the regression model, once again suggesting a small contribution towards the descriptor.

Both the PCA and RF methods result in similar conclusions, leading us to believe that the dopant's OS primarily governs the role of the descriptor, i.e., O vacancy formation energy, followed by a much smaller contribution of the IR and the EA. Upon revisiting the OS of the 6 promising dopants, they adopt either a +3 or +4 state. Therefore as a first measure, by understanding the coordination environment of

the dopant within the surface one can hazard a reasonable guess on its corresponding impact on the O vacancy formation energy. Even though many other elements such as Ti, V, Mn, Fe, Nb, Mo, Tc, Ru, Rh, Hf, Ta, Os, Ir adopt a similar OS state, the combination of the OS, IR and EA skews them out of the optimal regime.

8.5 Summary and Outlook

In this work, we considered a host of dopants in cerium oxide, that span the 4th, 5th and 6th period (specifically the *alkali*, *alkaline earth* and *d* series elements) of the Periodic Table, in order to understand the impact on the dissociation of water. Using a screening framework based on a first principles strategy augmented with data analysis methods, we successfully identified 6 promising dopants (Sc, Cr, Y, Zr, Pd and La), consistent with past experimental results, that are worthy of further inquiry. A dopant's oxidation state, ionic radius and electron affinity are found to be the dominant chemical factors that primarily govern the oxygen vacancy formation energy, which in turn governs the activity. The overall framework, we believe, can be easily extended for dopant selection in ceria and other oxides as well as for different chemical conversion processes (e.g., thermochemical CO_2 splitting, chemical looping, etc.).

Nevertheless, some open questions remain on the true measure of activity. First, kinetic factors, such as activation barriers, have been completely ignored in the present work. All the screening criteria were based on the thermodynamic requirements of the elementary steps, and serve as necessary but not sufficient conditions. Second, it is unclear what the impact of non-zero temperatures and gas phase component pressures would be on the computed quantities and final outcomes. Preliminary assessment based on first principles thermodynamics indicates that our main conclusions will be largely unchanged even when such factors are accounted for. However, by incorporating more of such metrics, along with the guidelines from the data analysis methods, we can systematically refine the screening framework.

References

1. G. Ceder, K. Persson, The stuff of dreams. Sci. Am. **309**, 36 (2013)
2. J. Neugebauer, T. Hickel, Density functional theory in materials science. Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**(5), 438–448 (2013)
3. G. Hautier, A. Jain, S.P. Ong, From the computer to the laboratory: materials discovery and design using first-principles calculations. J. Mater. Sci. **47**, 7317 (2012)

4. A.D. Becke, Perspective: fifty years of density-functional theory in chemical physics. J. Chem. Phys. **140**, 18A301 (2014)

5. T. Mueller, A.G. Kusne, R. Ramprasad, Machine learning in materials science: recent progress and emerging applications, in *Reviews in Computational Chemistry*, ed. by A.L. Parrill and K.B. Lipkowitz (Wiley, New York, 2016)

6. S. Srinivas, K. Rajan, property phase diagrams for compound semiconductors through data mining. Materials **6**, 279 (2013)

7. G. Hautier, C.C. Fisher, A. Jain, T. Mueller, G. Ceder, Finding natures missing ternary oxide compounds using machine learning and density functional theory. Chem. Mater. **22**, 3762 (2010)

8. C.C. Fischer, K.J. Tibbetts, D. Morgan, G. Ceder, Predicting crystal structure by merging data mining with quantum mechanics. Nat. Mat. **5**, 641 (2006)

9. X. Zhang, L. Yu, A. Zakutayev, A. Zunger, Sorting stable versus unstable hypothetical compounds: the case of multi-functional abx half-heusler filled tetrahedral structures. Adv. Funct. Mater. **22**, 1425 (2012)

10. P.V. Balachandran, S.R. Broderick, K. Rajan, Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning. Proc. R. Soc. A **467**, 2271 (2011)

11. E.W. Bucholtz, C.S. Kong, K.R. Marchman, W.G. Sawyer, S.R. Phillpot, S.B. Sinnot, K. Rajan, Data-driven model for estimation of friction coefficient via informatics methods. Tribol. Lett. **47**, 211 (2012)

12. I.E. Castelli, K.W. Jacobsen, Designing rules and probabilistic weighting for fast materials discovery in the perovskite structure. Model. Simul. Mater. Sci. Eng. **22**, 055007 (2014)

13. J. Carrete, W. Li, N. Mingo, S. Wang, S. Curtarolo, Finding unprecedentedly low-thermal-conductivity half-heusler semiconductors via high-throughput materials modeling. Phys. Rev. X **4**, 011019 (2014)

14. D.R. Hull, H. Prophet, Janaf thermochemical tables (2014), http://kinetics.nist.gov/janaf. Accessed 15 Jan 2014

15. S. Abanades, P. Charvin, G. Flamant, P. Neveu, Screening of water-splitting thermochemical cycles potentially attractive for hydrogen production by concentrated solar energy. Energy **31**, 2805 (2006)

16. T. Nakamura, Hydrogen production from water utilizing solar heat at high temperatures. Sol. Energy **19**, 467 (1977)

17. S. Abanades, G. Flamant, Solar hydrogen production from the thermal splitting of methane in a high temperature solar chemical reactor. Sol. Energy **80**, 1611 (2006)

18. L. D'Souza, Thermochemical hydrogen production from water using reducible oxide materials: a critical review. Mater. Renew. Sust. Energy **2**, 1 (2013)

19. W.C. Chueh, S.M. Haile, A thermochemical study of ceria: exploiting and old material for new modes of energy conversion and CO_2 mitigation. Philos. Trans. R. Soc. A **368**, 3269 (2010)

20. W.C. Chueh, S.M. Haile, Ceria as a thermochemical reaction medium for selectively generating syngas or methane from H_2O and CO_2. Chem. Sus. Chem. **2**, 735 (2009)

21. W.C. Chueh, C. Falter, M. Abbott, D. Scipio, P. Furler, S.M. Haile, A. Steinfeld, High-flux solar-driven thermochemical dissociation of CO_2 and H_2O using nonstoichiometric ceria. Science **330**, 1797 (2010)

22. A. Trovarelli, *Catalysis by Ceria and Related Materials* (World Scientific, London, 2002)

23. S. Kumar, P.K. Schelling, Density functional theory study of water adsorption at reduced and stoichiometric ceria (111) surfaces. J. Chem. Phys. **125**, 204704 (2006)

24. H.T. Chen, Y.M. Choi, M. Liu, M.C. Lin, A theoretical study of surface reduction mechanisms of CeO_2(111) and (110) by H_2. Chem. Phys. Chem. **8**, 849 (2007)

25. Z. Yang, Q. Wang, S. Wei, D. Ma, Q. Sun, The effect of environment on the reaction of water on the ceria(111) surface: a DFT+U study. J. Phys. Chem. C **114**, 14891 (2010)

26. M. Fronzi, S. Piccinin, B. Delley, E. Traversa, C. Stampfl, Water adsorption on the stoichiometric and reduced CeO_2 (111) surface: a first-principles investigation. Phys. Chem. Chem. Phys. **11**, 9188 (2009)

27. M. Molinari, S.C. Parker, D.C. Sayle, M.S. Islam, Water adsorption and its effect on the stability of low index stoichiometric and reduced surfaces of ceria. J. Phys. Chem. C **116**, 7073 (2012)

28. Q.L. Meng, C. Lee, T. Ishihara, H. Kaneko, Y. Tamaura, Reactivity of CeO_2-based ceramics for solar hydrogen production via a two-step water-splitting cycle with concentrated solar energy. Int. J. Hydrog. Energy **36**, 13435 (2011)

29. C. Lee, Q. Meng, H. Kaneko, Y. Tamaura, Solar hydrogen productivity of ceriascandia solid solution using two-step water-splitting cycle. J. Sol. Energy Eng. **1135**, 011062 (2013)

30. C. Lee, Q. Meng, H. Kaneko, Y. Tamaura, Dopant effect on hydrogen generation in two-step water splitting with CeO_2-ZrO_2MO_x reactive ceramics. Int. J. Hydrog. Energy **38**, 15934 (2013)

31. R. Bader, L.J. Venstrom, J.H. Davidson, W. Lipinski, Thermodynamic analysis of isothermal redox cycling of ceria for solar fuel production. Energy Fuels **27**, 5533 (2013)

32. L.J. Venstrom, N. Petkovich, S. Rudisill, A. Stein, J.H. Davidson, The effects of morphology on the oxidation of ceria by water and carbon dioxide. J. Sol. Energy Eng. **134**, 011005 (2012)

33. G. Hua, L. Zhang, G. Fei, M. Fang, Enhanced catalytic activity induced by defects in mesoporous ceria nanotubes. J. Mater. Chem. **22**, 6851 (2012)

34. J. Rossmeisl, W.G. Bessler, Trends in catalytic activity for SOFC anode materials. Solid State Ionics **178**, 1694 (2008)

35. P. Singh, M.S. Hegde, $Ce_{0.67}Cr_{0.33}O_2$: a new low−temperature O_2 evolution material and H_2 generation catalyst by thermochemical splitting of water. Chem. Mater. **22**, 762 (2010)

36. Y. An, M. Shen, J. Wang, Comparison of the microstructure and oxygen storage capacity modification of $Ce_{0.67}$. J. Alloy Compd. **441**, 305 (2007)

37. M. Zhao, M. Shen, X. Wen, J. Wang, Ce−Zr−Sr ternary mixed oxides structural characteristics and oxygen storage capacity. J. Alloy Compd. **457**, 578 (2008)

38. A.L. Gal, S. Abanades, N. Bion, T.L. Mercier, V. Harle, Reactivity of doped ceria-based mixed oxides for solar thermochemical hydrogen generation via two-step water-splitting cycles. Energy Fuels **27**, 6068 (2013)

39. A.L. Gal, S. Abanades, Dopant incorporation in ceria for enhanced water-splitting activity during solar thermochemical hydrogen generation. J. Phys. Chem. C **116**, 13516 (2012)

40. S. Abanades, A.L. Gal, CO_2 splitting by thermo-chemical looping based on $Zr_xCe_{1−x}O_2$ oxygen carriers for synthetic fuel generation. Fuel **102**, 180 (2012)

41. G. Kresse, J. Furthmuller, Efficient iterative schemes for ab initio total−energy calculations using a plane−wave basis set. Phys. Rev. B **54**, 11169 (1996)

42. G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented−wave method. Phys. Rev. B **59**, 1758 (1999)

43. J.P. Perdew, K. Burke, Y. Wang, Generalized gradient approximation for the exchange-correlation hole of a many−electron system. Phys. Rev. B **54**, 16533 (1996)

44. P.E. Blöchl, Projector augmented−wave method. Phys. Rev. B **50**, 17953 (1994)

45. V. Botu, R. Ramprasad, A.B. Mhadeshwar, Ceria in an oxygen environment: surface phase equilibria and its descriptors. Surf. Sci. **619**, 49 (2014)

46. M.B. Watkins, A.S. Foster, A.L. Shluger, Hydrogen cycle on CeO_2 (111) surfaces: density functional theory calculations. J. Phys. Chem. C **111**, 15337 (2007)

47. H. Kaneko, T. Miura, H. Ishihara, S. Taku, T. Yokoyama, H. Nakajima, Y. Tamaura, Reactive ceramics of CeO_2−MO_x (M = Mn, Fe, Ni, Cu) for H_2 generation by two−step water splitting using concentrated solar thermal energy. Energy **32**, 656 (2007)

48. M. Krcha, A.D. Mayernick, M.J. Janik, Periodic trends of oxygen vacancy formation and c−h bond activation over transition metal−doped CeO_2 (111) surfaces. J. Catal. **293**, 103 (2012)

49. Z. Hu, H. Metiu, Effects of dopants on the energy of oxygen−vacancy formation at the surface of ceria: local or global. J. Phys. Chem. C **115**, 17898 (2011)

50. V. Sharma, G. Pilania, G.A. Rossetti, K. Slenes, R. Ramprasad, Comprehensive examination of dopants and defects in $BaTiO_3$. Phys. Rev. B **87**, 134109 (2013)

51. D. Channei, B. Inceesungvorn, N. Wetchakun, S. Phanichphant, A. Nakaruk, P. Koshy, C.C. Sorrell, Photocatalytic activity under visible light of Fe− nanoparticles synthesized by flame spray pyrolysis. Ceram. Int. **39**, 3129 (2013)

52. T. Miki, T. Ogawa, M. Haneda, N. Kakuta, A. Ueno, S. Tateishi, S. Matsuura, M. Sato, Enhanced oxygen storage capacity of cerium oxides in $CeO_2/La_2O_3/Al_2O_3$ containing precious metals. J. Phys. Chem. **94**, 6464 (1990)
53. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. (Springer, New York, 2009)
54. I. Guyon, A. Elisseeff, An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
55. E.W. Bucholz, C.S. Kong, K.R. Marchman, W.G. Sawyer, S.R. Phillpot, S.B. Sinnott, K. Rajan, Data-driven model for estimation of friction coefficient via informatics methods. Tribol. Lett. **47**(2), 211–221 (2012)
56. S.C. Sieg, C. Suh, T. Schmidt, M. Stukowski, K. Rajan, W.F. Maier, Principal component analysis of catalytic functions in the composition space of heterogeneous catalysts. QSAR Comb. Sci. **26**(4), 528–535 (2007)
57. J.E. Jackson, *A User's Guide to Principal Components* (Wiley, New York, 1991)
58. I.T. Jolliffe, *Principal Component Analysis* (Springer, New York, 2002)
59. J. Shotton A. Criminisi, E. Konukoglu, Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical Report 114, Microsoft Research Technical Report (2011)
60. L. Breiman, Random forests. Mach. Learn. **45**, 5–32 (2001)
61. L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, *Classification and Regression Trees*, The Wadsworth and Brooks-Cole statistics-probability series (Taylor & Francis, Boca Raton, 1984)
62. MATLAB, *version 8.0.0.783 (R2012b)*. The MathWorks Inc., Natick, Massachusetts (2012)
63. 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, Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)