國立臺灣大學工學院化學工程學系

碩士論文

Department of Chemical Engineering

College of Engineering

National Taiwan University

Master Thesis

基於 CO 紅外光譜與深度學習研究二氧化鈰催化劑的
表面特性

Surface Characterization of Cerium Oxide Catalysts using

Deep Learning with Infrared Spectroscopy of CO

余欣諭

Hsin-Yu Yu

指導教授：李奕霈博士

Advisor: Yi-Pei Li, Ph.D.

中華民國112年7月

July, 2023

誌謝

在我讀碩士期間，因為有許多人的協助，包括我的指導老師-李奕霈教授、實驗室同學和家人，讓我現在有了這些成果。在這裡，我想一一感謝他們。

首先感謝李奕霈教授，在學術上給予我許多的幫助，總是能使用淺顯易懂的敘述或是簡單的例子，讓我快速了解，也在研究遇到困難時，給予適時的建議，使我能夠在這個領域取得了一些成就與自信，此外也讓我有機會與其他實驗室合作，接觸了很多不同領域的研究，非常感謝老師的信任，讓我在碩士短短的兩年內學習到很多經驗。

其次，我要謝謝 N415 的學長姐與同學們，謝謝世晟學長從我進實驗室開始帶領我熟悉量子模擬計算與機器學習的相關理論知識，在這兩年中，也經常很有熱忱與耐心的與我討論並解決研究上遇到的問題。感謝 Bala 在寫作上給我很多協助，讓我可以順利完成。謝謝隆義學長與筑宜學姐，在遇到 ML 的相關問題時，能給我建議。謝謝采合學姐經常分享很多有趣的事，邀約大家一起吃飯。謝謝岩均與育成在修課上與研究的幫助。謝謝彥翔和郁祺總是能讓實驗室的氣氛維持得很好，讓大家在愉快的環境中做事。

最後要感謝最支持我的家人，在低潮時傾聽我的煩惱陪伴在我身邊，他們理解我的壓力，鼓勵我克服一切，並支持我的每一個決定，也謝謝他們在背後默默的付出，讓我能更專注在學業上，無後顧之憂。也謝謝我的朋友們，雖然沒辦法經常見面，但偶而來台北約我出去玩，或是在電話中跟我分享工作上有趣的事情，有你們的陪伴，讓我在台北的這兩年感到不孤單。

<div align="right">

欣諭

2023.07 誌於國立台灣大學 計算化學研究室

</div>

中文摘要

　　材料表面的特性表徵對於理解其性質與行為至關重要。本研究利用 CO 作為探測分子的紅外光譜，結合深度學習技術，探索二氧化鈰催化劑的表面特性。通過對不同 CeO_2 表面 CO 衍生物的密度泛函理論 (Density Functional Theory) 系統性研究，獲得了包含 CO 在 CeO_2 表面的振動頻率、強度和吸附能等完整數據集。這些數據集被用來合成大量的複雜紅外光譜，用於訓練深度學習模型，以預測表面結構，包括 CeO_2 表面的分布、CO 衍生物和結合能的分佈。這些模型成功地分析了 CO 在不同類型 CeO_2 表面吸附的實驗紅外光譜，大多數情況下預測與實驗觀察結果一致。本研究提供了一種機器學習方法，以理解多種 CeO_2 材料的形態、局部環境排列、探測分子的交互作用行為和催化特性。
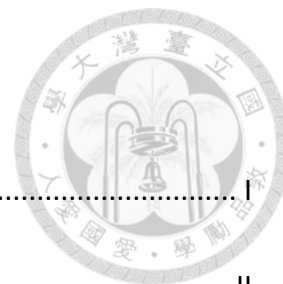
ABSTRACT

Characterization of material surfaces is crucial for understanding their properties and behavior. In this work, we utilized a deep learning technique, along with infrared (IR) spectrum of CO as a probe molecule, to explore the surface properties of cerium oxide (CeO_2) catalysts. Through systematic density functional theory investigation of CO-derived adspecies on various CeO_2 facets, we obtained an extensive dataset containing vibrational frequencies, intensities, and adsorption energies of CO on CeO_2. This dataset was used to synthesize large quantities of complex IR spectra to train deep learning models for predicting surface structures, including the distribution of CeO_2 facets, CO-derived adspecies, and binding energies. These models were successful in analyzing experimental IR spectra of CO adsorbed on different types of CeO_2, and their predictions were consistent with experimental observations in most cases. This work provides a machine learning approach in understanding the morphology, local environmental arrangement, interaction behavior of probe molecules, and catalytic characteristics of diverse CeO_2 materials.

目　錄

圖目錄

表目錄

1. Introduction

One of the most significant metal oxide (MO_2) catalysts is CeO_2, which exhibits an excellent ability to store and release oxygen gas and to transition between the +4 and +3 oxidation states of cerium [1,2]. CeO_2 is used as a catalyst in many gas phase redox reactions because of these characteristics, most notably in the water-gas shift [3,4], CO_2 reduction [5], and CO oxidation [6] reactions. Furthermore, by transferring its lattice oxygen to reactive adsorbates and so accelerating the oxidation reaction, it contributes significantly to the metal-support (M/CeO_2) interface reaction as support [6–11]. The acid-base properties of CeO_2 also plays a significant role by enhancing CO_2 chemisorption, thereby contributing to the improvement of catalytic performance and further amplifying the capacity to capture CO_2 [12,13]. Apart from serving as catalysts and supports, CeO_2 nanoparticles demonstrate significant potential in environmental sensing [14,15] and environmental remediation [16,17], with a successful implementation of green synthesis methods [17,18]. Therefore, understanding the characteristics of CeO_2 surfaces (morphology, coordination number, oxygen vacancy, and local atomic arrangement), as well as how gas molecules interact with the oxidized and reduced surface of CeO_2, at an atomic level, is crucial from a scientific and technical perspective.

Traditionally, X-ray based spectroscopies are used to examine the oxygen vacancy and local atomic arrangement of the Ce atom. For instance, X-ray Photoelectron
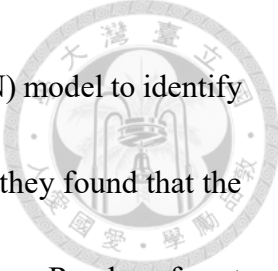
Spectroscopy (XPS) [19–24] and X-ray Absorption Near Edge Spectroscopy (XANES) [25–28] can be used to determine the chemical state of Ce atoms and reveal their local electronic structure. Imaging the crystalline structure and crystallographic defects at the atomic level is possible with high-resolution transmission electron microscopy (HRTEM) [29–32]. Adsorption of probe molecules is considered to be a convenient method for determining the surface structures of various CeO_2 facets. The combination of infrared (IR) spectroscopy [33–40] and cutting-edge computational modeling [41–48] became the all-encompassing method to study the chemisorption of CO or CO_2 on CeO_2. For instance, Chen *et al*. [40] used in-situ diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) to extensively investigate the chemisorption of CO and CO_2 on uniform CeO_2, TiO_2, and Cu_2O nanocrystals with varied morphologies. They observed that the formed adsorbates are morphology-dependent, surface-composition-sensitive, and coordination environment-sensitive. Through the use of first-principle calculations, Yang *et al*. [46] examined the CO adsorption on stoichiometric (110) and (111) CeO_2 surfaces. They observed that whereas both weak and strong interactions of CO exist on (110) and prefer both atop and bridged carbonate topologies, the weak interaction of CO on (111) prefers atop configuration. Density functional theory (DFT) and Fourier transform infrared (FTIR) spectroscopy were used by Vayssilov *et al*. [47] to study the surface species produced by the interaction of CO_2 or CO with activated, reduced, and hydroxylated CeO_2

2

nanoparticles and reassign various experimentally observed vibrational modes to different surface types. Additionally, stoichiometric and defective surfaces of single crystals of CeO_2 can be identified using IR spectroscopy and first-principle calculations in tandem [48–51] by measuring the shift in the vibrational frequency of the adsorbate, preferably under ultrahigh vacuum (UHV) conditions. Despite the fact that HRTEM and X-ray based spectroscopic techniques are excellent and frequently employed, they work in UHV settings and necessitate specialist expertise for in-depth investigations. While in IR spectroscopy, peak assignments are heuristic and challenging for complicated IR spectra. It is clear from the discussion above that there is still a lack of sophisticated and affordable methods for correctly interpreting the CeO_2 surface. In recent years, ML has gained popularity among researchers, who believe that this strategy can be used in conjunction with first-principles strategies to overcome the majority of the restrictions of experimental design.

Numerous studies have used ML techniques to interpret the spectra of metal catalysts, which contain a wealth of structure-dependent spectroscopic information about the material, to understand the surface characteristics of the catalysts, such as coordination number, atomic environment, and so forth. The XANES spectra of nanoparticles (NPs) were converted into the 3D geometry of metal catalysts using a supervised ML approach developed by Timoshenko *et al.* [52]. On the basis of a synthetic random training set,

3

Drera *et al.* [53] developed a deep convolutional neural network (CNN) model to identify and quantify chemical constituents from experimental XPS data, and they found that the accuracy of the model is on par with that of conventional XPS users. Random forest models were directly employed by Zheng *et al.* [54] to predict the coordination environment with high accuracy from the XANES spectra, and their model performance was demonstrated to outperform deep learning methods. Lansford *et al.* [55] recently proposed a method to characterize Pt surface microstructure using CO and NO as probe molecules by employing multinomial regression using neural network ensembles and generating synthetic IR spectra from first-principles computations.

In this study, we enhanced the approach of Lansford *et al.* to comprehend the CeO_2 surface properties utilizing CO as a probe molecule. On CeO_2 (100, 110, and 111) oxidized and reduced surfaces, we considered five CO adsorption configurations, including atop, bidentate carbonate (BDC), polydentate carbonate (PDC), bicarbonate, and bridged carbonate. This article primarily focuses on (i) creating a dataset of frequencies and intensities derived from DFT, taking into account a wide range of possible structural forms of CeO_2 surfaces with CO-derived adspecies, (ii) designing three deep learning models to independently predict CeO_2 crystal facets, surface species, and adsorption energy (E_{ads}), and (iii) evaluating the performance of our model using both synthetic and real-world complex IR spectra. We anticipate that by supplying solely

4

infrared spectra, our models will be able to comprehend the CeO_2 crystallographic plane

and measure the ratio between each facet, which will overcome the constraints of HRTEM

and provide crucial information for experimentalists. Additionally, the arrangement of

CO on the CeO_2 surface may be quickly determined, and the environment surrounding

the adsorption site can be examined. We think that our model offers a quick and simple

method for analyzing the characteristics of unidentified CeO_2 surfaces.

2. Methodology

2.1 Model overview

This work aims to develop machine learning models that can analyze the IR spectrum of CO adsorbed on CeO_2 and predict the characteristics of the CeO_2 surface. To achieve this, the input to the machine learning model should be the IR spectrum, while the output should be the surface information of interest. However, a limited amount of high-quality CO IR spectrum with detailed CeO_2 surface information is available, so it is challenging to train the machine learning model solely using experimental data. To overcome this issue, we used DFT calculations to create a dataset that covers the IR spectrum of a wide range of CO-adsorbed species on CeO_2. In this study, various active sites of CeO_2 surfaces were considered, where CO can adsorb in different configurations, such as atop, BDC, PDC, bicarbonate, and bridging carbonate. As shown in Fig. 1, DFT calculations were used to determine the frequencies, intensities, and E_{ads} for these configurations. To account for DFT errors, the computed frequencies were compared to available experimental data to derive scaling factors. Using the scaled DFT frequencies and selected experimental data, synthetic complex IR spectra were generated by combining individual IR spectra to simulate real-world conditions where different CO-adsorbed species and CeO_2 surfaces coexist in the sample. Three deep learning models were built to predict the distribution of CeO_2 facets, CO-derived species, and adsorption

6

energies. The model performance was evaluated using both synthetic and actual experimental IR spectra. The following subsections will provide detailed discussions of each step in the workflow.
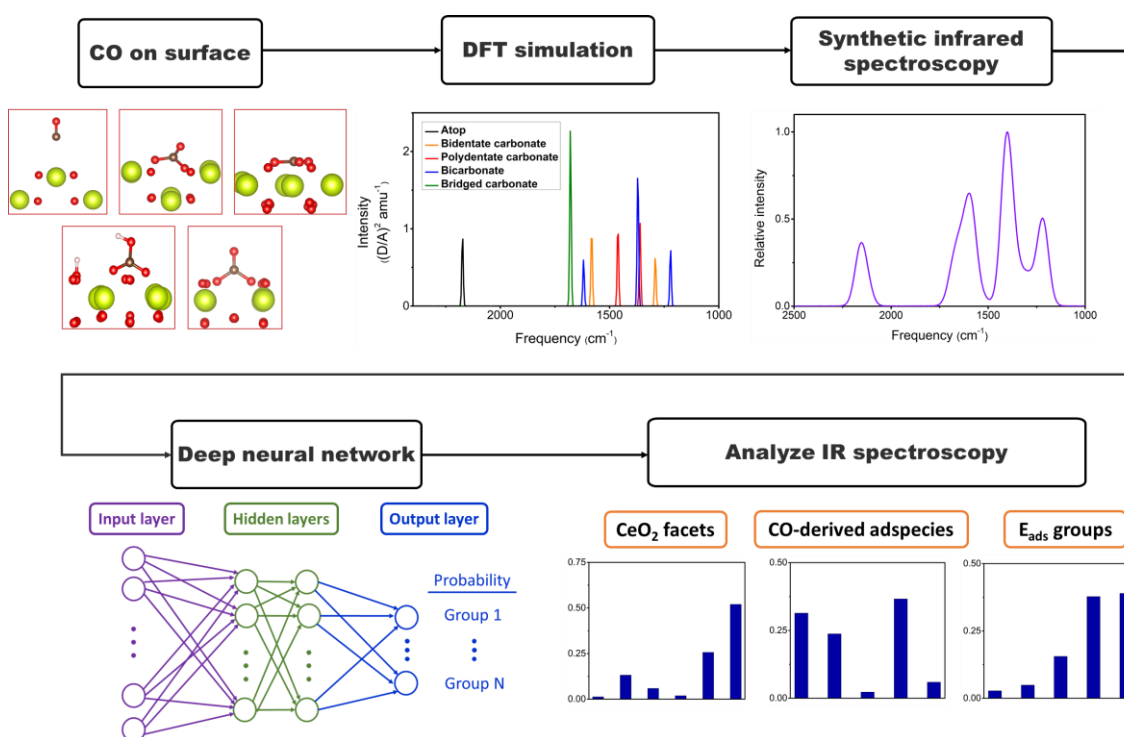
Fig. 1. Overview of the process of data preparation and model training. This study used DFT calculations to create a dataset covering a wide range of CO-adsorbed species on CeO_2, including various active sites and binding configurations. Synthetic complex IR spectra were generated using the computed frequencies and selected experimental data. Three deep learning models were built to predict the distribution of CeO_2 facets, CO-derived species, and adsorption energies.

7

2.2 DFT

All periodic DFT calculations were performed by the Vienna Ab initio Simulation Package (VASP) [56,57]. The core-electron interactions were expressed using the projector augmented wave (PAW) approach [58]. The generalized gradient approximation (GGA) of the Perdew-Burke-Ernzerhof (PBE) functional was utilized to model electron exchange and correlation [59]. The plane wave energy cutoff was set at 450 eV. The Hubbard U parameter was employed with a value of 5 eV on Ce, consistent with earlier studies [60,61] and corrected for Coulomb interactions. Monkhorst-Pack k-points were linearly changed from 2 x 2 x 2 to 12 x 12 x 12 to determine the optimal lattice parameter for bulk CeO_2. The different slabs (100, 110, and 111) were made using the optimized bulk CeO_2. The vacuum thickness was set to 15 Å for all slabs in order to ignore the z-direction interaction. In this study, all surfaces were modeled as (2 x 2) supercells, and the integral in the Brillouin zone was performed using samples taken from the surfaces (100), (110), and (111), respectively, using 3 x 3 x 1, 6 x 4 x 1, and 6 x 6 x 1. During the geometry optimization, all atoms in the bulk computation were allowed to relax, and for the (100), (110), and (111) slab models, the top 6, 3, and 6 layers were totally relaxed, while the remaining bottom layers remained fixed [44,61,62]. The CO molecule in the gas phase was optimized at Γ-point with a cell size of 15x15x15 $Å^3$. The systems of CO adsorbed slabs were optimized using the same criteria as free slab relaxation. We also

8

considered the slab with a hydroxyl group and a hydrogen atom on the surface because

H_2O can dissociate on CeO_2 surfaces [63]. The energy and force convergence parameters

were set at 10^{-6} eV and 0.01 eV/Å, respectively. Dipole corrections and spin-polarized

were taken into account in all slab calculations.

The finite difference approach with a displacement of 0.015 Å was used to calculate

the vibrational frequencies and electron densities of the modeled CO adsorbate, carbonate

(CO_3), and bicarbonate ($CO_2(OH)$. During these calculations, only the CO adsorbate, CO_3,

and $CO_2(OH)$ were relaxed, while the other atoms were kept frozen. The electron

densities obtained were integrated with CHARGEMOL [64,65] to compute the dipole-

moment of adsorbed species. The resulting dipole-moment was utilized to compute the

intensity (I) by utilizing Porezag and Pederson equation [66].

$$I_i^{IR} \propto \left| \frac{d\mu}{dQ_i} \right|^2 \qquad (1)$$

where μ is the dipole moment of the system, and Q_i is the coordinate of normal mode

i. The E_{ads} is computed using the following equation:

$$E_{ads} = E_{CO/slab} - E_{slab} - E_{CO} \qquad (2)$$

where $E_{CO/slab}$ is the total energy of CO adsorbed slab, E_{slab} is slab energy, and E_{CO} is CO

molecule energy in the gas phase.

9

For the surface energy (E_{surf}) calculations, the value can be determined as follows:

$$E_{surf} = (E_{slab} - nE_{bulk})/2S \qquad (3)$$

where E_{slab} and E_{bulk} are the total energies of the slab and the bulk CeO_2, n is the ratio of the number of slab atoms to the number of bulk atoms, and S denotes the surface area. Atomic Simulation Environment (ASE) [67] was used for the post-processing of the VASP output files.
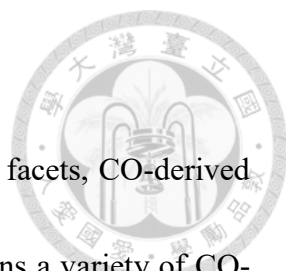
2.3 Data cleaning and preparation

Under specific conditions, it is possible to measure the frequency of particular adsorption modes of CO experimentally. For example, the frequencies of the atop site were measured using single crystal CeO_2 under UHV conditions [44,49,51]. However, high-quality UHV IR spectra with CeO_2 surface information are rare, especially for CO chemisorption, where defective surfaces are involved. For these cases, DFT calculated frequencies were used instead. To account for the DFT errors, we assigned several experimental frequencies that are close to the DFT results to obtain scaling factors. Therefore, in addition to using the experimental frequencies for the atop configuration measured with single crystal CeO_2 under UHV, the scaled DFT frequencies were used for other CO-derived adspecies in the generation of synthetic complex IR spectra. This not only eliminates DFT calculation errors but also results in a clearer dataset.

10

The synthetic complex IR spectra were generated following the procedures reported by Lansford *et al.* [55]. The frequencies and intensities of each CO-derived adspecies were convoluted using Fourier transforms with a Gaussian filter, resulting in the generation of single-adsorbed CO spectra. Leveraging the linearity between IR spectral intensity and the number of molecules [66], we obtained complex IR spectra by combining the single-adsorbed CO spectra through linear combinations. To eliminate noise from low frequencies, we only consider DFT frequency peaks in 1200 to 3000 cm^{-1} range. We systematically assessed the model performance by increasing the DFT data and computed the optimum needed data for the deep neural networks (DNN) models. The minimum required data for the CeO_2 facet model is 2980, whereas the minimum required data for both CO-derived adspecies and E_{ads} groups is 3280. We increased the number of data points by perturbing the frequencies and intensities to address the issue of uneven distribution of data and account for scaling factor errors. Fig. S1 depicts the training losses versus the number of data points for the DNN models. The DFT data was separated into 25% for testing and 75% for training and 3-fold cross-validation. We generated synthetic spectra by randomly selecting *n* single CO adsorption IR spectra, with a maximum *n* of 200, to ensure wide coverage of all possible distributions. The full width at half maximum (FWHM) was chosen at random from 2 to 75 cm^{-1}. The final spectra were convoluted with 4 cm^{-1} resolution using Gaussian and Lorentzian filters [68].

11

2.4 ML approach

This work employed DNN to predict the distributions of CeO_2 facets, CO-derived adspecies, and E_{ads}. Since the synthetic complex IR spectrum contains a variety of CO-derived adspecies, it is necessary to represent the E_{ads} for these adspecies as a distribution rather than a single value. In order to achieve this, the E_{ads} values were categorized into k groups using an unsupervised learning method known as K-means clustering. The elbow approach was used to determine the ideal k value, which was found to be 5. Based on the results of K-means clustering, group 1 to 5 are the E_{ads} ranges -(4.88 ~ 4.09 eV), -(4.09 ~ 3.25 eV), -(3.25 ~ 2.27 eV), -(2.27 ~ 0.94 eV), and -(0.94 ~ 0 eV). Fig. 2 clearly shows that the atop configuration belongs to Group 5. Other configurations fall into more than one group. There will be detailed discussions on the E_{ads} of different binding configurations in the subsection below.
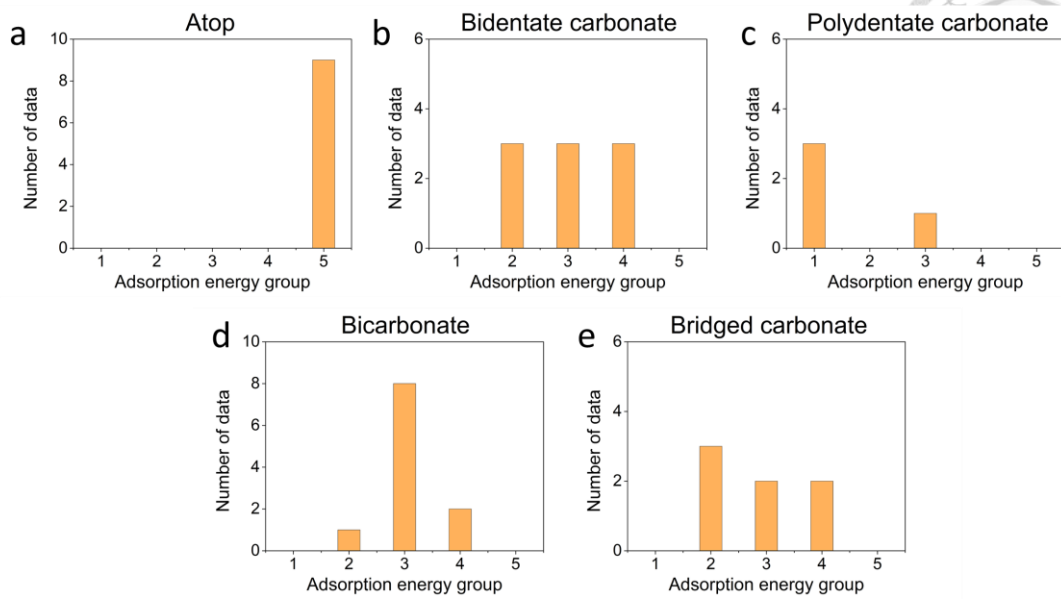
Fig. 2. Classification of the E_{ads} groups using the K-means clustering method for (a) atop, (b) BDC, (c) PDC, (d) bicarbonate, and (e) bridged carbonate.

Unless noted otherwise, the DNN models used in this work were trained with the Adam optimizer, and the rectified linear unit (ReLU) was used as the activation function of hidden layers. In the output layer, the softmax (S_i) activation function was employed to construct the probability distributions of each group:

$$S_i = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \ (i = 1, 2, \ldots, N) \tag{4}$$

where x_i and x_j are the outputs of the last hidden layer, and N is the total number of groups.

13

The loss function for measuring the similarity of probability distributions was Wasserstein distance (W) [69], and the equation follows from Lansford $et\ al.$ [55]:

$$W = \left[\frac{1}{C} \sum_{n=1}^{C} \left[\sum_{i=1}^{n} p_i - \sum_{i=1}^{n} t_i \right]^2 \right]^{\frac{1}{2}} \qquad (5)$$

where p_i and t_i are the predicted and actual probability values of the ith group, and C is the total number of groups. R squared (R^2) and Wasserstein measures were used to estimate model performance. The following is the formula for computing R^2 value:

$$R^2 = 1 - \frac{\sum_{k=1}(t_k - p_k)^2}{\sum_{k=1}(t_k - \bar{t})^2} \qquad (6)$$

We utilize random search to optimize hyperparameters, and the best hyperparameters of the models are summarized in Tables S2 and S3. We train 60 models with the best hyperparameter settings to derive an ensemble prediction and estimate uncertainty and confidence interval [70-72]. In this work, the predicted distributions are shown with error bars, which represent a 95% confidence interval estimated by the ensemble approach.

14

3. Results and Discussion

3.1 DFT

3.1.1 Optimization of bulk CeO$_2$ and free slabs

The bulk CeO$_2$ has a cubic fluorite structure, as seen in Fig. 3. Cerium and oxygen are 8-fold and 4-fold coordinated, respectively. The best lattice parameter determined for the (6x6x6)-Monkhorst-Pack k-point mesh is 5.48 Å, which agrees well with the experimental value of 5.41 Å [73]. We note that because CeO$_2$ (100) is Tasker Type 3 oxide [74], a stable (100) surface of CeO$_2$ needs to be created by deleting half of the oxygen from both the top and bottom of the O-terminated surface slabs [75,76].
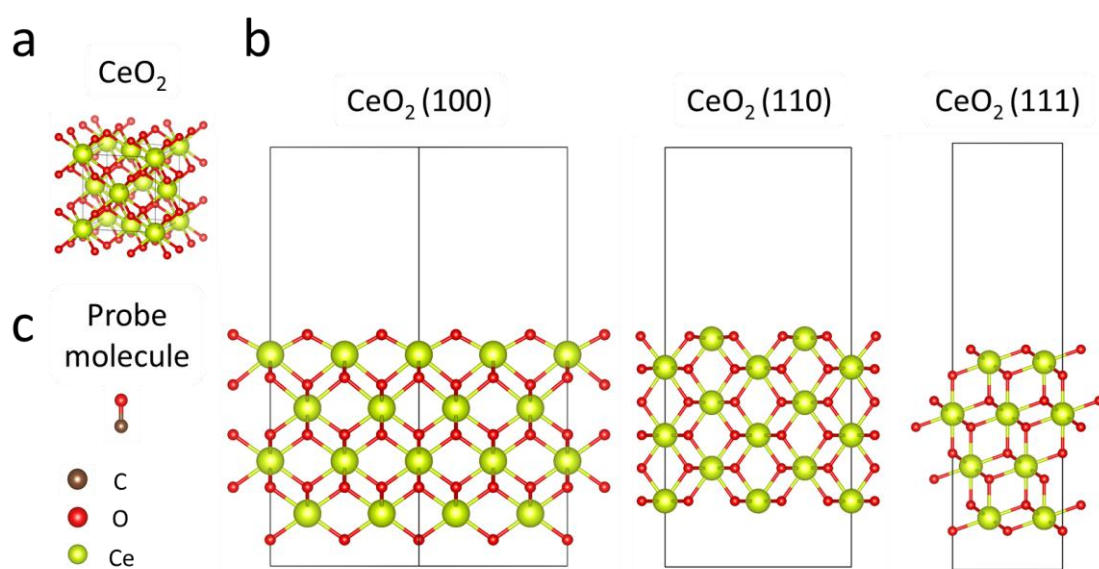
Fig. 3. (a) Cubic fluorite structure of bulk CeO$_2$. (b) Stoichiometric (100), (110), and (111) CeO$_2$ surfaces with (2x2) supercell. (c) Optimized CO as the probe molecule.
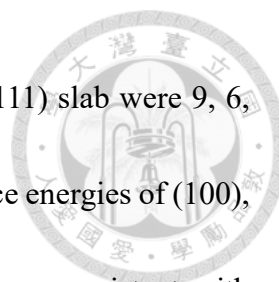
The number of atomic layers in each CeO_2 (100), (110), and (111) slab were 9, 6, and 12 accordingly, as reported in the literature [44,62,77]. The surface energies of (100), (110), and (111) are 1.76, 1.15, and 0.65 J/m^2, respectively, and are consistent with previously reported values [61,78], and the relative stability of the surface is in the order of (111) > (110) > (100). Table 1 compares the predicted surface parameters of bulk CeO_2 and its slabs to the literature.

Table 1. Properties of optimized bulk CeO_2 and CeO_2 (100), (110), and (111) surfaces.

	Layers	Volume ($Å^3$)	Surface energy (J/m^2)		Coordination number	
			This work	Previous study	O	Ce
Bulk CeO_2					4	8
CeO_2 (100)	9	2624.65	1.76	1.76 [a]	2	6
CeO_2 (110)	6	2101.79	1.15	1.01 [b]	3	6
CeO_2 (111)	12	1358.41	0.65	0.71 [a]	3	7

[a] ref [61]. [b] ref [78].

3.1.2 Optimization of CO adsorbed CeO$_2$ slabs

We considered stoichiometric CeO$_2$ slabs as oxidized (oxid-CeO$_2$) and introduced reduced (redu-CeO$_2$) slabs by removing one additional surface oxygen. This approach allowed us to investigate the influence of the presence of an extra oxygen vacancy on the system. As indicated in Fig. 4, five possible configurations of CO adsorbed CeO$_2$ were investigated in this work, all of which are common experimentally observed configurations. In theory, the direct adsorption of CO on an oxygen vacant site would theoretically lead to the formation of a carbonite species. However, Vayssilov *et al*. [47] conducted calculations and determined that carbonite species are unstable on a ceria surface, making their experimental formation and identification unlikely. As a result, the carbonite species was not considered in this study.

The possible configurations of CO adsorbed CeO$_2$ were optimized and then used to calculate vibrational frequencies. To account for differences in DFT-derived frequencies of each CO-derived adspecies, the frequency corresponding to each vibrational mode was scaled separately. The scaling technique is explained in detail in the Supplementary Information. We note that some of the binding configurations were only found on certain types of the CeO$_2$ surface. Tables 2-6 present a comparison between the scaled and unscaled DFT frequencies and the available experimental values, along with the intensity and E$_{ads}$ of CO on CeO$_2$ (100), (110), and (111). In practice, the frequencies and E$_{ads}$ can

17

exhibit variations at different surface coverages, especially at high coverages where the

interaction between neighboring CO molecules becomes substantial. However, it is

important to note that our DFT calculations were conducted on a fixed-size model with a

single CO molecule. Therefore, we did not consider this particular aspect in our

calculations.
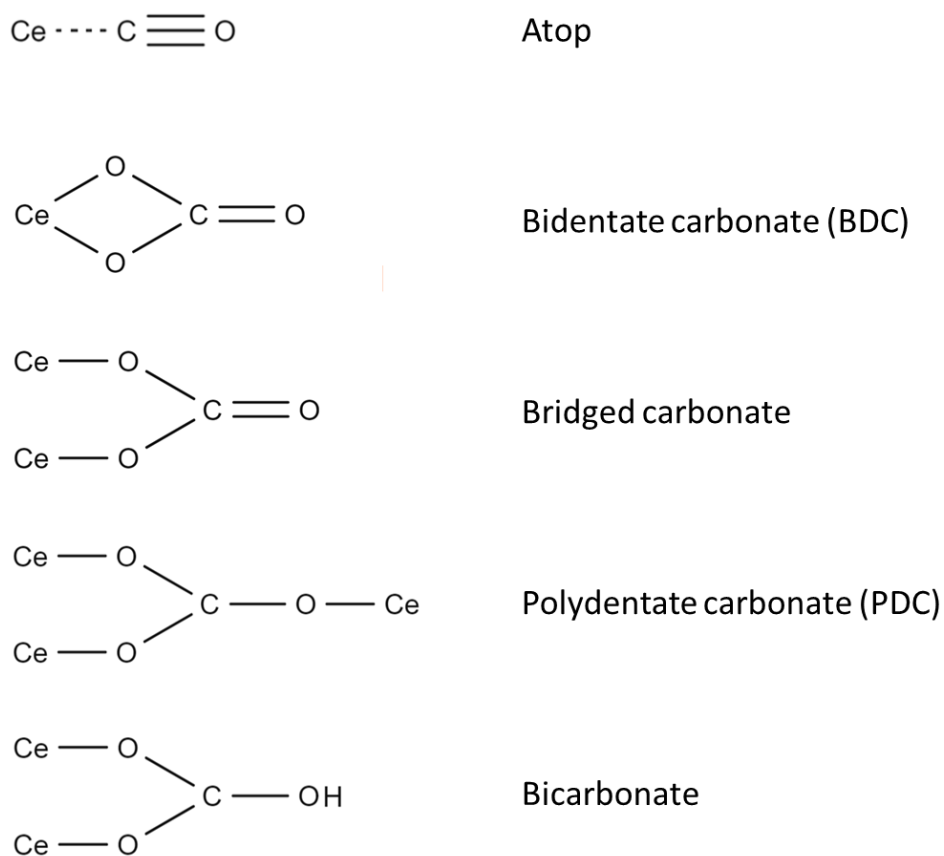
Fig. 4. Various binding sites on CeO_2 surfaces and patterns of CO adsorption [34,35].

3.1.2.1 Atop configuration

CO adsorbs on the top site of Ce with weak interactions on all CeO_2 (100), (110), and (111) slabs in atop arrangement, as shown in Fig. 5. The estimated E_{ads} of CO-adsorbed oxid- and redu-CeO_2 (100), (110), and (111) are in the range of -(0.33 to 0.18) eV, with an average E_{ads} of about -0.24 eV with the surface structure remaining nearly intact after adsorption. The adsorbed C-O bond length and stretching frequency are (1.14-1.15) Å and (2088-2165) cm^{-1}, respectively, which are in the same ranges of the experimentally measured bond length (1.13 Å) and stretching frequency (2143 cm^{-1}) of gas phase CO [79]. Especially for oxid-(110), Lustemberg *et al.* [44] observed two vibrational frequencies (2171 and 2160 cm^{-1}) in the IR spectrum, and assigned to the vertical and several tilted configurations of CO, respectively, by DFT calculations, as shown in Table 2. Similar to the finding of Lustemberg *et al.*, we also found that a single binding configuration may exhibit several slightly distinct stable structures, primarily because of differences in the orientation of the adsorbate (Fig. 5b-c). Therefore, a binding configuration could have multiple sets of slightly different frequencies. This phenomenon is also observed in the cases of CO chemisorption, as listed in Tables 3-6 below.
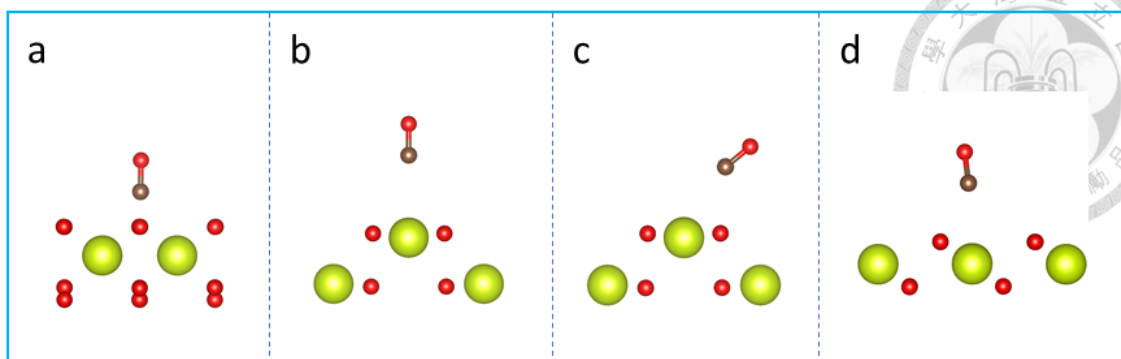
Fig. 5. (Side view) Atop configuration of CO on (a) oxid-CeO$_2$ (100), (b,c) oxid-CeO$_2$ (110), and (c) oxid-CeO$_2$ (111) slabs.

Table 2. Calculated vibrational frequencies, intensities, and corresponding E$_{ads}$ of atop site on CeO$_2$ (100), (110), and (111).

Facet	site	Frequency (cm^{-1})			Intensity (D/Å^2 amu^{-1})	E$_{ads}$ (eV)
		DFT	After scaling	Experimental		
100	oxidized	2141	2180	2147 [a]	13.36	-0.32
	reduced	2165	2204	2168 [a]	5.84	-0.29
110	oxidized	2141	2180	2171 [b]	8.03	-0.20
		2106	2144	2160 [b]	8.37	-0.21
		2118	2156	2160 [b]	6.60	-0.24
		2099	2137	2160 [b]	6.67	-0.23
	reduced	2088	2126	2175 [c]	10.84	-0.20
111	oxidized	2126	2164	2154 [b,c]	6.55	-0.18
	reduced	2122	2160	2163 [c]	16.20	-0.33

[a] ref [51]. [b] ref [44]. [c] ref [49].

20

3.1.2.2 BDC configuration

The BDC configuration was mostly found on CeO_2 (110) and (111) slabs where carbon of CO binds with two surface oxygen atoms. Our calculations did not find BDC on the CeO_2 (100) slab. As indicated in Fig. 6, two possible BDC sites, O-bridged and Ce-bridged, were considered. Both O-bridged and Ce-bridged sites exist on the CeO_2 (110) slab, but only the Ce-bridged site exists on the CeO_2 (111) slab. For Ce-bridged site on both CeO_2 (110) and (111) slabs, one of the surface oxygen atoms is dragged out of the plane. As shown in Table 3, we found more than one O-bridged or Ce-bridged site structure on both CeO_2 (110) and (111) slabs. Based on the definition of Vayssilov *et al.* [47]*,* these species can be further classified into 1.21, 1.2.1, and 1.3.1 types to describe the coordination numbers of the three O atoms with the Ce atoms.
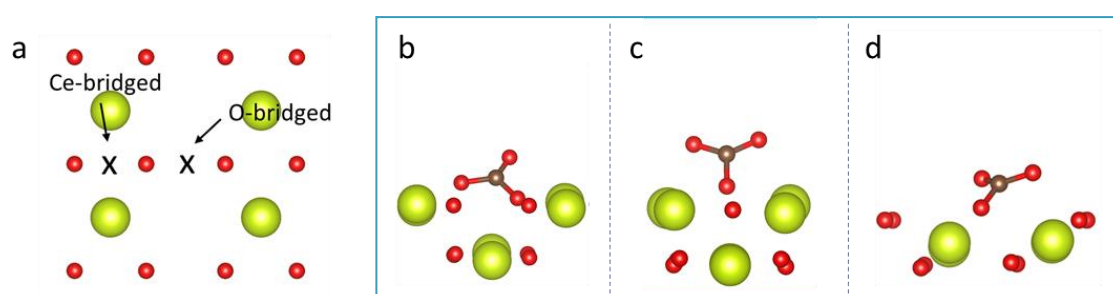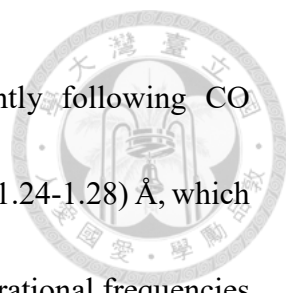
Fig. 6. (a) (Top view) Schematic representation of the adsorption sites of O-bridged and Ce-bridged on the CeO_2 (110) surface, labeled as 'X'. (Side view) BDC configuration of CO on (b) O-bridged site of oxid-CeO_2 (110), (c) Ce-bridged site of oxid-CeO_2 (110), and (d) Ce-bridged site of oxid-CeO_2 (111) slabs.

Unlike the atop site, the surface shape changed significantly following CO adsorption. The computed C-O bond lengths in BDC are in the range (1.24-1.28) Å, which is 10-14% longer than the free CO bond length. Furthermore, the vibrational frequencies are considerably red-shifted. Table 3 shows the E_{ads} and vibrational frequencies for the BDC arrangement on CeO_2 (110) and (111). Based on the CO E_{ads} and stretching frequencies, it is possible to conclude that the O-bridge site (average E_{ads} of -3.54 eV) has a stronger interaction than the Ce-bridged site in the case of CeO_2 (110) slab (average E_{ads} is -2.72 eV). For CeO_2 (111), we found that it has only one adsorption (Ce-bridged) site, and one of the surface oxygen atoms is dragged out of the plane. The vibrational frequencies are (1194 - 1234) cm^{-1} and (1557 - 1592) cm^{-1}, and the average E_{ads} is -1.79 eV. The Ce-bridged BDC configuration of CO on CeO_2 (111) is weaker than both the O- and Ce-bridged BDC arrangements on (110). In our calculations, the formation of BDC species is a highly exothermic process, which is in line with the DFT calculations performed by Chen *et al.* [80] and Song *et al.* [81].

22

Table 3. Calculated vibrational frequencies, intensities, and corresponding E_{ads} of BDC on CeO$_2$ (110) and (111).

Facet	site		Frequency (cm^{-1})			Intensity (D/Å^2 amu^{-1})	E_{ads} (eV)
			DFT	After scaling	Experimental		
110	oxidized	O-bridged	1228, 1539	1287, 1562	1296, 1563 [a]	6.41, 11.81	-3.73
			1270, 1474	1331, 1496		9.45, 8.97	-3.51
		Ce-bridged	1247, 1574	1306, 1598		9.77, 13.00	-2.91
	reduced	O-bridged	1223, 1566	1282, 1589		7.53, 13.88	-3.17
			1255, 1511	1315, 1534		7.93, 13.42	-3.78
		Ce-bridged	1246, 1562	1306, 1585	1292, 1587 [b]	9.65, 13.42	-2.53
111	oxidized	Ce-bridged	1234, 1557	1293, 1580	1296, 1579 [c]	5.62, 8.84	-1.65
			1202, 1592	1260, 1616		5.31, 7.69	-1.77
	reduced	Ce-bridged	1194, 1572	1251, 1596		6.01, 13.83	-1.96

[a] ref [36]. [b] ref [37]. [c] ref [40].

3.1.2.3 PDC configuration

Our simulations only found the PDC configurations on the oxid-CeO_2 (100) and redu-CeO_2 (100 & 111) slabs. Fig. 7 depicts the carbon and oxygen of CO, as well as two surface oxygens, on a plane parallel to the CeO_2 surface. As shown in Fig. 7, we optimized two orientations of CO_3 species which differs by 90° on redu-CeO_2 (100), and only one stable PDC configuration for oxid-CeO_2 (100) and redu-CeO_2 (111). The three C-O bonds are about the same length, around 1.30 Å, and form highly symmetric structures among various configurations. Similar to BDC, the PDC configuration exhibits two distinct stretching frequencies. However, due to the symmetry of PDC, these two peaks are closely spaced. In contrast, the two peaks in BDC are well-separated. The calculated vibrational frequencies correlate well with the experimental ones [34,38], and their values are shown in Table 4. Furthermore, in terms of E_{ads}, the PDC configuration exhibits a stronger interaction of CO with the CeO_2 surface than the BDC configuration.
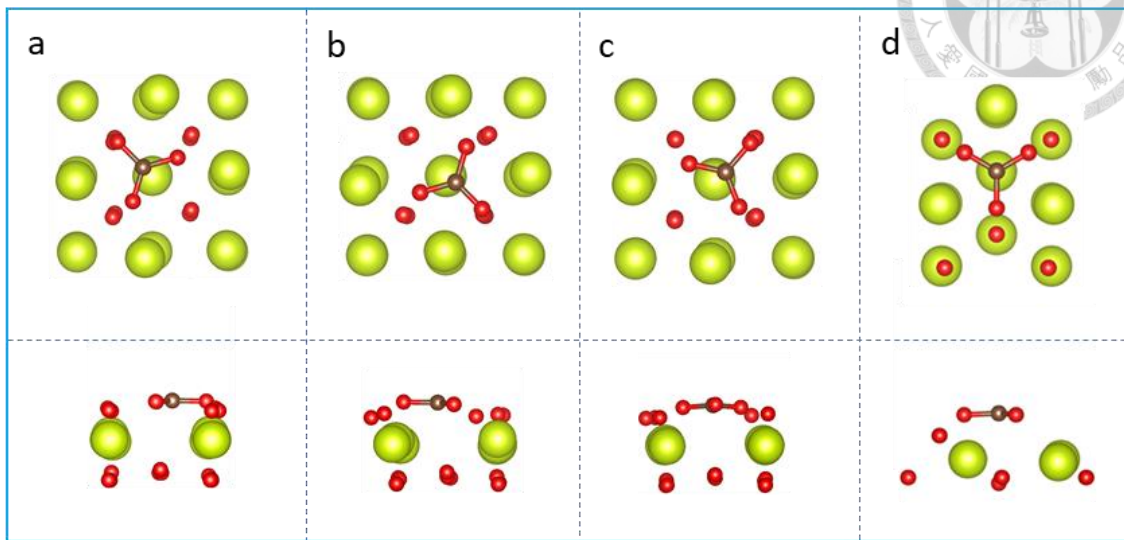
Fig. 7. (Upper: Top view, Lower: Side view) PDC configuration of CO on (a) oxid-CeO$_2$ (100), (b,c) two orientations of CO$_3$ on redu-CeO$_2$ (100), and (d) redu-CeO$_2$ (111) slabs.
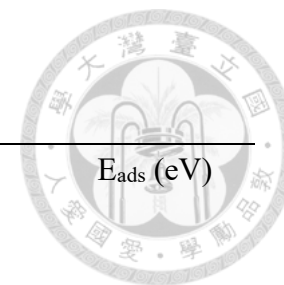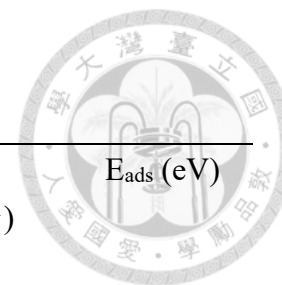
Table 4. Calculated vibrational frequencies, intensities, and corresponding E_{ads} of PDC on CeO_2 (100) and (111).

Facet	site	Frequency (cm^{-1})			Intensity ($D/Å^2$ amu^{-1})	E_{ads} (eV)
		DFT	After scaling	Experimental		
	oxidized	1386, 1425	1376, 1463	1367, 1476 [a]	10.29, 13.65	-4.87
100		1372, 1422	1362, 1460	1367, 1476 [a]	9.40, 9.26	-4.48
	reduced					
		1351, 1444	1342, 1483	1348, 1454 [b]	11.30, 10.46	-4.39
111	reduced	1364, 1365	1354, 1402		10.24, 10.19	-2.97

[a] ref [38]. [b] ref [34].

3.1.2.4   Bicarbonate configuration

The interaction of CO with the OH group present in all three low-index CeO_2 slabs results in the formation of the bicarbonate structure. Vayssilov *et al.* [47] conducted a study on bicarbonate species, exploring four distinct structures through DFT calculations. However, their analysis revealed that the vibrational frequencies of these species are highly similar, making it challenging to differentiate them based solely on vibrational frequencies. Considering this difficulty, in our study, we opted to group all $CO_2(OH)$ compounds together under the broader category of bicarbonates, without further differentiation.

Multiple stable structures of bicarbonate species were discovered on all oxidized surface. Oxid-(100) exhibits two distinct structures: one closely resembling the bridged form (Fig. 8a), while the other resembling the BDC form (Fig. 8b). Although oxid-(110) and (111) only has one adsorption site, there are two different hydrogen positions to be considered (Fig. 8c-d). The redu-(100) surface exhibits two orientations of $CO_2(OH)$, similar to PDC on redu-(100). On the redu-(110) surface, there are two distinct $CO_2(OH)$ structures corresponding to adsorption on O-bridged and Ce-bridged sites, respectively, resembling the BDC configuration. Table 5 shows the values of three peaks observed for the bicarbonate configuration: two $\nu(CO_3)$ and one $\delta(OH)$. The computed $\delta(OH)$ frequency is (1163 - 1192) cm^{-1}, which is well isolated from other $\nu(CO_3)$ peaks. The

27

calculated vibrational frequencies of two $\nu(CO_3)$ peaks are (1541 - 1636) cm^{-1} and (1349

- 1431) cm^{-1}, respectively. Though the two $\nu(CO_3)$ peaks are significantly distanced from

each other, these two vibrational ranges overlap with the BDC and PDC configurations.

Because of this overlapping, we believe that $\delta(OH)$ frequency is a key characteristic for

distinguishing bicarbonate species. The estimated bicarbonate frequencies agree with the

experimental values [33,35]. Table 5 shows the estimated vibrational frequencies,

intensities, and E$_{ads}$, as well as the available experimental values.
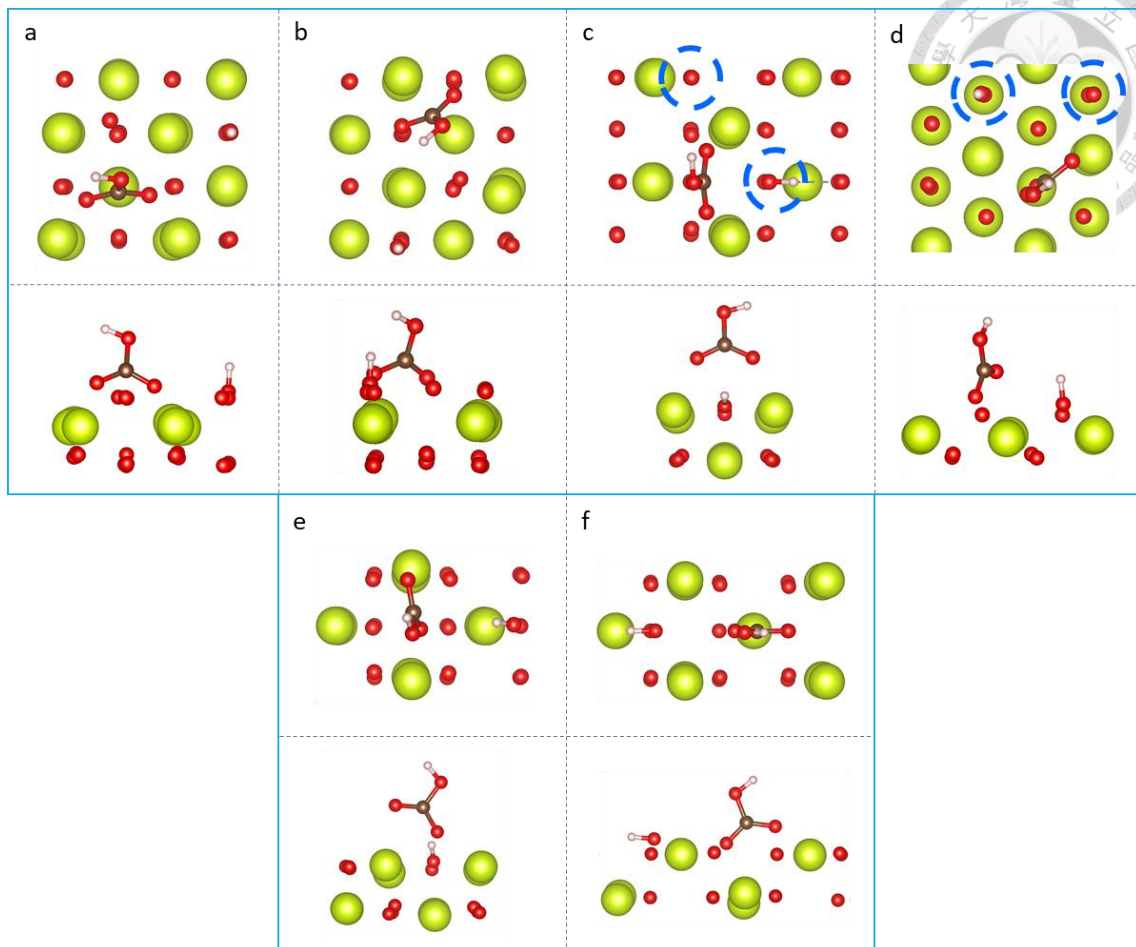
Fig. 8. (Upper: Top view, Lower: Side view) Bicarbonate configuration of CO on (a,b) oxid-CeO$_2$ (100), (c) oxid-CeO$_2$ (110), (d) oxid-CeO$_2$ (111), and (e,f) redu-CeO$_2$ (110). The blue dotted circles indicate the considered positions for the surface H atom.

Table 5. Calculated vibrational frequencies, intensities, and corresponding E_{ads} of bicarbonate on CeO_2 (100), (110), and (111).

Facet	site	Frequency (cm^{-1})			Intensity (D/Å² amu⁻¹)	E_{ads} (eV)
		DFT	After scaling	Experimental		
100	oxidized	1178, 1409, 1543	1223, 1433, 1540		8.83, 24.63, 3.68	-3.15
	O-bridged					
	Non Ce- or O-bridged	1163, 1431, 1541	1207, 1455, 1538		8.37, 23.54, 2.99	-3.33
	reduced O-bridged	1172, 1374, 1565	1217, 1397, 1562		6.35, 27.38, 3.64	-2.35
		1164, 1383, 1552	1207, 1407, 1549		10.43, 25.66, 1.82	-2.73
110	oxidized Ce-bridged	1192, 1357, 1636	1237, 1380, 1633		7.56, 29.66, 10.65	-2.81
		1191, 1349, 1629	1236, 1372, 1626		5.07, 35.50, 11.51	-3.12
	reduced O-bridged	1176, 1352, 1616	1221, 1375, 1613	1218, 1391, 1613 [a]	6.35, 13.24, 6.10	-2.40
	reduced Ce-bridged	1174, 1382, 1634	1219, 1405, 1631		13.98, 21.78, 14.47	-1.92
111	oxidized Non Ce- or O-bridged	1172, 1384, 1623	1217, 1407, 1620		10.47, 15.98, 11.58	-2.95
		1170, 1390, 1616	1214, 1414, 1613	1216, 1392, 1611 [b]	11.10, 18.79, 8.30	-3.04
	reduced Non Ce- or O-bridged	1173, 1381, 1599	1218, 1404, 1596	1218, 1413, 1599 [a]	10.08, 15.57, 11.60	-1.28

[a] ref [35]. [b] ref [33].

30

3.1.2.5 Bridged carbonate configuration

The bridged carbonate configuration closely resembles the BDC configuration, with the distinction that the plane is composed of two surface oxygen atoms and carbon and oxygen atoms of CO arranged perpendicular to the surface CeO_2. This specific arrangement is predominantly observed on CeO_2 (100) and (110) slabs. O-bridged and Ce-bridged sites, like BDC, were explored for this configuration. Both O-bridged and Ce-bridged sites exist on the CeO_2 (110) slab, but only the O-bridged site exists on the CeO_2 (100) slab. Among them, the O-bridging site on (110) has two possible nearest-neighbor oxygen vacancies to be considered (Fig. 9d). Table 6 shows that the frequency of bridged carbonate is greater than the BDC, around (1644 - 1711) cm^{-1}, which is consistent with experimental findings [33-35,39]. Furthermore, E_{ads} shows that CO interaction on CeO_2 (110) surface prefers O-bridge configuration over Ce-bridge configuration, which is consistent with Huang *et al.* [45]. Table 6 shows the E_{ads}, vibrational frequencies, and intensities for this configuration.
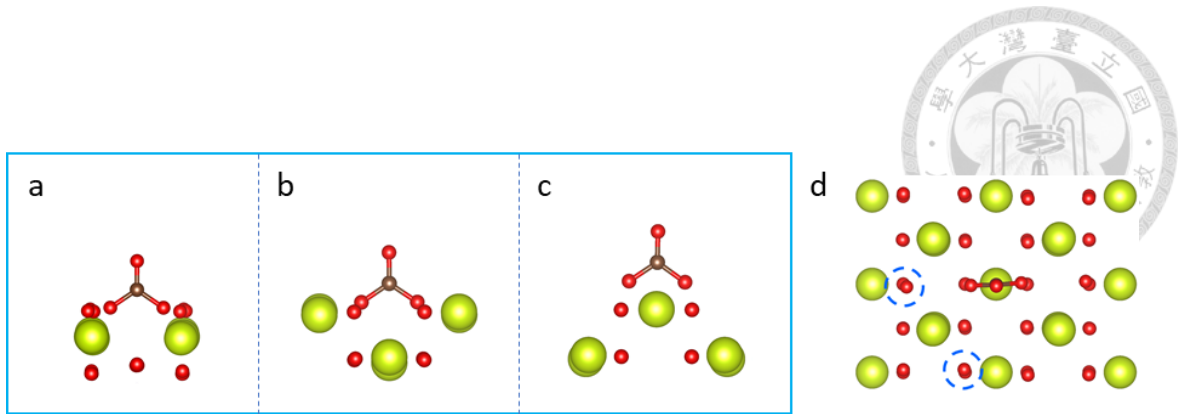
Fig. 9. (Side view) Bridged carbonate configuration of CO on (a) oxid-CeO$_2$ (100), (b) O-bridged site of oxid-CeO$_2$ (110), (c) Ce-bridged site of oxid-CeO$_2$ (110), and (d) two possible oxygen vacancies position (blue dotted circle) of O-bridged site on redu-CeO$_2$ (110).
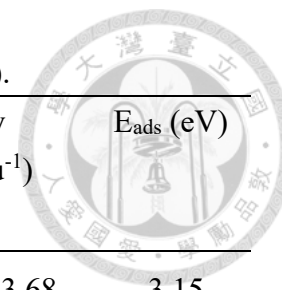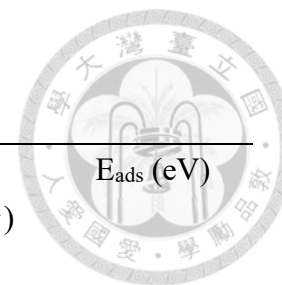
Table 6. Calculated vibrational frequencies, intensities, and corresponding E_{ads} of bridged carbonate on CeO$_2$ (100) and (110).

Facet	site		Frequency (cm^{-1})			Intensity (D/Å^2 amu^{-1})	E_{ads} (eV)
			DFT	After scaling	Experimental		
100	oxidized	O-bridged	1702	1734	1728 [a]	25.33	-3.56
	reduced	O-bridged	1711	1744	1736 [b]	27.51	-3.17
110	oxidized	O-bridged	1657	1688	1695 [c]	19.21	-3.63
		Ce-bridged	1655	1686	1695 [c]	23.43	-1.74
	reduced	O-bridged	1644	1675	1675 [d]	22.91	-3.73
			1651	1682		19.67	-3.18
		Ce-bridged	1705	1737	1736 [b]	27.52	-1.22

[a] ref [34]. [b] ref [35]. [c] ref [33]. [d] ref [39].

3.2 Assessment of DNN model

We employed various datasets produced from DFT and experimental IR spectrum to test the performance of our trained models on the prediction of distributions of CeO_2 facets, CO-derived adspecies, and E_{ads}, the details of which are discussed below.

3.2.1 DFT-based dataset

Around 200 complex IR spectra were produced at random and utilized as a testing dataset. R^2 and Wasserstein (W) were utilized as evaluation measures to examine the performance of the DNN models on the prediction of the distribution of oxid/redu-CeO_2 (100, 110, 111), CO-derived adspecies, and E_{ads}. As listed in Table 7, all three DNN models produce predictions with high R^2 and low W values, suggesting the models performed well on the prediction tasks. The predicted distributions were compared to the actual ones as shown in Fig. 10. The values of R^2 and root mean square error (RMSE) show that the predicted and actual values are in good agreement.

Our model can capture the majority of the IR spectral characteristics that correspond to CeO_2 facets (Fig. 10a-f), CO-derived adspecies (Fig. 10g-k), and E_{ads} groups (Fig. 10l-p). The prediction of CO-derived adspecies has demonstrated the highest accuracy, which is due to the fact that each surface species has a distinct frequency range, and some species even have two or three vibrational peaks, which can enable the model to recognize.

34
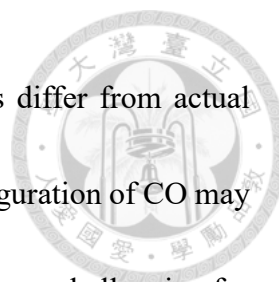
However, in the case of CeO$_2$ facets, some of the predicted values differ from actual values. This could be due to the fact that a particular adsorption configuration of CO may have similar IR features on various CeO$_2$ surfaces. As a result, it becomes challenging for the model to determine the exact CeO$_2$ surface based only on the IR spectra of that adsorption configuration. More discussions on this point can be found in the subsection below.

Table 7. Performance summary of the DNN models. R^2 and W obtained by testing 200 synthetic complex IR spectra.

Prediction Task	R^2	W
CeO$_2$ facets	0.97	0.026
CO-derived adspecies	0.99	0.0079
E$_{ads}$	0.98	0.016
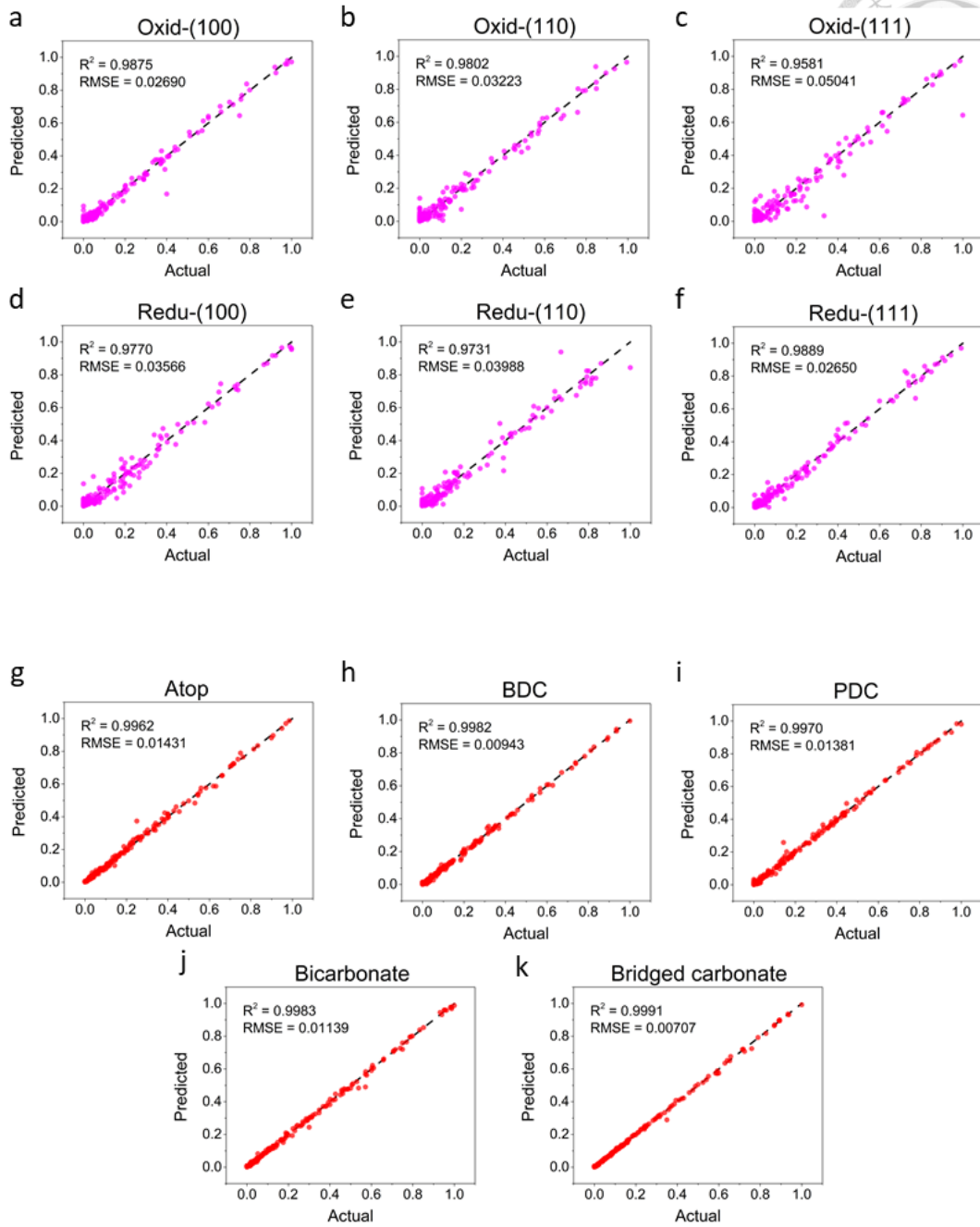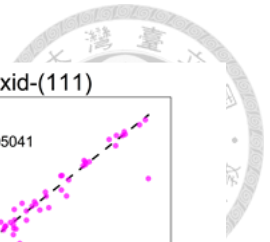
Fig. 10. Parity plots of the predicted and actual percentages of (a-f) CeO_2 facets, (g-k) CO-derived adspecies, and (l-p) E_{ads} composing the synthetic complex IR spectra in the test sets.

3.2.2 Experimental IR spectrum

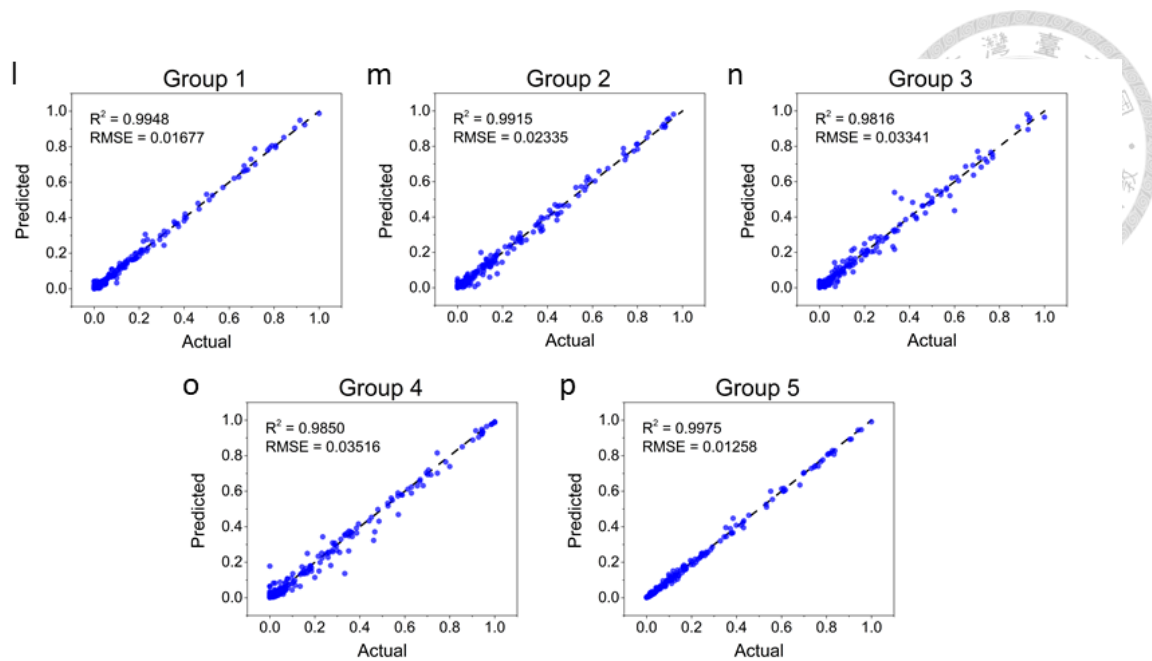To assess whether the DNN models capture the essential aspects of real-world IR spectra of CO adsorbed on CeO_2, we applied them to analyze a few experimental IR spectra derived from the literature [40,49]. First, we applied our model to the digitized [80] experimental spectra of CO adsorbed on four different single crystal CeO_2 (namely oxidized_110, reduced_110, oxidized_111, and reduced_111) to examine whether the predicted facets agree with the actual crystal structures. From the literature discussion [49], CO was in atop configuration with an adsorption energy of about -0.3 eV in all four infrared spectra. Fig. S2 shows that our model predictions of CO-derived adspecies and E_{ads} are in atop site and group 5 (-0.94 to 0 eV), which agree with the experimental findings. As shown in Fig. 11c-e, our model accurately predicted a high percentage of the correct facet (0.81, 0.65, and 0.80) in the case of oxidized_111, reduced_110, and reduced_111 CeO_2, respectively. However, for the oxidized_110 spectrum (Fig. 11b), the model misidentified a high proportion of reduced CeO_2 (110) and (100) facets. We believe this is because the frequency of the atop site of oxid-CeO_2 (110) (2171 cm^{-1}) is quite similar to redu-CeO_2 (110) and (100) (2175 and 2168 cm^{-1}), as listed in Table 2. Therefore, it is difficult for the model to distinguish between these CeO_2 surfaces based solely on the IR spectra of the atop site.
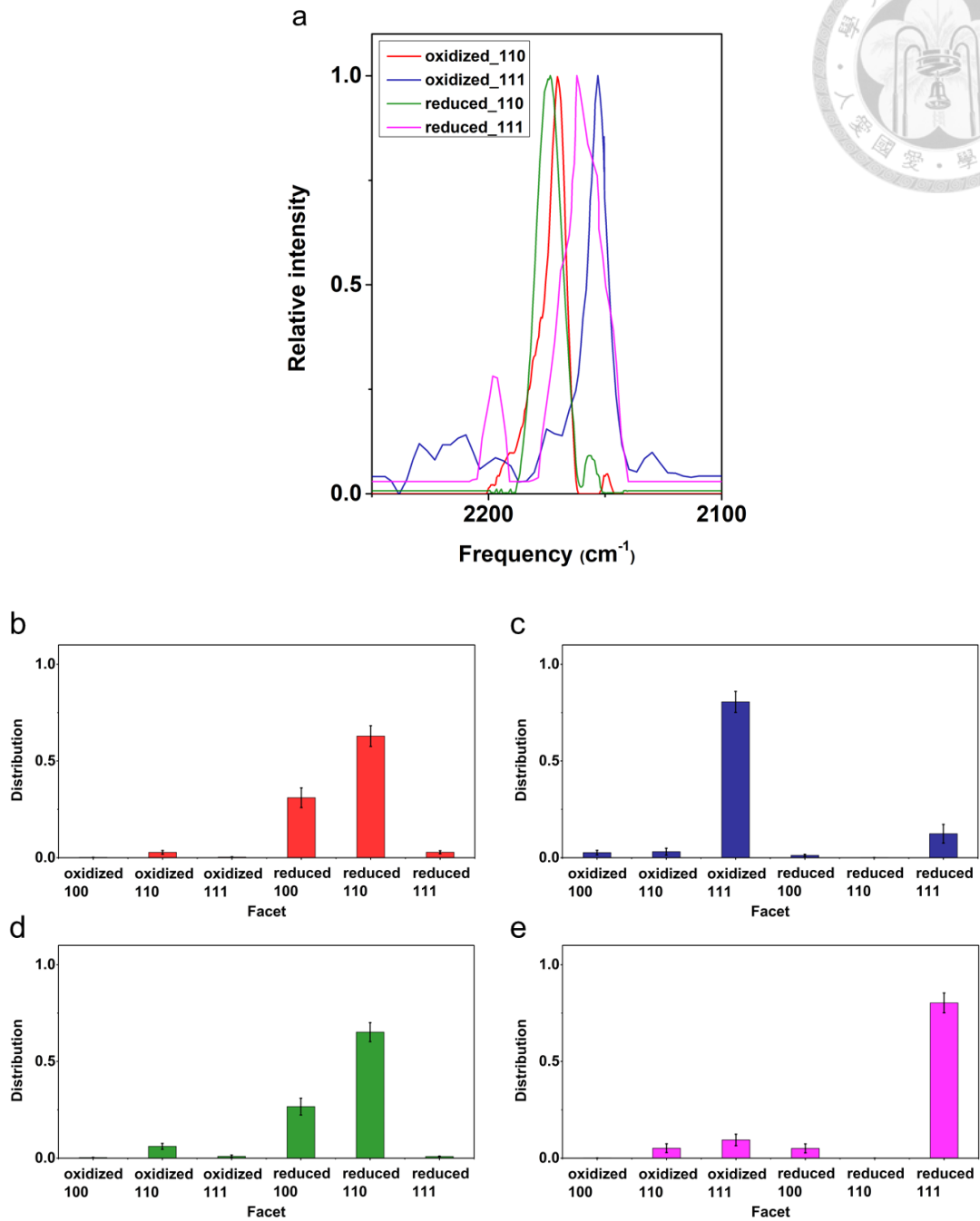
Fig. 11. Predicted facets of four CeO_2 single crystals. (a) The experimental IR spectra of

CO adsorption on oxidized_110 (red), oxidized_111 (navy), reduced_110 (green), and

reduced_111 (magenta) CeO_2 single crystals. Predicted facets using IR spectra of CO on

(b) oxidized_110, (c) oxidized_111, (d) reduced_110, and (e) reduced_111 CeO_2.

39

In addition to the single crystal spectrum, we also examined the performance of the model on CeO$_2$ nanorods. Yang *et al.* recorded an interesting experimental IR spectrum of CO adsorbed CeO$_2$ single crystals using infrared reflection absorption spectroscopy (IRRAS), and they discovered the presence of a high number of {111}-type nanofacets on the (110) planes of active CeO$_2$ nanorods. We applied our model to their IR spectrum and quantified the distribution of CeO$_2$ facets. As shown in Fig. 12b, our model predicted a substantial proportion of (111) and a smaller percentage of (110) facets, which is consistent with the qualitative findings from the experimental observations.
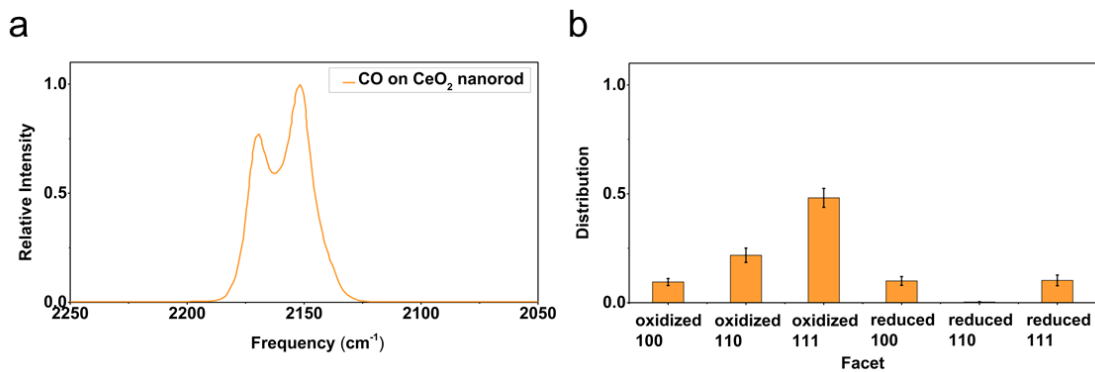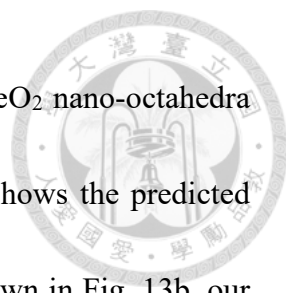
Fig. 12. (a) Experimental IR spectrum of CO on CeO$_2$ nanorods. (b) Predicted distribution of CeO$_2$ facets.

Finally, we used the complex IR spectra of CO adsorbed on CeO_2 nano-octahedra (o-CeO_2) recorded by Chen *et al.* [40] to test our model. Fig. 13 shows the predicted distributions of CeO_2 facets, CO-derived adspecies, and E_{ads}. As shown in Fig. 13b, our model predicted that the main exposed facet of o-CeO_2 is the (111) surface, which is consistent with the HRTEM experimental result of Chen *et al.* Furthermore, by utilizing IR peak assignments that were reported in the literature, Chen *et al.* established the adsorption of CO onto o-CeO_2 in the atop, BDC, and bicarbonate configurations. Based on our predictions of CO-derived adspecies (Fig. 13c), we estimate that around 0.83 of CO produces BDC, 0.16 at the atop site, and 0.012 for bicarbonate species, which is consistent with the analysis of Chen *et al.* Furthermore, the model also predicts that the E_{ads} is primarily dominated by group 4 (-2.27 to -0.94 eV). This finding aligns with the calculated E_{ads} obtained from the adsorption of BDC on the (111) surface, as listed in Table 3.

Fig. 13. Predicted facets and microstructure of CeO_2 octahedra. (a) Experimental IR spectra of CO on CeO_2 octahedra (o-CeO_2). Predicted distribution of (b) CeO_2 facets, (c) CO-derived adspecies, and (d) E_{ads} groups.
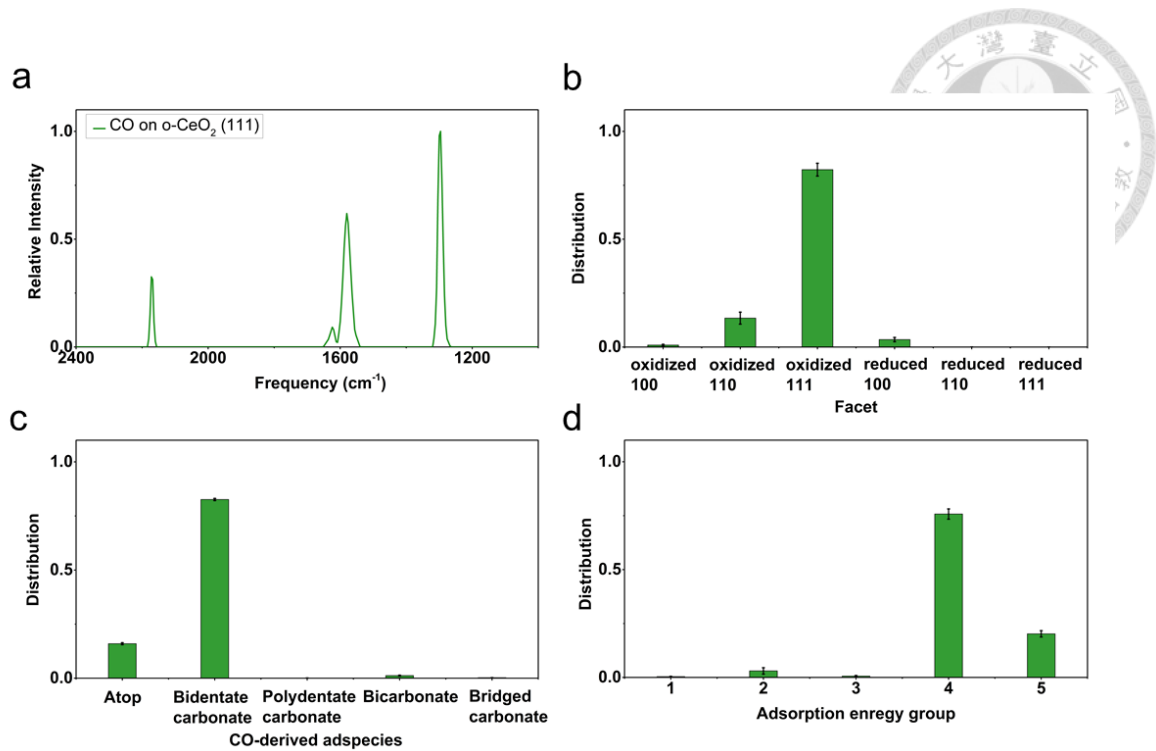
4. Conclusions

In this work, we utilized a deep learning approach along with an IR spectrum of CO to investigate the surface properties of CeO_2 catalysts. Using DFT calculations, we systematically examined the possible CO-derived adspecies on different CeO_2 facets, including atop, BDC, PDC, bridged carbonate, and bicarbonate configurations. Normal mode analysis was performed for each adspecies, and the resulting computed frequencies were combined with experimental data to create an extensive dataset containing vibrational frequencies, intensities, and binding energies of CO adsorbed on CeO_2. By utilizing this dataset, we generated synthetic IR spectra to develop deep learning models for predicting surface structure. The models estimate the distributions of CeO_2 facets, CO-derived adspecies, and binding energies based on IR spectrum data. We successfully applied the models to analyze experimental IR spectra of CO adsorbed on various types of CeO_2, including single crystal, nanorod, and nano-octahedron. We believe that our proposed method can assist in overcoming the challenges of assigning vibrational frequencies and determining the distribution of CeO_2 facets. This deep learning approach also has the potential for understanding the morphology, local environmental arrangement, interaction behavior of other probe molecules, e.g., CO_2 and NO, and catalytic characteristics of diverse CeO_2 facets.

Reference

[1] A. Trovarelli, Catalytic properties of ceria and CeO2-containing materials. Catal. Rev. 38 (1996) 439-520.

[2] H. C. Yao, Y. Y. Yao, Ceria in automotive exhaust catalysts: I. Oxygen storage. J. catal. 86 (1984) 254-265.

[3] Q. Fu, H. Saltsburg, M. Flytzani-Stephanopoulos, Active nonmetallic Au and Pt species on ceria-based water-gas shift catalysts. Science 301 (2003) 935-938.

[4] D. Andreeva, V. Idakiev, T. Tabakova, L. Ilieva, P. Falaras, A. Bourlinos, A. Travlos, Low-temperature water-gas shift reaction over Au/CeO2 catalysts. Catal. Today 72 (2002) 51-57.

[5] N. Kumari, M. A. Haider, M. Agarwal, N. Sinha, S. Basu, Role of reduced CeO2 (110) surface for CO2 reduction to CO and methanol. J. Phys. Chem. C 120 (2016) 16626-16635.

[6] G. Spezzati, A. D. Benavidez, A. T. DeLaRiva, Y. Su, J. P. Hofmann, S. Asahina, E. J. Olivier, J. H. Neethling, J. T. Miller, A. K. Datye, E.J.M. Hensen, CO oxidation by Pd supported on CeO2 (100) and CeO2 (111) facets. Appl. Catal. B: Environ. 243 (2019) 36-46.

[7] D. Widmann, R. Leppelt, R. J. Behm, Activation of a Au/CeO2 catalyst for the CO oxidation reaction by surface oxygen removal/oxygen vacancy formation. J. Catal.

251 (2007) 437-442.

[8] Y. Gao, W. Wang, S. Chang, W. Huang, Morphology effect of CeO2 support in the preparation, metal–support interaction, and catalytic performance of Pt/CeO2 catalysts. ChemCatChem 5 (2013) 3610-3620.

[9] M. Cargnello, V. V. Doan-Nguyen, T. R. Gordon, R. E. Diaz, E. A. Stach, R. J. Gorte, P. Fornasiero, C. B. Murray, Control of metal nanocrystal size reveals metal-support interface role for ceria catalysts. Science 341 (2013) 771-773.

[10] W. Xu, Z. Liu, A. C. Johnston-Peck, S. D. Senanayake, G. Zhou, D. Stacchiola, E. A. Stach, J. A. Rodriguez, Steam reforming of ethanol on Ni/CeO2: reaction pathway and interaction between Ni and the CeO2 support. ACS Catal. 3 (2013) 975-984.

[11] J. H. Lee, D. Y. Jo, J. W. Choung, C. H. Kim, H. C. Ham, K. Y. Lee, Roles of noble metals (M= Ag, Au, Pd, Pt and Rh) on CeO2 in enhancing activity toward soot oxidation: Active oxygen species and DFT calculations. J. hazard. mater. 403 (2021) 124085.

[12] C. Slostowski, S. Marre, P. Dagault, O. Babot, T. Toupance, & C. Aymonier, "CeO2 nanopowders as solid sorbents for efficient CO2 capture/release processes." J. CO2 Util. 20 (2017) 52-58.

[13] S. Sun, C. Zhang, S. Chen, X. Zhao, Y. Wang, S. Xu, C. Wu, "Integrated CO2 capture

and reverse water–gas shift reaction over CeO2-CaO dual functional materials." Royal Society Open Science 10.4 (2023): 230067.

[14] P. Li, B. Wang, C. Qin, C. Han, L. Sun, & Y. Wang, "Band-gap-tunable CeO2 nanoparticles for room-temperature NH3 gas sensors." Ceram. Int. 46.11 (2020) 19232-19240.

[15] W. Xie, B. Liu, S. Xiao, H. Li, Y. Wang, D. Cai, ... & T. Wang, "High performance humidity sensors based on CeO2 nanoparticles." Sens. Actuators B: Chem. 215 (2015) 125-132.

[16] R. Magudieshwaran, J. Ishii, K. C. N. Raja, C. Terashima, R. Venkatachalam, A. Fujishima, & S. Pitchaimuthu, "Green and chemical synthesized CeO2 nanoparticles for photocatalytic indoor air pollutant degradation." Mater. Lett. 239 (2019) 40-44.

[17] A. Muthuvel, M. Jothibas, V. Mohana, & C. Manoharan, "Green synthesis of cerium oxide nanoparticles using Calotropis procera flower extract and their photocatalytic degradation and antibacterial activity." Inorg. Chem. Commun. 119 (2020) 108086.

[18] S. Pansambal, R. Oza, S. Borgave, A. Chauhan, P. Bardapurkar, S. Vyas, & S. Ghotekar, Bioengineered cerium oxide (CeO2) nanoparticles and their diverse applications: a review. Appl. Nanosci. (2022) 1-26.

[19] A. Pfau, K. D. Schierbaum, The electronic structure of stoichiometric and reduced CeO2 surfaces: an XPS, UPS and HREELS study. Surf. Sci. 321 (1994) 71-80.

[20] E. Bêche, P. Charvin, D. Perarnau, S. Abanades, G. Flamant, Ce 3d XPS investigation of cerium oxides and mixed cerium oxide (CexTiyOz). Surf. Interface Anal. 40 (2008) 264-267.

[21] Y. Huang, C. F. Yan, C. Q. Guo, Y. Shi, Experimental and first-principles DFT study on oxygen vacancies on cerium dioxide and its effect on enhanced photocatalytic hydrogen production. Int. J. Hydrog. Energy 41 (2016) 7919-7926.

[22] F. Zhang, P. Wang, J. Koberstein, S. Khalid, S. W. Chan, Cerium oxidation state in ceria nanoparticles studied with X-ray photoelectron spectroscopy and absorption near edge spectroscopy. Surf. Sci. 563 (2004) 74-82.

[23] L. Qiu, F. Liu, L. Zhao, Y. Ma, J. Yao, Comparative XPS study of surface reduction for nanocrystalline and microcrystalline ceria powder. Appl. Surf. Sci. 252 (2006) 4931-4935.

[24] C. Yang, F. Bebensee, J. Chen, X.Yu, A. Nefedov, C. Wöll, Carbon dioxide adsorption on CeO2 (110): An XPS and NEXAFS study. ChemPhysChem 18 (2017) 1874-1880.

[25] H. Nakamatsu, T. Mukoyama, Assignment of Ce XANES spectra for CeO2 and CeO1.75 and effect of oxygen vacancy. Quantum Chem. 37 (2000) 111– 125.

[26] G. B. Della Mea, L. P. Matte, A. S. Thill, F. O. Lobato, E. V. Benvenutti, L. T. Arenas, ..., F. Bernardi, Tuning the oxygen vacancy population of cerium oxide

(CeO2− x, 0< x< 0.5) nanoparticles. Appl. Surf. Sci. 422 (2017) 1102-1112

[27] S. Y. Chen, C. H. Tsai, M. Z. Huang, D. C. Yan, T. W. Huang, A. Gloter, C. L. Chen, H.J. Lin, C. T. Chen, C. L. Dong, Concentration dependence of oxygen vacancy on the magnetism of CeO2 nanoparticles. J. Phys. Chem. C 116 (2012) 8707-8713.

[28] S. Y. Chen, Y. H. Lu, T. W. Huang, D. C. Yan, C. L. Dong, Oxygen vacancy dependent magnetism of CeO2 nanoparticles prepared by thermal decomposition method. J. Phys. Chem. C 114 (2010) 19576-19581.

[29] Z. L. Wang, X. Feng, Polyhedral shapes of CeO2 nanoparticles. J. Phys. Chem. B 107 (2003) 13563-13566.

[30] C.Hu, Z. Zhang, H. Liu, P. Gao, Z. L. Wang, Direct synthesis and structure characterization of ultrafine CeO2 nanoparticles. Nanotechnology 17 (2006) 5983.

[31] R. K. Hailstone, A. G. DiFrancesco, J. G. Leong, T. D. Allston, K. J. Reed, A study of lattice expansion in CeO2 nanoparticles by transmission electron microscopy. J. Phys. Chem. C 113 (2009) 15155-15159.

[32] Y. Lin, Z. Wu, J. Wen, K. R. Poeppelmeier, L. D. Marks, Imaging the atomic surface structures of CeO2 nanoparticles. Nano lett. 14 (2014) 191-196.

[33] Z. Wu, M. Li, S. H. Overbury, On the structure dependence of CO oxidation over CeO2 nanocrystals with well-defined surface planes. J. Catal. 285 (2012) 61-73.

[34] C. Li, Y. Sakata, T. Arai, K. Domen, K. I. Maruya, T. Onishi, Carbon monoxide and

carbon dioxide adsorption on cerium oxide studied by Fourier-transform infrared spectroscopy. Part 1.—Formation of carbonate species on dehydroxylated CeO_2, at room temperature. J. Chem. Soc., Faraday trans. I: Physical Chemistry in Condensed Phases 85 (1989) 929-943.

[35] C. Binet, M. Daturi, J. C. Lavalley, IR study of polycrystalline ceria properties in oxidised and reduced states. Catal. Today 50 (1999) 207-225

[36] C. Binet, A. Badri, M. Boutonnet-Kizling, J. C. Lavalley, FTIR study of carbon monoxide adsorption on ceria: CO 2–2 carbonite dianion adsorbed species. J. Chem. Soc., Faraday trans.   90 (1994) 1023-1028.

[37] Z. J. Gong, Y. R. Li, H. L. Wu, S. D. Lin, W. Y. Yu, Direct copolymerization of carbon dioxide and 1, 4-butanediol enhanced by ceria nanorod catalyst. Appl. Catal. B: Environ. 265 (2020) 118524.

[38] K. Yoshikawa, H. Sato, M. Kaneeda, J. N. Kondo, Synthesis and analysis of CO2 adsorbents based on cerium oxide. J. CO2 Util. 8 (2014) 34-38.

[39] Z. Wu, A. K. Mann, M. Li, S. H. Overbury, Spectroscopic investigation of surface-dependent acid–base property of ceria nanoshapes. J. Phys. Chem. C 119 (2015) 7340-7350.

[40] S. Chen, T. Cao, Y. Gao, D. Li, F. Xiong, W. Huang, Probing surface structures of CeO2, TiO2, and Cu2O nanocrystals with CO and CO2 chemisorption. J. Phys.

Chem. C 120 (2016) 21472-21485.

[41] B. Herschend, M. Baudin, K. Hermansson, CO adsorption on CeO2 (1 1 0) using hybrid-DFT embedded-cluster calculations. Chem. phys. 328 (2006) 345-353.

[42] M. Nolan, S. C. Parker, G. W. Watson, Vibrational properties of CO on ceria surfaces. Surf. sci. 600 (2006) 175-178.

[43] Z. Cheng, B. J. Sherman, C. S. Lo, Carbon dioxide activation and dissociation on ceria (110): A density functional theory study. J. Chem. Phys. 138 (2013) 014702.

[44] P. G. Lustemberg, P. N. Plessow, Y. Wang, C. Yang, A. Nefedov, F. Studt, C. Wöll, M. V. Ganduglia-Pirovano, Vibrational frequencies of cerium-oxide-bound CO: A challenge for conventional DFT methods. Phys. Rev. Lett. 125 (2020) 256101.

[45] M. Huang, S. Fabris, CO adsorption and oxidation on ceria surfaces from DFT+ U calculations. J. Phys. Chem. C 112 (2008) 8643-8648.

[46] Z. Yang, T. K. Woo, K. Hermansson, Strong and weak adsorption of CO on CeO2 surfaces from first principles calculations. Chem. phys. lett. 396 (2004) 384-392.

[47] G. N. Vayssilov, M. Mihaylov, P. S. Petkov, K. I. Hadjiivanov, K. M. Neyman, Reassignment of the vibrational spectra of carbonates, formates, and related surface species on ceria: a combined density functional and infrared spectroscopy investigation. J. Phys. Chem. C 115 (2011) 23435-23454.

[48] C. Yang, L. L. Yin, F. Bebensee, M. Buchholz, H. Sezen, S. Heissler, J.Chen, A.
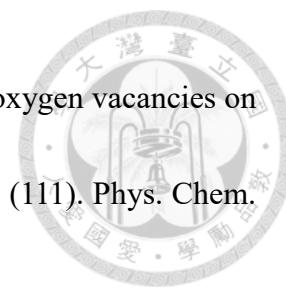
Nefedov, H. Idriss, X. Q. Gong, C. Wöll, Chemical activity of oxygen vacancies on ceria: a combined experimental and theoretical study on CeO 2 (111). Phys. Chem. Chem. Phys. 16 (2014) 24165-24168.

[49] C. Yang, X. Yu, S. Heißler, A. Nefedov, S. Colussi, J. Llorca, A. Trovarelli, Y. Wang, C. Wöll, Surface faceting and reconstruction of ceria nanoparticles. Angew. Chem. Int. Ed. 56 (2017) 375-379.

[50] C. Yang, X. Yu, P. N. Pleßow, S. Heißler, P. G. Weidler, A. Nefedov, F. Studt, Y. Wang, C. Wöll, Rendering photoreactivity to ceria: The role of defects. Angew. Chem. Int. Ed. 56 (2017) 14301-14305

[51] C. Yang, M. Capdevila-Cortada, C. Dong, Y. Zhou, J. Wang, X. Yu, A. Nefedov, S. Heißler, N. López, W. Shen, C. Wöll, Y. Wang, Surface refaceting mechanism on cubic ceria. J. Phys. Chem. Lett. 11 (2020) 7925-7931.

[52] J. Timoshenko, D. Lu, Y. Lin, A. I. Frenkel, Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles. J. Phys. Chem. Lett. 8 (2017) 5091-5098.

[53] G. Drera, C. M. Kropf, L. Sangaletti, Deep neural network for x-ray photoelectron spectroscopy data analysis. Mach. Learn.: Sci. Technol. 1 (2020) 015008.

[54] C. Zheng, C. Chen, Y. Chen, S. P. Ong, Random forest models for accurate identification of coordination environments from X-ray absorption near-edge
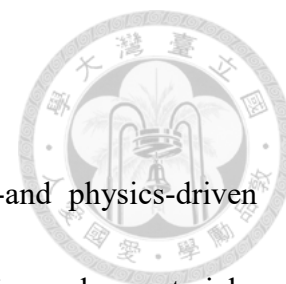
structure. Patterns 1 (2020) 100013.

[55] J. L. Lansford, D. G. Vlachos, Infrared spectroscopy data-and physics-driven machine learning for characterizing surface microstructure of complex materials. Nat. commun. 11 (2020) 1-12.

[56] G. Kresse, J. Hafner, Ab initio molecular dynamics for liquid metals. Phys. rev. B 47 (1993) 558.

[57] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput. mater. sci. 6 (1996) 15-50.

[58] P. E. Blöchl, Projector augmented-wave method. Phys. rev. B 50 (1994) 17953.

[59] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. Phys. rev. lett. 77 (1996) 3865.

[60] M. Nolan, S. Grigoleit, D. C. Sayle, S. C. Parker, G. W. Watson, Density functional theory studies of the structure and electronic structure of pure and defective low index surfaces of ceria. Surf. Sci. 576 (2005) 217-229.

[61] A. Beste, S. H. Overbury, Pathways for ethanol dehydrogenation and dehydration catalyzed by ceria (111) and (100) surfaces. J. Phys. Chem. C 119 (2015) 2447-2455.

[62] H. T. Chen, Y. Choi, M. Liu, M. C. Lin, A first-principles analysis for sulfur tolerance of CeO_2 in solid oxide fuel cells. J. Phys. Chem. C 111 (2007) 11117-11122.

[63] K. Yoshikawa, M. Kaneeda, H. Nakamura, Development of Novel CeO_2-based CO_2 adsorbent and analysis on its CO_2 adsorption and desorption mechanism. Energy Procedia 114 (2017) 2481-2487.
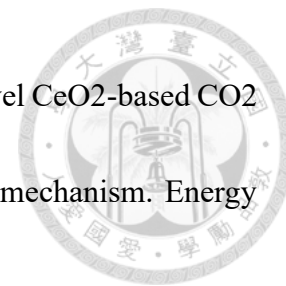
[64] T. A. Manz, N. G. Limas, Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. RSC adv. 6 (2016) 47771-47801.

[65] T. A. Manz, N. G. Limas, Chargemol program for performing DDEC analysis. http://ddec. sourceforge. net. (2017).

[66] D. Porezag, M. R. Pederson, Infrared intensities and Raman-scattering activities within density-functional theory. Phys. Rev. B 54 (1996) 7830.

[67] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, ... & K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms. J. Phys.: Condens. Matter 29 (2017) 273002.
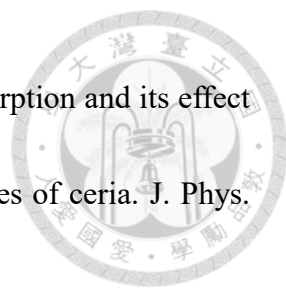
[68] G. K. Wertheim, M. A. Butler, K. A. West, D. N. E. Buchanan, Determination of the Gaussian and Lorentzian content of experimental line shapes. Rev. Sci. Instrum. 45 (1974) 1369-1371.

[69] E. Levina, P. Bickel,. The earth mover's distance is the mallows distance: Some insights from statistics. In Proc. Eighth IEEE Int. Conf. Comput. Vis. ICCV 2 (2001). 251-256.

[70] Y. P. Li, K. Han, C. A. Grambow, W. H. Green, Self-evolving machine: A
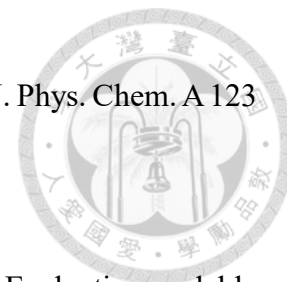
continuously improving model for molecular thermochemistry. J. Phys. Chem. A 123 (2019) 2142-2152.

[71] G. Scalia, C. A. Grambow, B. Pernici, Y. P. Li, W. H. Green, Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J. chem. inf. model. 60 (2020: 2697-2717.

[72] C. I. Yang, Y. P. Li, Explainable uncertainty quantifications for deep learning-based molecular property prediction. J. Cheminformatics 15 (2023) 13.

[73] H. J. Whitfield, D. Roman, A. R. Palmer, X-ray study of the system ThO2-CeO2-Ce2O3. J. Inorg. Nucl. Chem. 28 (1966) 2817-2825.

[74] P. W. Tasker, The stability of ionic crystal surfaces. J. Phys. C: Solid State Phys. 12 (1979) 4977.

[75] M. Capdevila-Cortada, N. López, Entropic contributions enhance polarity compensation for CeO2 (100) surfaces. Nat. mater. 16 (2017) 328-334.

[76] Y. Pan, N. Nilius, C. Stiehler, H. J. Freund, J. Goniakowski, C. Noguera, Ceria nanocrystals exposing wide (100) facets: Structure and polarity compensation. Adv. Mater. Interfaces 1 (2014) 1400404.

[77] H. Zhou, D. Wang, X. Q. Gong, Clarifying the impacts of surface hydroxyls on CO oxidation on CeO 2 (100) surfaces: a DFT+ U study. Phys. Chem. Chem. Phys. 22 (2020) 7738-7746.

[78] 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 (2012) 7073-7082.

[79] G. A. Mekhemer, M. I. Zaki, Low-temperature IR spectroscopy of CO adsorption on calcined supported CeO2: Probing adsorbed species and adsorbing sites. Adsorpt. Sci. Technol. 15 (1997) 377-389.

[80] F. Chen, D. Liu, J. Zhang, P. Hu, X. Q. Gong, & G. Lu, "A DFT+ U study of the lattice oxygen reactivity toward direct CO oxidation on the CeO2 (111) and (110) surfaces." Phys. Chem. Chem. Phys. 14.48 (2012): 16573-16580.

[81] Y. L. Song, L. L. Yin, J. Zhang, P. Hu, X. Q. Gong, & G. Lu, "A DFT+ U study of CO oxidation at CeO2 (110) and (111) surfaces with oxygen vacancies." Surface science 618 (2013): 140-147.

[82] A. Rohatgi, WebPlotDigitalizer version 4.4. Austin, Texas, USA. https://automeris. io/WebPlotDigitizer, (2018).

Supporting Information

Scaling factor

To account for the discrepancy between experimental and computed frequencies, the frequencies derived from DFT were calibrated using the scaling procedure as proposed in the literature [1]. The equation for the scaling factor (SF) is given as,

$$SF = \frac{\sum_{i=1}^{n}(\nu_i * \omega_i)}{\sum_{i=1}^{n}\omega_i{}^2}$$

where n is the amount of data, ν_i is the experimental frequency, and ω_i is the frequency derived from DFT calculations. The scaling factors for each vibrational mode are listed in Table S1. For each CO-derived adspecies, the highest frequency is the first mode, the second highest frequency is the second mode, and so on.

Table S1. The scaling factor for each vibrational mode.

CO-derived adspecies	1st mode	2nd mode	3rd mode
Atop	1.018		
Bidentate carbonate	1.015	1.048	
Polydentate carbonate	1.027	0.993	
Bicarbonate	0.998	1.017	1.038
Bridged carbonate	1.019		

Hyperparameter optimization

As listed in Table S2, hyperparameters including number of training data per training set, number of training sets, number of hidden layers, number of neurons per layer, batch size, initial learning rate, epsilon, regularization loss (norm), and regularization parameter were tuned to find the optimal hyperparameter set. The optimal hyperparameters for the DNN models are listed in Table S3.

Table S2. Hyperparameter search space.

	Range
Hidden layers	2 or 3
Neuron	50~151
Regularization parameter	$10^{-1} \sim 10^{-6}$
Initial learning rate	$10^{-4} \sim 10^{-3}$
Batch size	$10 \sim 100$
Epsilon	$10^{-10} \sim 10^{-14}$
Regularization Loss (norm)	L1 or L2
Number of training data per training set	$500 \sim 5000$
Number of training sets	$100 \sim 1000$

Table S3. The optimum hyperparameter settings for the DNN models.

Model tasks / Hyperparameter	CeO$_2$ facet	CO-derived adspecies	Adsorption energy groups
Hidden layer structure	(91, 96)	(65, 123, 89)	(71, 148)
Regularization parameter	7.7×10^{-6}	4.1×10^{-5}	5.3×10^{-6}
Initial learning rate	4.5×10^{-4}	2.2×10^{-4}	2.7×10^{-4}
Batch size	62	51	76
Epsilon	7.8×10^{-13}	1.1×10^{-12}	1.9×10^{-12}
Regularization Loss (norm)	L1	L2	L1
Number of training data per training set	4665	2219	4590
Number of training sets	706	987	900

Impact of training size on model performance

By introducing noise to the frequency and intensity calculated by DFT and adjusting the composition of CO-adsorbed species and surface in the spectra, it is possible to generate a large amount of synthetic IR spectra. Our analysis, as shown in Fig. S1, demonstrates that training models for E_{ads} and CO-derived adspecies require at least 3280 data, whereas the model for the CeO_2 facet requires at least 2980 data to reach stable performance.
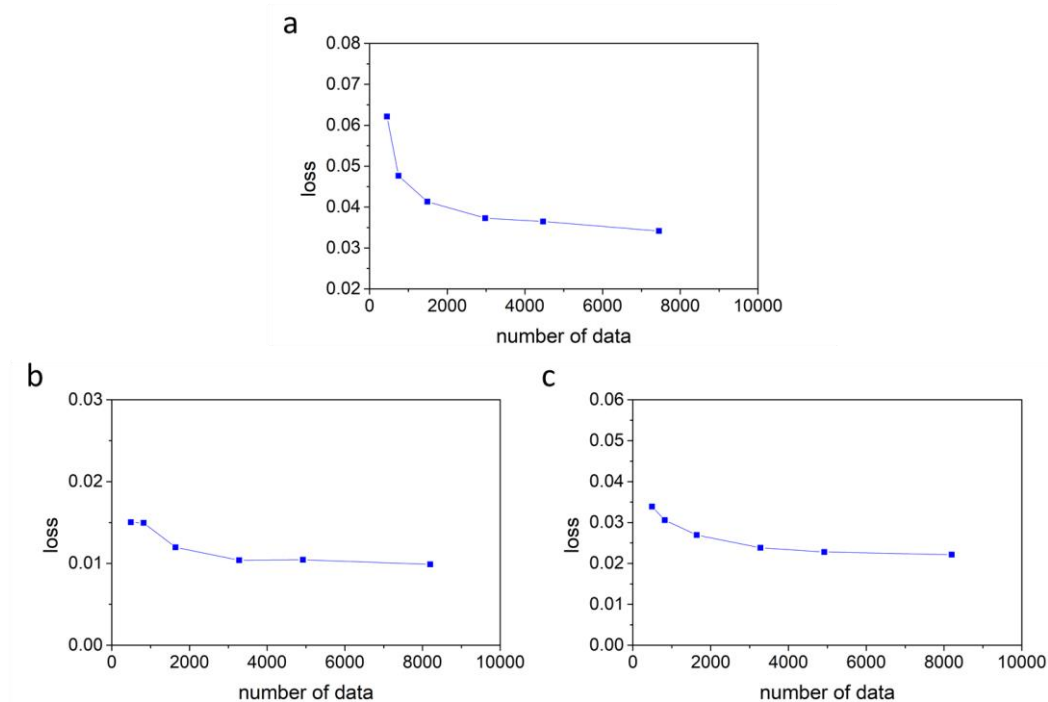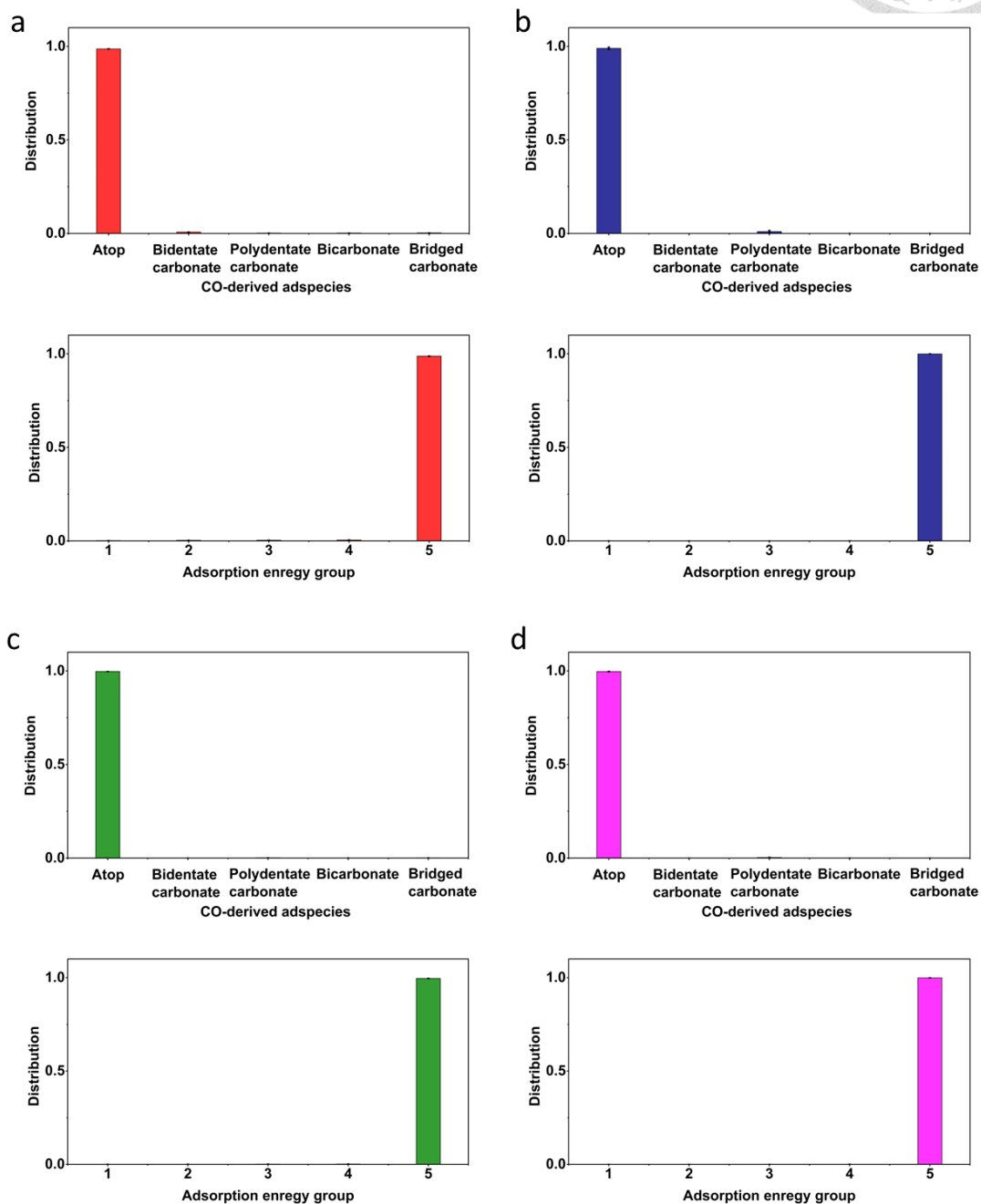

Fig. S1. The relationship between the number of training data and the accuracy of the model predictions for the distribution of (a) CeO_2 facets, (b) CO-derived adspecies, and (c) adsorption energy.

Predicted CO-derived adspecies and adsorption energy on IR shown in
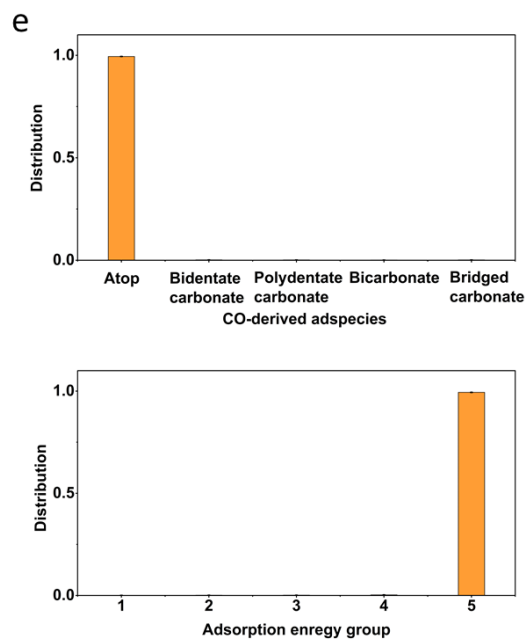
Fig. 10a

Fig. S2. The predicted distribution of CO-derived adspecies and adsorption energy on (a) oxidized_110, (b) oxidized_111, (c) reduced_110, and (d) reduced_111 CeO$_2$.