

Article

# The Role of Structural Representation in the Performance of a Deep Neural Network for X-ray Spectroscopy

Marwah M.M. Madkhali <sup>1,2</sup>, Conor D. Rankine <sup>1</sup>  and Thomas J. Penfold <sup>1,\*</sup>

<sup>1</sup> Chemistry—School of Natural and Environmental Sciences, Newcastle University, Newcastle-Upon-Tyne NE1 7RU, UK; m.m.madkhali2@newcastle.ac.uk (M.M.M.M.); conor.rankine@newcastle.ac.uk (C.D.R.)

<sup>2</sup> Department of Chemistry, College of Science, Jazan University, Jazan 45142, Saudi Arabia

\* Correspondence: tom.penfold@ncl.ac.uk

Academic Editor: Anan Yaghmur

Received: 5 May 2020; Accepted: 8 June 2020; Published: 11 June 2020



**Abstract:** An important consideration when developing a deep neural network (DNN) for the prediction of molecular properties is the representation of the chemical space. Herein we explore the effect of the representation on the performance of our DNN engineered to predict Fe K-edge X-ray absorption near-edge structure (XANES) spectra, and address the question: How important is the choice of representation for the local environment around an arbitrary Fe absorption site? Using two popular representations of chemical space—the Coulomb matrix (CM) and pair-distribution/radial distribution curve (RDC)—we investigate the effect that the choice of representation has on the performance of our DNN. While CM and RDC featurisation are demonstrably robust descriptors, it is possible to obtain a smaller mean squared error (MSE) between the target and estimated XANES spectra when using RDC featurisation, and converge to this state a) faster and b) using fewer data samples. This is advantageous for future extension of our DNN to other X-ray absorption edges, and for reoptimisation of our DNN to reproduce results from higher levels of theory. In the latter case, dataset sizes will be limited more strongly by the resource-intensive nature of the underlying theoretical calculations.

**Keywords:** machine learning; deep neural network; Coulomb matrix; radial distribution curve; X-ray absorption spectroscopy; XANES; K-edge

## 1. Introduction

Structural techniques, such as X-ray diffraction (XRD) and spectroscopy (XS), have made it possible to determine directly the structures of molecules and condensed matter systems and have had a huge influence across physics, chemistry, and biology. The proliferation of high-brilliance light sources such as 3<sup>rd</sup> generation synchrotrons and X-ray free-electron lasers (XFELs) is helping to increase this influence of these techniques by facilitating the measurement of increasingly challenging systems such as operating catalysts [1,2] and short-lived reaction intermediates [3,4].

Besides providing element- and site-specific information on geometric structure, X-ray absorption spectroscopy (XAS) is also able to provide direct information on the electronic structure around the absorption site [5,6]. Indeed, XAS spectra are characterised by X-ray absorption edges that correspond to the excitation of core electrons to the ionisation threshold. The electrons are initially excited to unoccupied or partially-occupied orbitals at energies just below the ionisation potential (IP); these bound transitions, which form the pre-edge spectral features, provide detailed information about the nature of the unoccupied valence orbitals. At energies above the IP, resonances occur due to

interference of the electron wave originating from the absorbing site with the electron wave scattered back from the neighbouring atoms. When the kinetic energy of the electrons is large, the scattering cross-section of the electrons is small, and information on the short-range geometric structure around the absorption site is accessible. This information is encoded in the extended X-ray absorption fine structure (EXAFS) region of the XAS spectrum, and is typically found  $>50$  eV above the X-ray absorption edge. EXAFS encodes information about coordination numbers and the distance of nearest neighbours to the absorption site. In contrast, the X-ray absorption near-edge structure (XANES) region of the XAS spectrum, found at lower photoelectron energies ( $<50$  eV above the X-ray absorption edge), is dominated by the interference of scattering pathways between multiple atoms. The XANES region encodes detailed information about coordination numbers, bond distances, and bond angles.

Extracting this information via analysis of the XAS spectrum demands detailed theoretical treatments [7,8]. The requisite theory for the analysis of EXAFS spectra is better developed than that required for the analysis of XANES spectra; the reason for this is that the higher-energy photoelectron is largely insensitive to the details and response of the potential [9,10]. The analysis of XANES spectra is more challenging and, despite significant progress [11–13], quantitative interpretation still necessitates resource-intensive theoretical treatments that can take hours/days to yield a single spectrum (cf. a few minutes for the calculation of an EXAFS spectrum). While the computational cost of such theoretical treatments is not prohibitive for a few calculations, it becomes so for the most challenging systems, e.g., disordered/amorphous systems, such as the surface of an operando catalyst [14–16], or dynamical processes [17–21]. A huge number of different and dynamically-evolving absorption sites must be modelled for a quantitative analysis, but to treat theoretically every possible chemically-inequivalent absorption site (or even to sample a meaningful number of such sites) is computationally challenging, resource-intensive, and time-consuming. This is a considerable challenge in the EXAFS analyses of such systems, which routinely necessitate molecular dynamics (MD) simulations with nuclear ensemble approaches [22] and, at present, an insurmountable one in many XANES analyses. Such analyses are out of reach for the majority of XAS users and, for the most complex systems, even expert theoreticians.

To address this, a number of recent works have explored supervised machine learning/deep learning algorithms with a view towards mapping the relationship between XANES spectra and the underlying electronic and geometric structure of materials [23–31]. In the majority of these recent works, the authors have focused on mapping the spectrum onto a property or structure, and these works have generally been system-specific or restricted to a narrow class of systems. Indeed, generally-applicable machine learning models capable of predicting XANES spectra for arbitrary absorption sites (i.e., machine learning models capable of learning ‘reverse’ mappings of the structure onto the spectrum) are lacking. We have addressed this by introducing a deep neural network (DNN) for the instantaneous prediction of Fe K-edge XANES spectra in a recent publication [32]. Our motivation is to accelerate the XANES analysis of complicated, disordered, and dynamically-evolving systems and, in these cases, it is unlikely that deriving a single structure from a XANES spectrum would prove useful or physically meaningful. In such cases, an accurate, high-throughput DNN for XANES simulation has great potential for revealing structure-spectrum relationships. Our DNN, which requires no input beyond the local geometric structure around an arbitrary Fe absorption site, is able to predict Fe K-edge XANES spectra with a mean squared error (MSE) below ca. 3% (evaluated on ca. 2000 unseen XANES spectra) after training with ca. 7000 theoretical examples [32]. We have demonstrated sub-eV accuracy on spectral peak positions and accuracy on peak positions that is an order of magnitude smaller than the spectral variations that our model was engineered to reproduce [32].

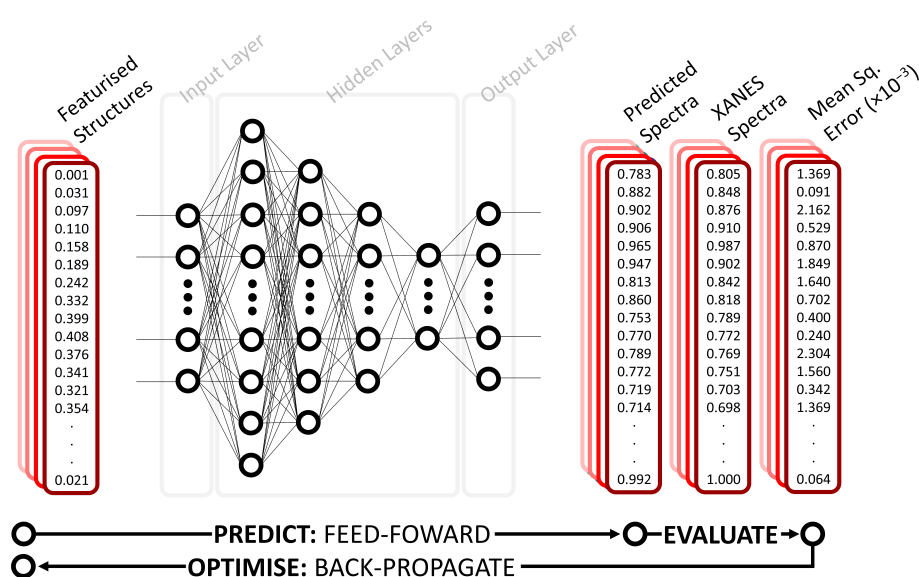
An important consideration when engineering a DNN for this task is the representation, or ‘featurisation’, of the input data: How should the local environment—the ‘chemical space’—around an arbitrary Fe absorption site be best represented to maximise the performance of the DNN? In this contribution, we explore the effect of representation on the performance of our DNN [32]. Using the Coulomb matrix (CM) [33–35] and pair-distribution/radial distribution curve (RDC) [36–39],

we demonstrate that the latter not only leads to smaller mean squared error (MSE) but also achieves this faster and with smaller training set sizes, which greatly supports the development of a DNN which is generalisable across the whole periodic table.

## 2. Theory and Computational Details

### 2.1. Deep Neural Network

A schematic of our DNN is shown in Figure 1. Our DNN is based on the multilayer perceptron model (MLP); an MLP is a class of feed-forward neural network comprising an input layer,  $n$  hidden layers, and an output layer. The dimensions of the input layer in our DNN are determined by the representation used (see the “Representation” section). The first hidden layer comprises 1200 neurons and every subsequent layer is reduced in size by 30% relative to the preceding hidden layer; our DNN uses four hidden layers. The output layer comprises  $X$  neurons, defined by the discretization of our target XANES spectra.



**Figure 1.** Schematic representation of the deep neural network (DNN) used in this work. The DNN takes the local environment around an atomic absorption site (featurised using either Coulomb matrix (CM) or radial distribution curve (RDC)) as input. This is passed through the network which consists of four hidden layers to output a predicted spectrum and mean squared error between the theoretical and predicted XANES spectra.

The layers are all fully connected, or ‘dense’, i.e., each neuron in an arbitrary layer,  $k$ , is connected to every neuron in the preceding layer,  $k - 1$ , via a matrix of weights,  $w_{i,j}^{(k)}$ . The pre-activation value of an arbitrary neuron in an arbitrary layer,  $z_{k,j}$ , is given by the linear combination of the input activations,  $x_{(k-1)}$  (these being the output activations of each neuron in the preceding layer), and their respective weights:  $z_{k,j} = \sum_i w_{ij}^{(k)} x_{(k-1),i}$ . A non-linear activation function,  $g(z)$ , is then applied to the pre-activation values to compute output activations for the neuron in the layer,  $\hat{y}_{k,j}$ ; our DNN uses a hyperbolic tangent (tanh) activation function which constrains the possible values of  $\hat{y}$  between  $-1$  and  $+1$ . The resulting activations, obtained for an arbitrary neuron in an arbitrary layer as  $\hat{y}_{k,j} = g(z_{k,j})$ , then serve as the input activations for the next layer, unless subject to an intermediate transformation. Information propagates through the MLP via a ‘feed-forward’ process until it arrives at the output layer. The activations of the output layer are then compared against target activations via evaluation of a cost function,  $J(\mathbf{W}) = \frac{1}{n} \sum_i \ell\{f(x^{(i)}, \mathbf{W}), y^{(i)}\}$ , that quantifies the difference between the obtained,  $f(x^{(i)}, \mathbf{W})$ , and expected,  $y^{(i)}$ , activations over a dataset of  $n$  samples as a function of the weights,  $\mathbf{W}$ ,

and input activations,  $x^{(i)}$ . Our DNN uses a mean-squared error (MSE) cost function of the general form  $J(\mathbf{W}) = \frac{1}{n} \sum_i^n \{y^{(i)} - f(x^{(i)}, \mathbf{W})\}$ .

The derivatives of  $J(\mathbf{W})$  with respect to the internal weights,  $\frac{\delta J(\mathbf{W})}{\delta \mathbf{W}}$ , can be calculated cost-effectively and used to adjust the internal weights such that  $J(\mathbf{W})$  is minimised; succinctly, the objective is to find a set of internal weights,  $\mathbf{W}^*$ , for which  $\mathbf{W}^* = \underset{\mathbf{W}}{\operatorname{argmin}} J(\mathbf{W})$ .

Our DNN optimizes c.a. three million internal weights via sequential feed-forward and back propagation cycles. Gradients of the MSE cost function with respect to the internal weights are updated iteratively according to the Adaptive Moment Estimation (ADAM) algorithm. Gradients are estimated over minibatches of 100 samples. The learning rate for the ADAM algorithm,  $\eta$ , is set to  $3 \times 10^{-4}$ .

The performance of our DNN is assessed via  $K$ -fold cross validation [40]. The data are randomly partitioned into  $K$  folds with  $K - 1$  folds kept in-sample to train the DNN and the remaining fold left out-of-sample to evaluate the performance of the DNN on unseen data.  $K$  evaluations are made such that every data sample appears in the out-of-sample testing set once, and in the in-sample training set  $K - 1$  times. The entire procedure can be repeated any number of times with different random  $K$ -fold partitions. The repeated evaluations of performance can be used to estimate an error. We mitigate the risk of overfitting our DNN to the training set by assessing the performance of each  $K$ -fold on the out-of-sample data only. We use five-fold cross-validation, i.e., an 80:20 in-sample/out-of-sample split, with five repetitions.

Our DNN also utilises dropout [41]; dropout is a regularisation technique that sets the activations of a certain fraction of the neurons in each layer are set to zero during the feed-forward/back-propagation procedure. Utilising dropout encourages a DNN to distribute weights probabilistically, and works to mitigate layers adapting to correct for mistakes in other layers; the latter behaviour would otherwise lead to overfitting, as these adaptations will not generalise well beyond the in-sample dataset. Our DNN uses a dropout of 15%.

Our DNN is programmed in Python 3 with the TensorFlow/Keras [42,43] API. All hyperparameters were determined via Bayesian optimisation using the GPyOpt [44,45] module, as in Reference [32].

## 2.2. Representation

We trial two alternative representations of chemical space: The CM and RDC.

The CM representation ( $M_{IJ}$ ) [33–35] is constructed as:

$$M_{IJ} = \begin{cases} \frac{1}{2} Z_I^2 Z_J^2 & \forall I = J \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \forall I \neq J \end{cases} \quad (1)$$

$Z$  is the nuclear charge of an atom and  $\mathbf{R}$  is the position of the atom in Cartesian space.  $M_{IJ}$  is a symmetric matrix of dimensions  $N \times N$  where  $N$  is the upper limit on the number of atoms designated for inclusion in the CM. The off-diagonal elements of  $M_{IJ}$  correspond to a Coulombic repulsion term between atoms  $I$  and  $J$  and the on-diagonal elements of  $M_{IJ}$  correspond to the Coulomb potentials of the free atoms. The rows of  $M_{IJ}$  are sorted in descending order according to their Euclidean ( $L^2$ ) norms,  $\|M_I\|$ , i.e., a permutation of the rows and columns is found that satisfies the inequality:

$$\|M_I\| \geq \|M_{I+1}\| \quad \forall I \quad (2)$$

The upper triangle of  $M_{IJ}$  is then taken and flattened row-wise to yield a feature vector of length  $\frac{1}{2}(N^2 + N)$ . Practicably, this feature vector should have the same length, regardless of the size of the system it encodes, if it is to be input into a neural network. If the system contains more than  $N$  atoms, the closest  $N$  atoms to the absorbing atom are used to construct the CM and the rest are discarded; if the system contains less than  $N$  atoms, the remaining rows and columns of the CM are zero-filled.

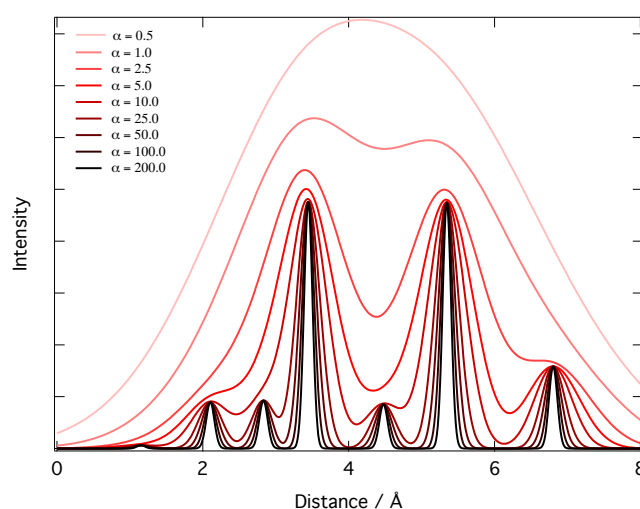
The sorted CM representation is unique, invariant with respect to atomic indexing, translations, and rotations of the chemical space that it describes, and its construction requires no explicit information on chemical bonding [33–35].

Where CM featurisation is used in this work,  $N = 20$ .

The RDC representation [36–39] encodes local chemical space as an intensity distribution ( $f_{RDC}$ ) as a function of equally-distributed values of  $R$ , where the intensity is defined:

$$f_{RDC} = \sum_I^n \sum_{J>I}^n Z_I Z_J \exp^{-\alpha(r_{IJ}-R)^2} \quad (3)$$

$Z_I$  and  $Z_J$  are the nuclear charges of atoms  $I$  and  $J$ , respectively.  $r_{IJ}$  is the distance between atoms  $I$  and  $J$ , and  $R$  is a vector obtained by discretizing a linear interpolation between zero and twice the cutoff radius around the absorption site (defining the maximum pairwise distance that can be encoded by the RDC).  $\alpha$  is a smoothing parameter that controls the resolution of the RDC. As  $\alpha$  is increased, so too is the detail that is visible in RDC, but if  $\alpha$  is too large, the RDC starts to become sparse. This is illustrated in Figure 2.



**Figure 2.** An RDC for an arbitrary system; the intensity (the probability of finding an interatomic distance,  $r_{IJ}$  at some arbitrary distance,  $r$ ) is plotted as a function of the distance,  $r$ , for nine values of  $\alpha$  between 0.5 and 200.0. A larger value of  $\alpha$  increases the decay of the exponential at each value of  $r$ , increasing the resolution of the RDC; very large values of  $\alpha$  yield sparse RDCs.

Like the CM representation, the RDC representation is invariant with respect to atomic indexing, translations, and rotations of the chemical space that it describes, and its construction requires no explicit information on chemical bonding [39]. It does not have to be weighted by  $Z_I$  and  $Z_J$  alone; indeed, it is possible to construct property-weighted RDCs using any relevant atomic property (e.g., electron affinity, electronegativity, van der Waals radius, ect.) [46,47] to engineer the descriptor for a specific purpose.

An additional advantage of the RDC is that  $f_{RDC}$  can be discretised to yield a feature vector of constant length [39,46,47] regardless of the size of the chemical space it describes, i.e., CM featurisation encodes information on a fixed number of atoms and, consequently, a fixed number of interatomic distances, while RDC featurisation flexibly encodes information on all atoms and interatomic distances below a cutoff radius as per specification of  $R$ .

Where RDC featurisation is used in this work,  $\alpha = 10.0$  and  $R = 0.0 \xrightarrow{1.2} 800.0$  pm.

The CM and RDC representations are useful descriptors on account of their simplicity; both require very little space in memory, and the requisite operations for their construction are easily vectorisable; large datasets can be featurised quickly, and can typically fit in the memory in their entirety.

### 2.3. Dataset

Our dataset comprises 9040 unique Fe-containing structures harvested from the Materials Project Library via the Materials Project API. Fe K-edge XANES spectra for one arbitrary absorption site per structure have been calculated using multiple scattering theory as implemented in the FDMNES package [12]. The Fe K-edge XANES calculations employed a self-consistent muffin-tin-type potential of radius 6.0 Å around the absorbing site. The interaction with the X-ray field was described using the electric quadrupole approximation, and scalar relativistic effects were included. To transform the computed cross-sections into XANES spectra that can be compared to experiment, the cross-sections need to be convoluted with a function that accounts for the core-hole-lifetime broadening, instrument response, and many-body effects, e.g., inelastic losses. Throughout this work, this convolution has been performed using an energy-dependent arctangent function via an empirical model close to the Seah-Dench formalism [48]:

$$\Gamma = \Gamma_i + \Gamma_f \left( \frac{1}{2} + \frac{1}{\pi} \arctan \left( \frac{\pi \Gamma_f}{3 E_w} \left( \frac{E - E_f}{E_c} - \frac{E_c^2}{(E - E_f)^2} \right) \right) \right) \quad (4)$$

$\Gamma$  is defined over the energy scale,  $E$ , of the XANES spectrum as per specification of the core-level and final-state widths ( $\Gamma_i$  and  $\Gamma_f$ , respectively), and the centre and width of the arctangent function ( $E_c$  and  $E_w$ , respectively). The arctangent convolution is performed as implemented in the FDMNES package [12].

The arctangent convolution is only applied as a post-processing step on XANES spectra estimated by our DNN; our dataset comprises only unconvoluted cross-sections, and our DNN learns from these unconvoluted cross-sections.

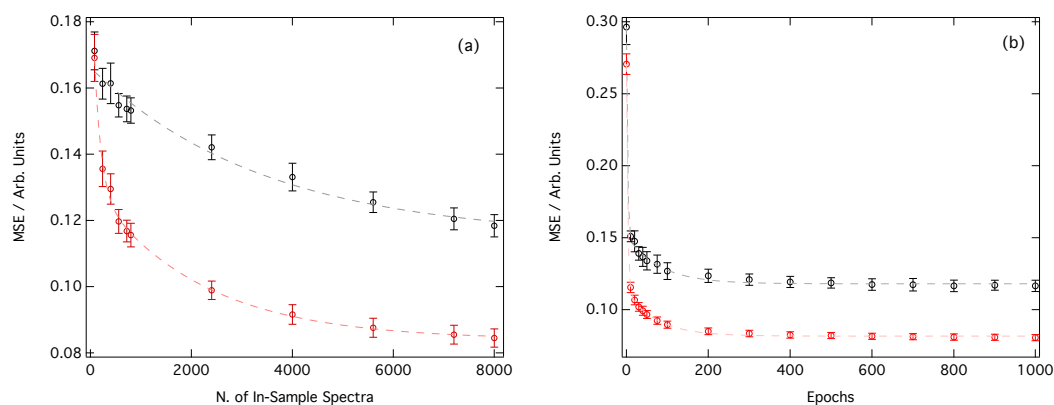
## 3. Results

### 3.1. Performance of the Deep Neural Network

Figure 3a shows the MSE as a function of the number of in-sample spectra accessible to our DNN during the learning process. The local environment around each Fe absorption site has been featurised either as a CM (black) or RDC (red). In the small-sample limit (ca. 100 in-sample spectra), both representations exhibit similar performance with the MSE of 0.17. However, as the number of in-sample spectra accessible to our DNN is increased, an almost linear improvement in MSE is seen when CM featurisation is used and a MSE of 0.12 is obtained for the large-sample limit (ca. 9000 in-sample spectra). In contrast, RDC featurisation gives a rapid initial improvement, delivering a much smaller MSE than can be achieved via CM featurisation. Beyond ca. 2000 in-sample spectra, the improvement in the MSE begins to slow, but the final MSE in the large-sample limit (ca. 0.08) is still significantly lower than the MSE that can be achieved using CM featurisation.

Figure 3b illustrates the performance during the training of the DNN, shown as a function of the number of forward passes through our dataset. While, as seen in Figure 3a, we achieve a lower MSE for the RDC representation, in both cases it is observed that the DNN can be optimised in <500 forward passes through the dataset. This is achievable in as little as five to ten minutes if graphical processing unit (GPU) acceleration is used.

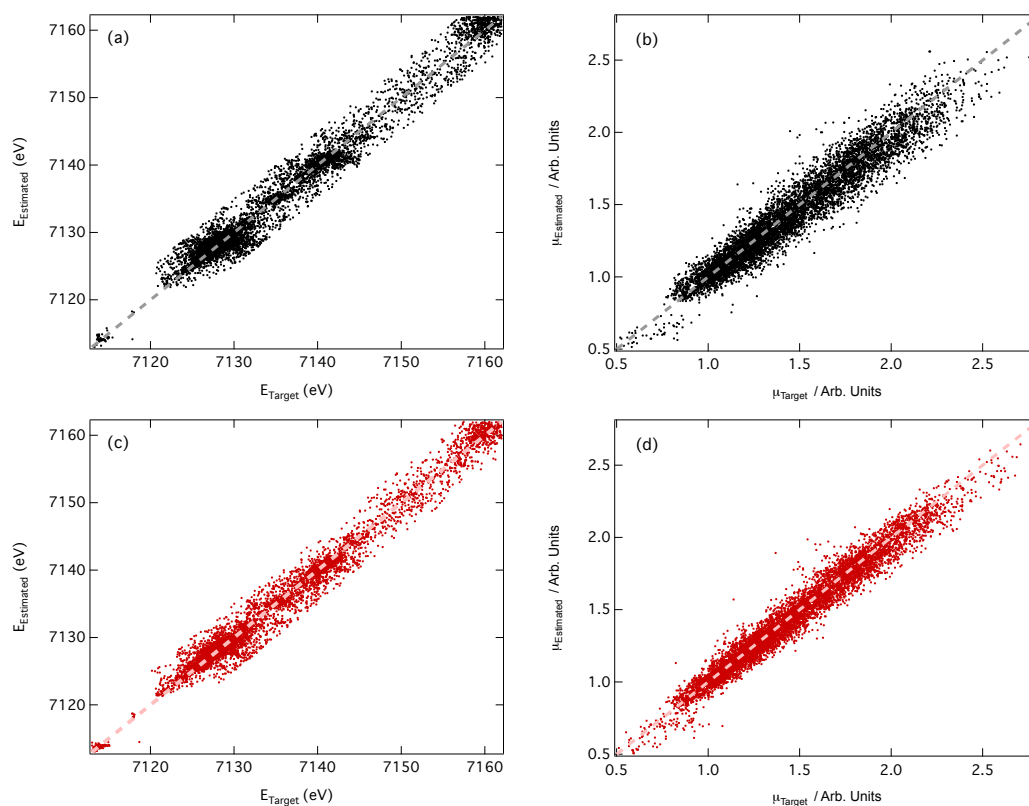




**Figure 3.** (a) Evolution of the mean squared error (MSE) as a function of the number of in-sample spectra accessible to the DNN. (b) Evolution of the MSE as a function of the number of forward passes through our dataset ('epochs'). The local environment around each Fe absorption site has been featurised either as a CM (black) or RDC (red). Data points are averaged over 100 K-fold cross-validated evaluations; error bars indicate one standard deviation.

### 3.2. Predictions of Peak Position and Intensity

When predicting XANES spectra, accurate reproduction of the positions and intensities of above-ionisation resonances is crucial, as these directly encode the structural information in the spectrum. Figure 4 shows parity plots of the difference between the estimated and target peak positions on the energy ( $E_{\text{Target}}$  and  $E_{\text{Est.}}$ ) and intensity ( $\mu_{\text{Target}}$  and  $\mu_{\text{Est.}}$ ) scales. The upper (Figure 4a,b) and lower panels (Figure 4c,d) display the results from CM and RDC featurisation, respectively.



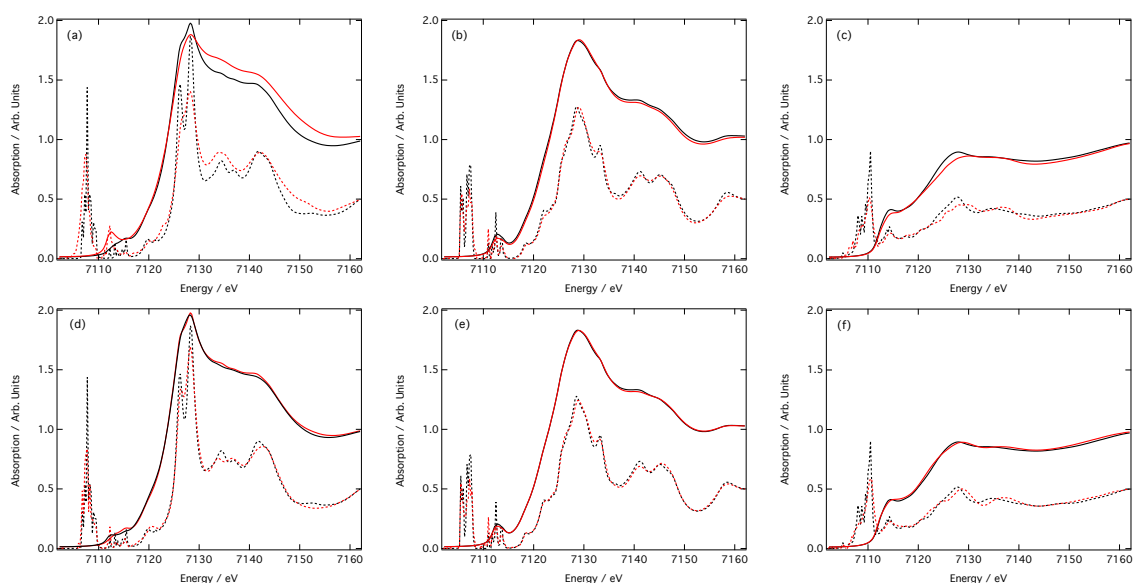
**Figure 4.** Parity plots of estimated and target peak positions on the (a,c) energy ( $E_{\text{Target}}$  and  $E_{\text{Est.}}$ , respectively) and (b,d) intensity ( $\mu_{\text{Target}}$  and  $\mu_{\text{Est.}}$ , respectively) scales; (a,b) use the CM representation, while (c,d) use the RDC representation.

In both cases, a strong linear relationships are evidenced by the coefficients of determination,  $R^2$ , which are 0.974 and 0.930 for energy and intensity, respectively, if CM featurisation is used, and 0.986 and 0.973 for energy and intensity, respectively, if RDC featurisation is used. As expected from the training curve shown in Figure 3a, the RDC representation performs slightly better exhibiting a lower  $R^2$  in both cases as expected from the narrower spread which is visible in Figure 4.

### 3.3. Predictions of Spectra

Figure 5 compares six computed XANES spectra with their corresponding out-of-sample DNN estimations. The dashed lines represent the computed and predicted cross-sections (scaled by 50% for clarity) and the solid lines represent the computed and predicted XANES spectra post-convolution of the cross-sections with the arctangent function (Equation (4)). These XANES spectra all belong to the first centile when performance is ranked over all out-of-sample DNN estimations by MSE.

The top three XANES spectra (Figure 5a–c) were obtained using CM featurisation, while the bottom three (Figure 5d–f) were obtained using RDC featurisation. In the latter case, the DNN-estimated XANES spectra can hardly be distinguished from the target XANES spectra; while discrepancies can be observed in the unconvoluted cross-sections on which our DNN is trained, the differences are negligibly small and can be considered insignificant once the arctangent convolution has been applied. In contrast, differences between the DNN-estimated and target XANES spectra are amplified when using CM featurisation. Inspection of the unconvoluted cross-sections suggests that these differences have their origin in the estimated intensities of peaks; this is most apparent in Figure 5a. CM featurisation performs less effectively than RDC featurisation on estimated peak intensities, as also evidenced in Figure 4.

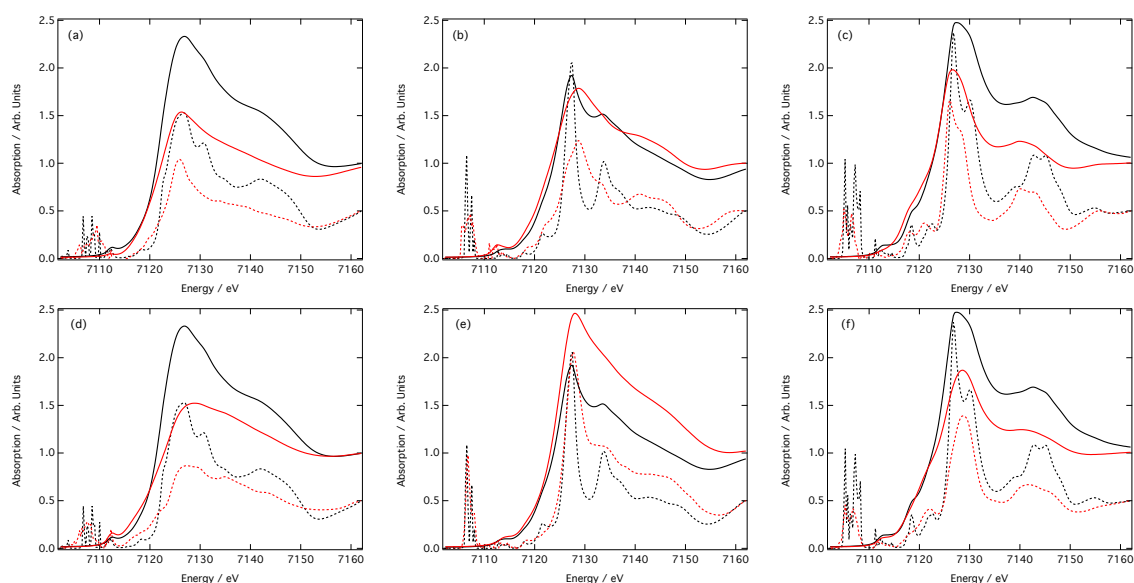


**Figure 5.** Arctangent-convoluted (solid) and unconvoluted (dashed) target (black) and out-of-sample DNN-estimated (red) Fe K-edge X-ray absorption near-edge structure (XANES) spectra for absorption sites in (a,d)  $C_6Al_2Fe_4O_{15}$ , (b,e) FeOF, and (c,f)  $Sm_2Fe_{17}H_3$ . Spectra belong to the first centile when performance is ranked over all out-of-sample DNN estimations by MSE. Spectra in panels a–c and d–f were obtained using the CM and RDC representations, respectively. Amplitudes of all unconvoluted spectra have been reduced by half for clarity.

Figure 6 shows the three samples spectra, optimised using the CM (panels a–c) and the RDC (panels d–f) representation, drawn from the ninety-ninth centile, i.e., the mostly poorly predicted spectra. In this case, as with previous work [32], the principle reason that these are in the lowest centile is due to their underestimation of the spectral intensity that compounds across the energy scale. This is especially true in the case of the RDC representation which finds peaks in the right position. This is less



so for the CM representation, which is especially apparent in Figure 6b,c, for which larger deviations are observed.



**Figure 6.** Arctangent-convoluted (solid) and unconvoluted (dashed) target (black) and out-of-sample DNN-estimated (red) Fe K-edge XANES spectra for absorption sites in (a,d)  $\text{Fe}_3\text{H}_{36}\text{C}_{12}\text{S}_6(\text{BrO}_2)_3$ , (b,e)  $\text{KFe}_2\text{F}_6$ , and (c,f)  $\text{Li}_7\text{Fe}_3\text{O}_{10}$ . Spectra belong to the ninety-ninth centile when performance is ranked over all out-of-sample DNN estimations by MSE. Spectra in panels a–c and d–f were obtained using the CM and RDC representations, respectively. Amplitudes of all unconvoluted spectra have been reduced by half for clarity.

#### 4. Discussion and Conclusions

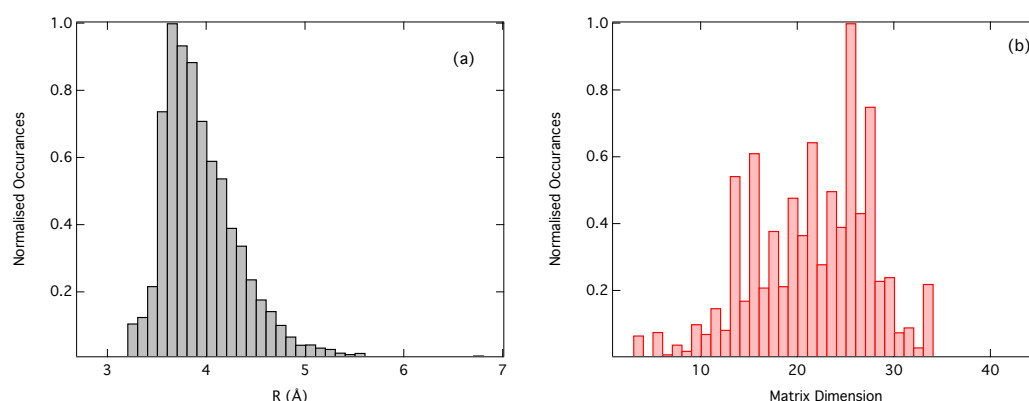
Appropriate featurisation is crucial for achieving best-in-class performance from a DNN. In this contribution, we have outlined the effect that the choice of featurisation (CM or RDC featurisation) has on the performance of our DNN [32] engineered for the prediction of XANES spectra at Fe K-edge. In both cases, a MSE of  $\leq 0.12$  is readily achievable in only a few minutes of real-time learning, and the example estimations of out-of-sample XANES spectra shown in Figures 5 and 6 demonstrate that both representations are able to deliver qualitative predictions of out-of-sample XANES spectra, even in the ninety-ninth centile. However, Figure 3 demonstrates that convergence of the DNN during the learning process is faster, better if one is restricted to the small-sample limit, and ultimately achieves a lower final MSE in the large-sample limit. These results evidence that the RDC is to be preferred over the CM as a representation of chemical space for this particular problem and DNN architecture.

At this point, it is important to highlight an important difference between the CM and RDC representations. Throughout this work we have limited the dimensions of  $M_{IJ}$  to  $20 \times 20$ . Optimisation of our DNN has led us to identify  $N = 20$  as the optimal CM dimension, i.e., that which gives the lowest MSE as evaluated on out-of-sample examples, within the limit of the possible values of  $N$  that are large enough to capture the necessary structural information, but not so large as to increase the propensity for overfitting. The maximum radius around an absorption site encoded into a CM of dimensions  $20 \times 20$  is consequently system-dependent, e.g., it depends on the identities of the neighbouring atoms around the absorption site and, by extension, the packing density of the system to be featurised.

Figure 7a shows a histogram of the maximum radius around the absorption site encoded into  $M_{IJ}$  when the dimensions are limited to  $20 \times 20$ . The modal radius is ca.  $3.5 \text{ \AA}$ , which is in close agreement with, albeit slightly smaller than, the optimal cutoff radius identified for RDC featurisation ( $4.0 \text{ \AA}$ ) and this cutoff radius encodes approximately two coordination spheres around the absorption site. The smaller cutoff radius suggests that it is harder for CM featurisation to encode effectively all of the

geometric information required to reproduce the XANES spectra as accurately as when using RDC featurisation. Figure 7b shows the reverse, i.e., a histogram of the CM dimensions when the radius around the absorbing atoms is set to 4.0 Å. Here the model dimension is around  $25 \times 25$ , meaning that the RDC featurisation, on average, can describe the effect of a larger number of atoms around the absorbing atom.

In summary, the performance of our DNN demonstrates that it is possible to develop a highly generalisable neural network for the prediction of XANES spectra at a specific absorption edge for an arbitrary absorption site, and that the RDC is a robust local descriptor for this purpose. This represents a highly encouraging starting point for our proof-of-principle demonstration which can be developed in a number of ways. Firstly, our theoretical XANES spectra (from which our DNN learns) are calculated under the muffin-tin approximation, and although this represents a computationally cost-effective choice for developing the underlying method, it is clear that the usefulness of our DNN can be considerably improved by moving beyond this. Secondly, the training set from which our DNN learns is composed of perfectly-ordered, homogeneous crystalline systems. While we have previously demonstrated [32] that it can be applied to situations outside of this scope, the sensitivity of our DNN to irregularities in the bulk such as vacancies, defects, undercoordinated sites, and the effects of lattice stress remains unclear. Finally, our DNN primarily considers only the local geometric environment around the absorption site of interest, and its ability to describe the changes in electronic charge state of the absorbing atom and therefore reproduce edge shifts is still uncertain. This could be incorporated as a post-processing step by simply shifting the predicted XANES spectra; this is commonly used for good approximation in time-resolved XAS experiments [49,50]. It remains desirable, however, to have this included from first principles. These aspects will be the focus of future work.



**Figure 7.** (a) Histograms of (a) the maximum radii around each Fe absorption site encoded in a CM of dimensions  $20 \times 20$ , and (b) the necessary CM dimension,  $N$ , required to encode a radius of 4.0 Å around each Fe absorption site in our dataset. Histograms are normalised.

**Author Contributions:** Conceptualization, T.J.P.; methodology, T.J.P. and C.D.R.; investigation and analysis, M.M.M.M., T.J.P. and C.D.R.; writing—review and editing, T.J.P., C.D.R. and M.M.M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research described in this paper was funded by the Leverhulme Trust (Project RPG-2016-103) and EPSRC (EP/S022058/1, EP/R021503/1, and EP/R51309X/1). CDR is supported by a Doctoral Prize Fellowship (EP/R51309X/1). M.M.M.M. thanks Jazan University (KSA) for supporting her study and funding. This research made use of the Rocket High Performance Computing (HPC) service at Newcastle University. CDR additionally thanks the Alan Turing Institute, via which access to the EPSRC-supported (EP/T022205/1) Joint Academic Data Science Endeavour (JADE) HPC cluster was provided under Project JAD029.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Asakura, H.; Hosokawa, S.; Ina, T.; Kato, K.; Nitta, K.; Uera, K.; Uruga, T.; Miura, H.; Shishido, T.; Ohyama, J.; et al. Dynamic Behavior of Rh Species in a Rh/Al<sub>2</sub>O<sub>3</sub> Model Catalyst During a Three-Way Catalytic Reaction: An Operando X-ray Absorption Spectroscopy Study. *J. Am. Chem. Soc.* **2018**, *140*, 176–184. [[CrossRef](#)] [[PubMed](#)]
2. Fabbri, E.; Abbott, D.F.; Nachttegaal, M.; Schmidt, T.J. Operando X-ray Absorption Spectroscopy: A Powerful Tool Toward Water-Splitting Catalyst Development. *Curr. Opin. Electrochem.* **2017**, *5*, 20–26. [[CrossRef](#)]
3. Penfold, T.; Milne, C.; Chergui, M. Recent Advances in Ultrafast X-ray Absorption Spectroscopy of Solutions. *Adv. Chem. Phys.* **2013**, *153*, 1–41.
4. Milne, C.; Penfold, T.; Chergui, M. Recent Experimental and Theoretical Developments in Time-Resolved X-ray Spectroscopies. *Coord. Chem. Rev.* **2014**, *277*, 44–68. [[CrossRef](#)]
5. Koningsberger, D.; Prins, R. *X-ray Absorption: Principles, Applications, and Techniques of EXAFS, SEXAFS, and XANES*; Wiley: Chichester, UK, 1988.
6. Bunker, G. *Introduction to XAFS: A Practical Guide to X-ray Absorption Fine Structure Spectroscopy*; Cambridge University Press: Cambridge, UK, 2010.
7. Ankudinov, A.; Ravel, B.; Rehr, J.; Conradson, S. Real-Space Multiple-Scattering Calculation and Interpretation of X-ray Absorption Near-Edge Structure. *Phys. Rev. B* **1998**, *58*, 7565. [[CrossRef](#)]
8. Rehr, J.J.; Albers, R.C. Theoretical Approaches to X-ray Absorption Fine Structure. *Rev. Mod. Phys.* **2000**, *72*, 621. [[CrossRef](#)]
9. O’day, P.; Rehr, J.; Zabinsky, S.; Brown, G.J. Extended X-ray Absorption Fine Structure (EXAFS) Analysis of Disorder and Multiple Scattering in Complex Crystalline Solids. *J. Am. Chem. Soc.* **1994**, *116*, 2938–2949. [[CrossRef](#)]
10. Mustre, J.; Yacoby, Y.; Stern, E.A.; Rehr, J.J. Analysis of Experimental Extended X-ray Absorption Fine Structure (EXAFS) Data Using Calculated Curved-Wave, Multiple-Scattering EXAFS Spectra. *Phys. Rev. B* **1990**, *42*, 10843. [[CrossRef](#)]
11. Rehr, J.; Ankudinov, A. Progress in the Theory and Interpretation of XANES. *Coord. Chem. Rev.* **2005**, *249*, 131–140. [[CrossRef](#)]
12. Bunău, O.; Joly, Y. Self-Consistent Aspects of X-ray Absorption Calculations. *J. Phys. Condens. Matter* **2009**, *21*, 345501. [[CrossRef](#)]
13. Kas, J.J.; Jorissen, K.; Rehr, J.J. Real-Space Multiple-Scattering Theory of X-ray Spectra. In *X-ray Absorption and X-ray Emission Spectroscopy: Theory and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
14. Zhou, Y.; Doronkin, D.E.; Zhao, Z.; Plessow, P.N.; Jelic, J.; Detlefs, B.; Pruessmann, T.; Studt, F.; Grunwaldt, J.D. Photothermal Catalysis over Nonplasmonic Pt/TiO<sub>2</sub> Studied by Operando HERFD-XANES, Resonant XES, and DRIFTS. *ACS Catal.* **2018**, *8*, 11398–11406. [[CrossRef](#)]
15. Hu, H.; Wachs, I.E.; Bare, S.R. Surface Structures of Supported Molybdenum Oxide Catalysts: Characterization by Raman and Mo L<sub>3</sub>-Edge XANES. *J. Phys. Chem.* **1995**, *99*, 10897–10910. [[CrossRef](#)]
16. Alayon, E.M.; Nachttegaal, M.; Ranocchiari, M.; van Bokhoven, J.A. Catalytic Conversion of Methane to Methanol Over Cu Mordenite. *Chem. Commun.* **2012**, *48*, 404–406. [[CrossRef](#)]
17. Capano, G.; Milne, C.; Chergui, M.; Rothlisberger, U.; Tavernelli, I.; Penfold, T. Probing Wavepacket Dynamics Using Ultrafast X-ray Spectroscopy. *J. Phys. B At. Mol. Opt. Phys.* **2015**, *48*, 214001. [[CrossRef](#)]
18. Penfold, T.J.; Szlachetko, J.; Santomauro, F.G.; Britz, A.; Gawelda, W.; Doumy, G.; March, A.M.; Southworth, S.H.; Rittmann, J.; Abela, R.; et al. Revealing Hole Trapping in Zinc Oxide Nanoparticles by Time-Resolved X-ray Spectroscopy. *Nat. Commun.* **2018**, *9*, 478. [[CrossRef](#)]
19. Northey, T.; Norell, J.; Fouda, A.E.; Besley, N.A.; Odellius, M.; Penfold, T.J. Ultrafast Nonadiabatic Dynamics Probed by Nitrogen K-edge Absorption Spectroscopy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 2667–2676. [[CrossRef](#)] [[PubMed](#)]
20. Hayes, D.; Hadt, R.G.; Emery, J.D.; Cordones, A.A.; Martinson, A.B.; Shelby, M.L.; Fransted, K.A.; Dahlberg, P.D.; Hong, J.; Zhang, X.; et al. Electronic and Nuclear Contributions to Time-Resolved Optical and X-ray Absorption Spectra of Hematite and Insights into Photoelectrochemical Performance. *Energy Environ. Sci.* **2016**, *9*, 3754–3769. [[CrossRef](#)]

21. Cannelli, O.; Bacellar, C.; Ingle, R.; Bohinc, R.; Kinschel, D.; Bauer, B.; Ferreira, D.; Grolimund, D.; Mancini, G.; Chergui, M. Toward Time-Resolved Laser T-Jump/X-ray Probe Spectroscopy in Aqueous Solutions. *Struct. Dyn.* **2019**, *6*, 064303. [[CrossRef](#)]
22. Kuzmin, A.; Timoshenko, J.; Kalinko, A.; Jonane, I.; Anspoks, A. Treatment of Disorder Effects in X-Ray Absorption Spectra Beyond the Conventional Approach. *Radiat. Phys. Chem.* **2018**. doi:10.1016/j.radphyschem.2018.12.032. [[CrossRef](#)]
23. Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A.I. Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8*, 5091–5098. [[CrossRef](#)] [[PubMed](#)]
24. Timoshenko, J.; Halder, A.; Yang, B.; Seifert, S.; Pellin, M.J.; Vajda, S.; Frenkel, A.I. Subnanometer Substructures in Nanoassemblies Formed from Clusters under a Reactive Atmosphere Revealed Using Machine Learning. *J. Phys. Chem. C* **2018**, *122*, 21686–21693. [[CrossRef](#)]
25. Timoshenko, J.; Ahmadi, M.; Cuenya, B.R. Is There a Negative Thermal Expansion in Supported Metal Nanoparticles? An in Situ X-ray Absorption Study Coupled with Neural Network Analysis. *J. Phys. Chem. C* **2019**, *123*, 20594–20604. [[CrossRef](#)]
26. Ahmadi, M.; Timoshenko, J.; Behafarid, F.; Cuenya, B.R. Tuning the Structure of Pt Nanoparticles through Support Interactions: An in Situ Polarized X-ray Absorption Study Coupled with Atomistic Simulations. *J. Phys. Chem. C* **2019**, *123*, 10666–10676. [[CrossRef](#)] [[PubMed](#)]
27. Timoshenko, J.; Wrasman, C.J.; Luneau, M.; Shirman, T.; Cargnello, M.; Bare, S.R.; Aizenberg, J.; Friend, C.M.; Frenkel, A.I. Probing Atomic Distributions in Mono- and Bimetallic Nanoparticles by Supervised Machine Learning. *Nano Lett.* **2019**, *19*, 520–529. [[CrossRef](#)]
28. Timoshenko, J.; Frenkel, A.I. “Inverting” X-ray Absorption Spectra of Catalysts by Machine Learning in Search for Activity Descriptors. *ACS Catal.* **2019**, *9*, 10192–10211. [[CrossRef](#)]
29. Liu, Y.; Marcella, N.; Timoshenko, J.; Halder, A.; Yang, B.; Kolipaka, L.; Pellin, M.J.; Seifert, S.; Vajda, S.; Liu, P.; et al. Mapping XANES Spectra on Structural Descriptors of Copper Oxide Clusters Using Supervised Machine Learning. *J. Chem. Phys.* **2019**, *151*, 164201. [[CrossRef](#)]
30. Carbone, M.R.; Yoo, S.; Topsakal, M.; Lu, D. Classification of Local Chemical Environments from X-ray Absorption Spectra Using Supervised Machine Learning. *Phys. Rev. Mater.* **2019**, *3*, 033604. [[CrossRef](#)]
31. Carbone, M.R.; Topsakal, M.; Lu, D.; Yoo, S. Machine-Learning X-Ray Absorption Spectra to Quantitative Accuracy. *Phys. Rev. Lett.* **2020**, *124*, 156401. [[CrossRef](#)]
32. Rankine, C.D.; Madkhali, M.M.; Penfold, T.J. A Deep Neural Network for the Rapid Prediction of X-ray Absorption Spectra. *J. Phys. Chem. A* **2020**, *124*, 4263–4270. [[CrossRef](#)]
33. Rupp, M.; Tkatchenko, A.; Müller, K.R.; Von Lilienfeld, O.A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301. [[CrossRef](#)]
34. Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.R.; Von Lilienfeld, O.A. Learning Invariant Representations of Molecules for Atomization Energy Prediction. In *Advances in Neural Information Processing Systems*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 25, pp. 440–448.
35. Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.R.; Von Lilienfeld, O.A. Machine Learning of Molecular Electronic Properties in Chemical Compound Space. *New J. Phys.* **2013**, *15*, 095003. [[CrossRef](#)]
36. Gasteiger, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. Finding the 3D Structure of a Molecule in its IR Spectrum. *Fresenius J. Anal. Chem.* **1997**, *359*, 50–55. [[CrossRef](#)]
37. Hemmer, M.C.; Steinhauer, V.; Gasteiger, J. Deriving the 3D Structure of Organic Molecules from their Infrared Spectra. *J. Vib. Spectrosc.* **1999**, *19*, 151–164. [[CrossRef](#)]
38. Hemmer, M.C.; Gasteiger, J. Prediction of Three-Dimensional Molecular Structures using Information from Infrared Spectra. *Anal. Chim. Acta* **2000**, *420*, 145–154. [[CrossRef](#)]
39. von Lilienfeld, O.A.; Ramakrishnan, R.; Rupp, M.; Knoll, A. Fourier Series of Atomic Radial Distribution Functions: A Molecular Fingerprint for Machine Learning Models of Quantum Chemical Properties. *Int. J. Quantum Chem.* **2015**, *115*, 1084–1093. [[CrossRef](#)]
40. Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; Von Lilienfeld, O.A.; Tkatchenko, A.; Müller, K.R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419. [[CrossRef](#)] [[PubMed](#)]

41. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. *arXiv* **2012**, arXiv:1207.0580.
42. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Available online: <http://tensorflow.org/> (accessed on 11 June 2020).
43. Keras. 2015. Available online: <http://github.com/keras-team/keras> (accessed on 11 June 2020).
44. GPy: A Gaussian Process Framework in Python. 2012. Available online: <http://github.com/SheffieldML/GPy> (accessed on 11 June 2020).
45. GPyOpt: A Bayesian Optimization Framework in Python. 2016. Available online: <http://github.com/SheffieldML/GPyOpt> (accessed on 11 June 2020).
46. Fernandez, M.; Trefiak, N.R.; Woo, T.K. Atomic-Property-Weighted Radial Distribution Functions Descriptors of Metal–Organic Frameworks for the Prediction of Gas Uptake Capacity. *J. Phys. Chem. C* **2013**, *27*, 14095–14105. [[CrossRef](#)]
47. Krykunov, M.; Woo, T.K. Bond-Type-Restricted Property-Weighted Radial Distribution Functions for Accurate Machine Learning Prediction of Atomization Energies. *J. Chem. Theory Comput.* **2018**, *14*, 5229–5237. [[CrossRef](#)]
48. Seah, M.; Dench, W. *NPL Report Chem.* **1978** *82*, 1.
49. Reinhard, M.; Penfold, T.; Lima, F.; Rittmann, J.; Rittmann-Frank, M.; Abela, R.; Tavernelli, I.; Rothlisberger, U.; Milne, C.; Chergui, M. Photooxidation and Photoaquation of Iron Hexacyanide in Aqueous Solution: A Picosecond X-ray Absorption Study. *Struct. Dyn.* **2014**, *1*, 024901. [[CrossRef](#)]
50. Penfold, T.J.; Karlsson, S.; Capano, G.; Lima, F.A.; Rittmann, J.; Reinhard, M.; Rittmann-Frank, M.; Bräm, O.; Baranoff, E.; Abela, R.; et al. Solvent-Induced Luminescence Quenching: Static and Time-Resolved X-ray Absorption Spectroscopy of a Copper(I) Phenanthroline Complex. *J. Phys. Chem. A* **2013**, *117*, 4591–4601. [[CrossRef](#)] [[PubMed](#)]

**Sample Availability:** Samples of the compounds are not available from the authors.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).