Machine Learning: Science and Technology

Neural networks and kernel ridge regression for excited states dynamics of $CH_2NH_2^+$: From single-state to multi-state representations and multi-property machine learning models

IOP Publishing Mach. Learn.: Sci. Technol. 1 (2020) 025009 https://doi.org/10.1088/2632-2153/ab88d0

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
18 December 2019

REVISED
31 March 2020

ACCEPTED FOR PUBLICATION
7 April 2020

PUBLISHED
19 May 2020

PAPER

Neural networks and kernel ridge regression for excited states dynamics of $CH_2NH_2^+$: From single-state to multi-state representations and multi-property machine learning models

Julia Westermayr[1], **Felix A Faber**[2], **Anders S Christensen**[2], **O Anatole von Lilienfeld**[2] and **Philipp Marquetand**[1,3,4]

[1] University of Vienna, Faculty of Chemistry, Institute of Theoretical Chemistry, Währinger Str. 17, 1090 Wien, Austria
[2] Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), Department of Chemistry, University of Basel, Klingelbergstr. 80, CH-4056 Basel, Switzerland
[3] Vienna Research Platform on Accelerating Photoreaction Discovery, University of Vienna, Währinger Str. 17, 1090 Wien, Austria
[4] University of Vienna, Faculty of Chemistry, Data Science @ Uni Vienna,Währinger Str. 29, 1090 Wien, Austria

E-mail: anatole.vonlilienfeld@unibas.ch **and** philipp.marquetand@univie.ac.at

Keywords: machine learning, photodynamics, excited states, quantum chemistry, neural networks, kernel ridge regression

Supplementary material for this article is available online

Abstract

Excited-state dynamics simulations are a powerful tool to investigate photo-induced reactions of molecules and materials and provide complementary information to experiments. Since the applicability of these simulation techniques is limited by the costs of the underlying electronic structure calculations, we develop and assess different machine learning models for this task. The machine learning models are trained on *ab initio* calculations for excited electronic states, using the methylenimmonium cation ($CH_2NH_2^+$) as a model system. Two distinct strategies for modeling excited state properties are tested in this work. The first strategy is to treat each state separately in a kernel ridge regression model and all states together in a multiclass neural network. The second strategy is to instead encode the state as input into the model, which is tested with both models. Numerical evidence suggests that using the state as input yields the best performance. An important goal for excited-state machine learning models is their use in dynamics simulations, which needs not only state-specific information but also couplings, i.e. properties involving pairs of states. Accordingly, we investigate how well machine learning models can predict the couplings. Furthermore, we explore how combining all properties in a single neural network affects the accuracy. Finally, machine learning predicted energies, forces, and couplings are used to carry out excited-state dynamics simulations. Results demonstrate the scopes and possibilities of machine learning to model excited-state properties.

1. Introduction

Many fundamental processes in nature and life are direct consequences of excitation of molecules by light. For example, photosynthesis [1], vision with photo-receptors in the eye [2, 3], or the root cause of diseases such as skin cancer [4, 5] are all based on a photo-induced process. The excited-state dynamics and kinetics of compounds can give insight into why and how these processes occur or can be used to help designing new drugs [6] or materials [7–9].

After a molecule is irradiated with light, it can enter a higher electronic state. Several processes, radiationless or radiative ones, may follow. These processes give rise to the photostability or photodamage of a molecule, and hence photoreactions in general. Knowing the high-dimensional potential energy surfaces of a molecule makes a comprehensive photochemical study possible. However, it is a challenge to find meaningful and accurate potential energy surfaces in advance to execute excited-state molecular dynamics simulations. A possible solution for this problem is the use of on-the-fly *ab initio* molecular dynamics

simulations [10–12]. The mixed-quantum classical methods—such as the surface-hopping methodology that is used in this work—often remain the methods of choice and are a good compromise between accuracy and computational efficiency. In this way, large molecules, i.e. with up to hundreds of atoms, can be treated. Still, the large number of costly electronic structure calculations limits the simulation times of nonadiabatic dynamics to the range of femto- to picoseconds [10, 13–15].

With the rise of machine learning (ML) and the amount of data and computational power available, the fitting of potential energy surfaces of a molecule was put into spotlight. Although such approaches were used already more than 25 years ago [16–20], the interest to speed up simulations in the field of nonadiabatic molecular dynamics simulations has increased only in the last three years [21–30, 30–36].

The main advantage of ML models is that, at least in principle, they can predict any molecular property, typically with much improved efficiency when compared to their quantum chemical counterpart. This can be achieved by learning relations between a molecular structure (in the form of some translation- and rotation-invariant representation) and some target property (provided by quantum chemistry, usually real-valued or complex numbers) [37, 38]. The application of ML models for dynamics simulations in the electronic ground state [39–45] or excited states [21–30, 30–36, 46, 47] already exist, and show the potential for further developing this research field.

Nevertheless, the challenges to model excited states of molecules have not always been tackled successfully with ML: Not only one potential energy surface has to be learned, but several, including the couplings between them, that should also be treated in an ML model [29]. Independent works applied kernel ridge regression (KRR) and neural networks (NNs) to enhance quantum chemical calculations in nonadiabatic molecular dynamics simulations and show the following trend: Models based on KRR need to resort to intermediate quantum chemistry steps in critical regions to obtain correct dynamics simulations [22, 23], whereas NNs are able to completely replace quantum chemistry during the dynamics [24, 26, 28, 29]. This observation raises the question why some models fail for some excited-state properties and others do not. Our goal is to investigate relevant properties for nonadiabatic molecular dynamics simulations by using KRR and NNs. The aim of this work is thus twofold: 1) a comparison of KRR and NNs using different representations for the molecular structure and 2) possible improvements of existing ML techniques for excited-state property prediction.

The methylenimmonium cation, $CH_2NH_2^+$, is used as model system for this purpose. This cation, like the larger homologue retinal, belongs to the member of protonated Schiff bases, and shows isomerization of the double bond after light excitation [48–51]. Similarly, the absorption of light in retinal leads to an isomerization process, that is fundamental for vision [52]. The ultrafast dynamics is particularly challenging to reproduce with ML models, but, at the same time, it remains feasible to provide quantum chemistry reference simulations. Unlike in the isoelectronic molecule ethylene, the excited-state calculations are not hampered by low-lying Rydberg states [53]. Therefore, the methylenimmonium cation is well suited as a test system for this study.

2. Methods

2.1. Surface-hopping molecular dynamics

The program SHARC (Surface-Hopping including ARbitrary Couplings) [54] is used for surface-hopping molecular dynamics simulations with interfaced ML models. Due to the stochastic nature of this method, reliable results can only be obtained when considering an ensemble of independent trajectories. To this aim, Wigner sampling [55] is used to obtain 200 initial configurations of the methylenimmonium cation, to start dynamics simulations. SHARC determines nonadiabatic transitions from one potential energy surface to another from the NAC vectors between each set of singlet states. Those transitions, or so-called hops in surface-hopping, usually take place in conformational regions of the molecule, where two potential energy surfaces are in close proximity. Those regions are termed conical intersection and are not only difficult to model with ML models [23, 24, 29], but also provide a challenge to converge a quantum chemical calculation. This can be attributed to the inverse proportionality of the NAC vectors between two coupled states i and j, NAC_{ij}, to the energy gap between these two states [14]:

$$NAC_{ij} \approx \frac{\langle \Psi_i \mid \nabla_{\mathbf{R}} \hat{H}_{el} \mid \Psi_j \rangle}{E_i - E_j} \text{for } i \neq j. \quad (1)$$

Ψ_i and Ψ_j are the eigenstates of the electronic Schrödinger equation with E_i and E_j being the respective eigenvalues, i.e. the potential energies of state i and j. $\nabla_{\mathbf{R}}$ denotes the spatial derivatives with respect to the nuclear coordinates. When the two electronic states are degenerate, NACs become infinitely large and consequently show sharp peaks around conical intersections. In contrast, NAC values are almost zero elsewhere. An example is illustrated in figure 1.

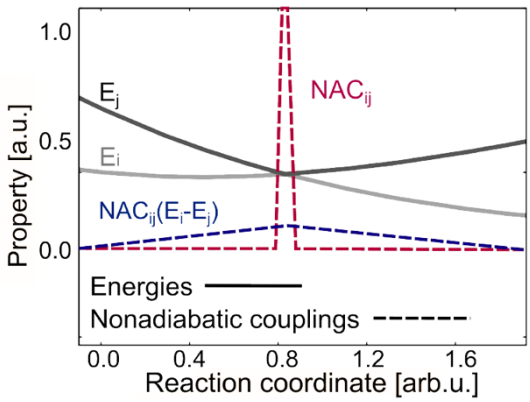

Figure 1. Scan along a reaction coordinate of $CH_2NH_2^+$ that shows a conical intersection of the E_i and E_j states. The corresponding NAC values, NAC_{ij}, taken as the norm of the NAC vectors between those states, are large at the conical intersection, but almost zero elsewhere (red dashed line). Multiplied with the energy-difference, smoother properties can be obtained (blue dashed line).

Multiplication of NACs obtained from quantum chemistry (labelled as QC) with the corresponding energy gaps can remove such sharp spikes and can provide smoother quantities (also referred to as interstate couplings) [28].

$$C_{ij}^{\text{reference}} = NAC_{ij}^{QC} \cdot \mid E_i^{QC} - E_j^{QC} \mid \tag{2}$$

For prediction, it is necessary to divide the learned quantity by the energy gap [28] obtained from ML models.

$$NAC_{ij}^{\text{predict}} = \frac{C_{ij}^{ML}}{\mid E_i^{ML} - E_j^{ML} \mid} \tag{3}$$

The outcome is given to the surface-hopping molecular dynamics program to compute the hopping probability. However, this approach requires very accurate ML potentials for energies and those are also challenging to obtain in regions near conical intersections.

2.2. Training set

The training set used here is taken from reference [29] and represents a conformational subspace of the methylenimmonium cation, $CH_2NH_2^+$. This training set involves three singlet states and covers the relevant conformational space that is visited during excited-state molecular dynamics simulations after excitation to the bright second excited state (S_2, $\pi\pi^*$). It is thus considered to constitute an optimal set for analysis of different ML models as well as common molecular representations. Triplet states are not assumed to play a major role in the excited state dynamics of $CH_2NH_2^+$ [29, 48–51]. In case of triplet states, additional couplings between triplet-triplet states (NAC vectors) and singlet-triplet states (spin-orbit couplings) would arise, that could be similarly modelled as NACs between singlet-singlet states.

The quantum chemical reference method is the multi-reference configuration interaction method accounting for single and double excitations with the basis set aug-cc-pVDZ (MR-CISD(6,4)/aug-cc-pVDZ). The active space consists of 4 electrons in 6 orbitals. The data set for training and validation contains 4000 data points that are obtained by randomly shuffling the complete set of 4770 points. The rest of the data set (770 data points) is held back as a test set. Each data point contains the xyz-coordinates of a molecular structure as well as energies for three singlet states, corresponding gradients, (transition) dipole moments, and NACs between each state. In total, 3 energy values, 54 gradient values, 27 values for (transition) dipole moments, and 54 values for NACs have to be predicted. Noticeably, calculating dipole moments and transition dipole moments to reproduce absolute experimental values is difficult even with high-level *ab initio* methods [56]. However when calculating transition probabilities, e.g. oscillator strengths for the simulation of UV spectra, absolute values are rarely important but rather relative values are of interest. The (transition) dipole moments in this work are modelled in the simplest way—that is the direct fitting of values obtained from quantum chemistry, without an *a posteriori* determination of nuclear charges [57]. For more details on the training set and its generation, see reference [29].

2.3. Machine learning models

Learning curves and scatter plots between reference data and predicted data show the quality of each ML model. By plotting the prediction error against the training set size, N, in logarithmic scale, the learning efficiency can be assessed [58–61].

Multi-layer feed forward NNs and KRR using the quantum machine learning (QML) toolkit [62] are chosen as ML models. 5-fold cross-validation is applied to optimize the hyperparameters of each model with the training set of 4000 data points. The MAE is reported on the test set of 770 data points as the mean of 10 calculations along with the standard deviations obtained from models trained on 90% of the 4000 data points. 10% are necessary to employ early-stopping for the NNs. An analogous procedure is applied for comparability when using KRR. A detailed description can be found in the Supplementary Information (SI), which is available online at (stacks.iop.org/MLST/1/025009/mmedia).

The gradients are treated as derivatives of ML potentials for energies as described in reference [63] for NNs and reference [61] for KRR. This is necessary to conserve the energy in nonadiabatic molecular dynamics simulations. For comparison, gradients are directly trained and predicted too. Details on chosen parameters for KRR and NNs are given in the SI in chapter S2.2 and S2.3, respectively. Learning (transition) dipole moments and NACs in addition to energies and gradients with one model can give insights into the influence of of joint learning of different properties. This is straightforward with NNs, therefore this effect is investigated using NNs.

2.3.0.1. Molecular representations

As a molecular representation, the matrix of inverse distances is chosen, as it gave fair results in reference [29] for NNs and other ML models, see for example references [45, 64, 65]. The FCHL (Faber-Christensen-Huang-Lilienfeld) representation is used for KRR [61, 66] and a development version of the same representation is also tested for NNs [67].

The molecule is treated as a whole with the matrix of inverse distances. The distance of an atom to all other atoms is computed, these distances are inverted and arranged in matrix format. This representation is probably the simplest and cheapest representation to use for ML and can be used for very efficient training and evaluation of ML models. The FCHL representation, however, provides a more accurate description and is computationally more expensive to apply. It does not treat the molecule as a whole, but describes an atom in its chemical and structural environment within a pre-defined cut-off region [66]. Not only the distances from one atom to the other atoms are taken into account as two-body terms, but also one- and three-body terms. These account for chemical composition as well as angular contributions, respectively.

An encoding for the quantum energy level is implemented in addition to the aforementioned representations to predict several electronic state energies at once. Several possibilities are tested to describe the electronic state. For KRR, a representation for each electronic singlet state, $S = \{1, 2, 3\}$, containing simply numbers of 1, 2, and 3 for the three states turns out to be beneficial. Other representations of the state do not result in improved learning and only change the additional hyperparameter, the width of the state kernel. Also for the NNs, several state-encoding representations are tested. Duplication of a a molecular representation N_S-times and multiplication of each copy with the corresponding state-number – 1, 2 or 3 in this case, turns out to be best. Due to the existing implementations, gradients are only treated as response properties for KKR with the FCHL representation [61] and as derivatives of NN potentials for energies with the matrix of inverse distances.

2.3.0.2. Kernel ridge regression (KRR)

In KRR, a kernel basis function is placed on each compound (each molecule) in the training set, $\{M_k\}$, and related to a property of a query compound M, $p^{KRR}(M)$, by:

$$p^{KRR}(M) = \sum_{k=1}^{N_M} \alpha_k K(M, M_k) \tag{4}$$

with N_M being the number of molecules in the training set, K the kernel, and $\{\alpha_k\}$ the regression coefficients, which are obtained through linear regression:

$$\alpha = (K + \lambda I)^{-1} p^{\text{reference}}. \tag{5}$$

The regularizer, λ (multiplied with the unit matrix I), is usually small assuming that the noise in the training set is negligible [38, 66]. Using this standard implementation, only one molecular property can be fitted at a time, which we call single-state fitting. An additional Gaussian kernel, $K_2(S, S_l)$, which relates information from a query state S to the set of available states $\{S_l\}$, is used to extended the representations and

to predict all electronic states, N_S, at once. This kernel is subsequently combined with the original kernel (now denoted as K_1) that maps a compound to its property.

$$p^{KRR}(M,S) = \sum_{k=1}^{N_M}\sum_{l=1}^{N_S}\alpha_k^l K_1(M,M_k)K_2(S,S_l). \tag{6}$$

Here, $\mathbf{p}^{KRR}(M,S)$ is a vector of length $N_M \cdot N_S$, which can be recast as a matrix of size $N_M \times N_S$. In contrast, the predicted property $p^{KRR}(M)$ from equation (4) is only dependent on the molecule and is a vector of length N_M – representing only one electronic state. N_S versions of KRR models with the standard representation in equation (5) have to be used to predict N_S energetic values. With the new state representation, only one KRR model has to be trained.

2.3.0.3. Multi-layer feed forward neural networks (NNs)

Compared to KRR, NNs possess more hyperparameters and, thus, are more difficult to optimize with respect to error convergence. However, due to the NN architecture, a benefit lies in the flexibility and possibility to relate a molecular structure to a many-state output. In principle, this many-state output can be obtained without additional encodings like the state kernel $K_S(S,S_l)$ in the KRR approach. For better comparison to the KRR approach, we also used state-encoding representations as detailed above.

All NN models use the numpy [68] and theano [69] distribution implemented in python. To find optimal hyperparameters of the models to represent the relation between a molecular geometry and its multi-dimensional output, random grid search of different sets of hyperparameters is carried out, see reference [29] for details. In all cases, the stochastic gradient descent optimization algorithm adaptive moment estimation (Adam) [70] is applied and the learning rate is annealed during training.

The NN models are trained by optimizing parameters such that mean squared errors of the predicted properties p^{NN} and the reference properties p^{QC} are minimized. Note that p is a vector that contains three values, corresponding to the electronic states, in case of fitted energies. It contains 84 values if all properties are treated together. We use the scheme of reference [40] to additionally include the forces, F, as NN derivatives in the loss function, L_2:

$$\begin{aligned} L_2 &= \frac{1}{N_M}(p^{NN}-p^{QC})^2 \\ &\quad + \frac{1}{N_M}\frac{1}{3N_a}\sum_{a}^{3N_a}(F_\alpha^{NN}-F_\alpha^{QC})^2. \end{aligned} \tag{7}$$

F_α^{NN} are the values of forces predicted with NNs and F_α^{QC} are corresponding reference values, where a runs over all atoms, N_a.

3. Results

ML models are trained on energies, forces, and NACs to speed up simulations and successfully reproduce surface-hopping molecular dynamics simulations of the reference method. The populations obtained from surface-hopping molecular dynamics simulations with chosen ML models that take gradients as derivatives from ML energy potentials are given in figure 2.

The reference dynamics in figure 2 (A) is taken from reference [29], and so is the population scheme in figure 2 (B), that results from an NN model with the inverse distance matrix as the molecular representation. The average populations of the ensemble show that the time evolution is governed by fast population transfer from the second excited state to the first excited state and back to the ground state. The agreement of these methods is fairly good.

Panel (C) shows dynamics computed with an NN model that treats all properties together. Up to around 10 fs, the dynamics agree to the reference scheme. During this time, the population is transferred from the S_2 state to the S_1 state. Afterwards, there are less transitions than expected and in the end of the simulation time only a small fraction of the population is in the S_0 state. For the KRR model with the FCHL representation given in panel (D), all of the population is transferred to the electronic ground state, S_0, within the first 10 fs. After that, there are hops from lower lying states to higher energetic states. Those hops, especially in cases of large potential energy gaps between states, are considered to be implausible and the trajectories are not reliable anymore. Furthermore, the molecule atomizes during the course of the simulation, which is not the case in the quantum chemistry reference dynamics. The premature population transfer leads the molecule to regions of the conformational subspace that are not visited with the reference method and are also not

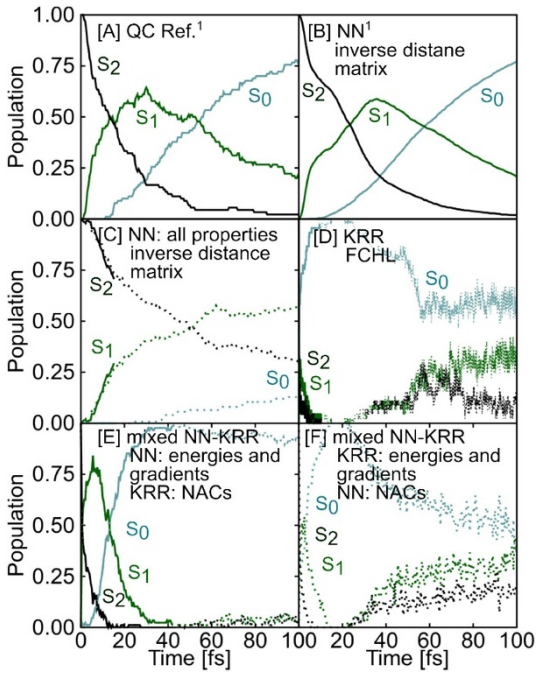

Figure 2. Populations of three singlet states as a function of time obtained from surface-hopping molecular dynamics simulations using six different methods. Dynamics based on [A] the quantum chemistry reference[1] (labeled as QC), [B] NN[1] models using the inverse distance as a descriptor, [C] NNs that treat all properties in one model, [D] KRR with the FCHL representation, and [E]-[F] mixed NN-KRR models, where energies and gradients are taken either from the NN model of panel [B] and NACs from KRR of panel [D] or vice versa. Dotted lines are populations that are considered to be wrong with respect to the reference or exhibit large energy fluctuations ([1] taken from reference [29]).

considered in the training set. Panels (E) and (F) show results from mixed models, i.e. KRR and NN models, and will be discussed later.

The way ML models learn shows why dynamics is erroneous. A correct learning behaviour is given, when the out of sample error of an ML model decreases with increasing training set size, which has been shown by Vapnik and coworkers for KRR [58] and by Müller and coworkers for NNs [59]. The learning curves are given in figure 3(A). While KRR (blue) and the NN model (red) for energies and gradients yield very similar accuracy, the NNs that treat all properties together (green) are far from being accurate.

The learning curves for the NACs (left panel in figure 3(B)) as directly obtained from quantum chemistry, i.e. as non-smooth properties, show that the learning efficiency of KRR is much higher than the one of the NNs, which is dictated by the slope of the learning curve. The NN models are comparable in their accuracy. The learning of the smooth NACs (as described in equation (2)) is evaluated in the middle panel with KRR and NN models that account for each property separately. The inaccurate potential energies of the NN model that treats all properties together prohibit the prediction of smooth NACs in this case. The accuracy of the actual NACs obtained from the smooth couplings and the corresponding ML energy gaps (as given in panel (A)), is already high for a small training set size and is comparable to the MAE for direct NAC prediction with a large training set. However, increasing the training set size cannot improve prediction accuracy anymore. For KRR, the MAE can be reduced with increasing training set size, but the learning curve is not linear. These findings indicate that the energies must be predicted with high accuracy in order to use smooth NACs, compare also figure S4 panels (B) and (C) in the SI. The learning curves for (transition) dipole moments show similar trends as the ones for energies, gradients and NACs and are given in the SI in figure S5. Furthermore, NNs with the FCHL representation as well as KRR with the inverse distance matrix as molecular representations are discussed along with scatter plots for energies and gradients (see SI figures S1, S2, S4, and S5).

Having analyzed the learning behavior for energies and NACs separately, their interplay is investigated in the following. We thus return to the dynamics depicted in figure 2 in panels (E) and (F). In these panels, population plots from mixed models are shown—once with energies and gradients obtained from NNs and NACs obtained from KRR and once vice versa. Surprisingly, none of the cases leads to accurate population schemes. Nethertheless, the trend of the populations in panel (E) – with energies and gradients from NNs and NACs from KRR—is similar to the reference scheme. The dynamics only happens on a much shorter time scale, indicating too large NACs. Plotting the potential energy curves along two different reaction

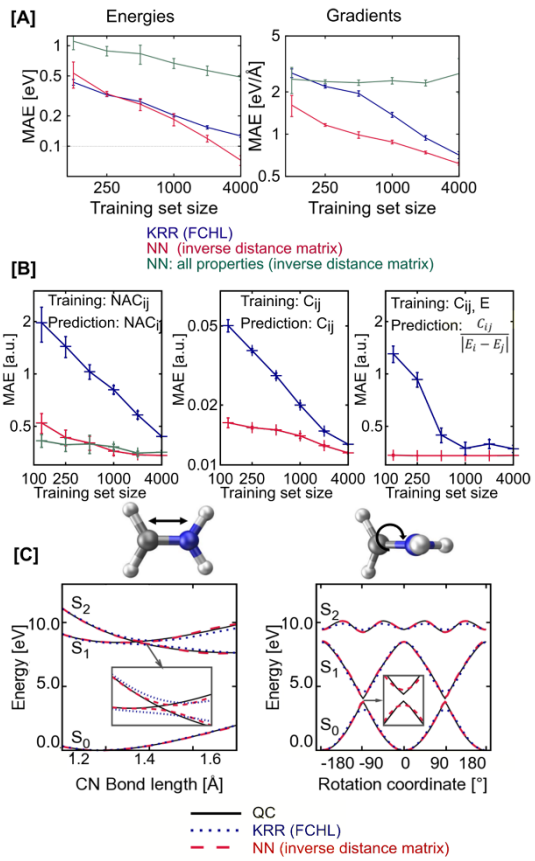

Figure 3. Learning curves showing the mean absolute error (MAE) for the energies and gradients [A] and NACs [B] averaged over all three singlet states (MAEs obtained for the separate energy levels S_0, S_1, and S_2 are shown in the SI in figure S1). [C] Scans along the CN-bond elongation of the methylenimmonium cation (left panel) and the rotation along the dihedral angle (right panel) computed with quantum chemistry (QC, black continuous lines), KRR (blue dotted lines), and NNs (red dashed lines).

coordinates including a critical region (panel C in figure 3) reveals that both ML models, KRR and NNs, can correctly reproduce the shape of the curves. However, there are small deviations in the critical regions of the PESs. A comparison of all these findings implies that an accurate prediction of energies is more important for reproducing the dynamics than an accurate prediction of NACs. Nevertheless, it is intuitively clear that surface-hopping molecular dynamics requires all properties to be accurate enough: Having the correct potentials, but wrong NAC values does also result in wrong dynamics.

3.1. Multi-state representation

A major difference between NN models and KRR models is, that the NNs can predict all electronic states at once, whereas three independent KRR models are used for the three states. However, the PESs are not independent from each other. Consequently, it should also be favourable to learn all states together in one ML model. Encoding of the quantum energy level in an additional state-representation makes multi-state outputs possible. This representation is tested for KRR with the FCHL representation and NNs with the inverse distance matrix.

A comparison of the standard representations and the encoded state representations is given in figure 4(A). All ML models show that learning of all state energies at once is possible and favorable. Encoding of the quantum energy level improves the accuracy of all ML models, whereby this effect is significant for KRR models and small for NNs. With the state representation for KRR, the kernel matrix size increases from $N_m \times N_m$ to $N_S \cdot N_m \times N_S \cdot N_m$. Mapping only a subset of molecules to the complete dataset for training can reduce the memory consumption and can make the training process more efficient without a major loss in accuracy.

Principal component analysis provides an explanation to the above observed results. Therefore, the first principal component is plotted against the second principal component of an ML model in figure 5. The ML models used for surface-hopping dynamics are compared to the ML models with a state representation. Remarkably, a state representation for KRR (figures 5 (B) and (C)) leads to a clear ordering of the data corresponding to different electronic states. Moreover, within one state, a better ordering can be obtained

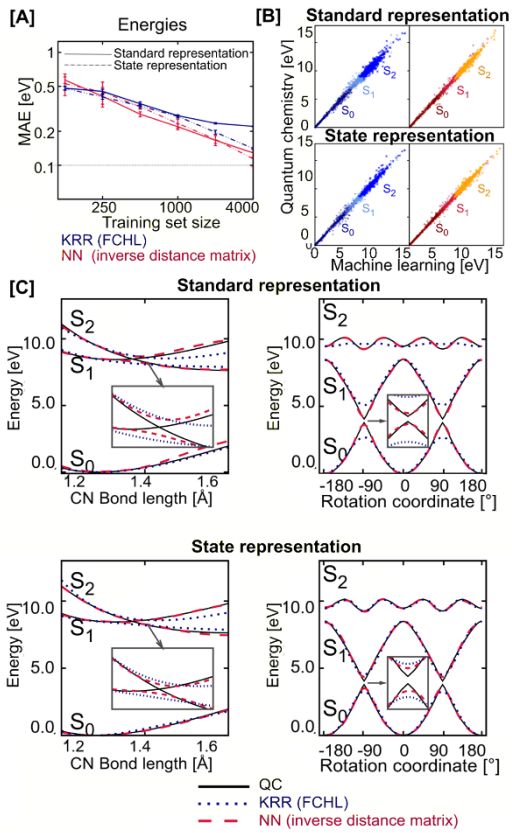

Figure 4. Learning curves showing the mean absolute error (MAE) for the energies obtained from standard and state representations [A] along with corresponding scatter plots [B]. [C] Scans along the CN-bond elongation of the methylenimmonium cation (left panel) and the rotation along the dihedral angle (right panel) computed with quantum chemistry (QC, black continuous lines), KRR (blue dotted lines), and NNs (red dashed lines).

than for KRR without a representation for the electronic states (panel (A)). Similarly, NNs illustrated in panels (D) and (E) show an improved ordering of data, when a representation for the electronic state is used. The application of those models in surface-hopping dynamics, is assumed to lead to even more accurate energy predictions, but requires the gradients as derivatives from ML potentials and as response properties from KRR models. This approach needs further considerations, especially due to the high memory consumption of KRR models, and will be the subject of future developments. Nevertheless, the results clarify how an improved ordering of data can lead to higher prediction accuracy of ML models and thus highlight the importance of the molecular representation in addition to the type of regressor.

4. Conclusion

In this paper, two frequently used ML regressors, namely KRR and NNs, are compared for their application in excited-state molecular dynamics of $CH_2NH_2^+$. The role of the ML model in combination with different representations of the molecular structure for the prediction of energies of the ground state as well as excited states, corresponding forces, NACs between different states, and (transition) dipole moments, is investigated. All ML models are able to learn the relation between a quantum chemical property and the molecular structure, when the properties are treated separately from each other. Learning all properties at once leads to significantly worse results and the learning of single properties can even be impeded when the cost function includes all properties at once.

It is shown that the FCHL representation is in most cases superior to the matrix of inverse distances. Encoding of the quantum energy level in the representation can further improve results and make multiple outputs for KRR possible. The state encoding is shown for three electronic singlet states for KRR as well as NNs. In both cases, the modification of the representation is necessarily accompanied by an enlargement of the ML model, a larger kernel matrix in the case of KRR, and a larger input layer in the case of NNs. Principal component analyses further show that an enhanced ordering of data points is obtained.

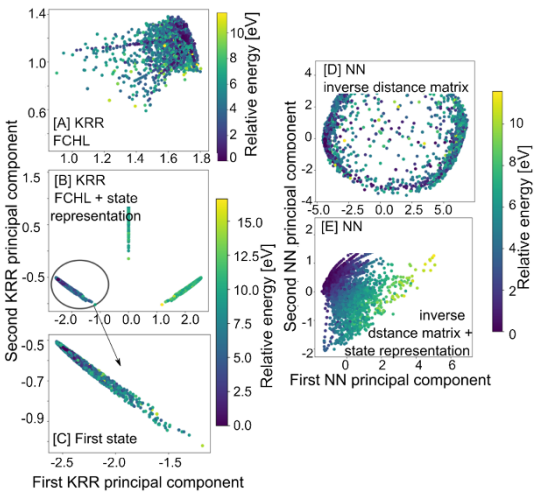

Figure 5. Principal component analysis of the kernel matrices using the FCHL representation and the matrices obtained from the last NN-layers with the inverse distance matrix. The left side gives results for KRR, whereas the right side shows results obtained from NNs.

The NAC vectors pose a real challenge for ML models due to their peaked nature. The inclusion of the energy gap of the coupled pair of states can improve the accuracy of KRR models, but does not allow for more accurate NAC vectors in general. This is due to deteriorating effects of the errors in the energy gaps.

Finally, it is not obvious that one ML model outperforms the other. Different options, such as the size of the training set and the computational power available, may favour different models. The faster hyperparameter optimization for KRR and the higher model flexibility of NNs lets us recommend to use KRR for first exploratory runs and NNs for final production runs of excited-state dynamics simulations. The concept of *wide and deep learning* [71] is interesting for future applications, in the sense that different ML models can be applied within one application to combine their distinct benefits.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

The data that support the findings of this study are openly available at http://dx.doi.org/10.103 9/C9SC01 742A.

Acknowledgments

This work was financially supported by the Austrian Science Fund, W 1232 (MolTag), the uni:docs program of the University of Vienna (J W), and the University of Basel. We additionally want to thank Michael Gastegger from TU Berlin and Sebastian Mai from University of Vienna for their discussion concerning nonadiabatic couplings and suggestions to improve on their prediction. P.M. thanks the University of Vienna for continuous support, also in the frame of the Vienna Research Platform on Accelerating Photoreaction Discovery and the Research Platform Data Science @ Uni Vienna.

O A v L acknowledges funding from the Swiss National Science foundation (No. PP00P2_138932 and 407540_167186 NFP 75 Big Data) and from the European Research Council (ERC-CoG grant QML). This work was partly supported by the NCCR MARVEL, funded by the Swiss National Science Foundation.

ORCID iDs

Julia Westermayr https://orcid.org/0000-0002-6531-0742
O Anatole von Lilienfeld https://orcid.org/0000-0001-7419-0466
Philipp Marquetand https://orcid.org/0000-0002-8711-1533

References

[1] Vass I, Cser K and Cheregi O 2007 *Ann. NY Acad. Sci.* **1113** 114–22
[2] Schönlein R, Peteanu L, Mathies R and Shank C 1991 *Science* **254** 412–15
[3] Garavelli M, Negri F and Olivucci M 1999 *J. Am. Chem. Soc.* **121** 1023–9
[4] Schreier W J, Schrader T E, Koller F O, Gilch P, Crespo-Hernández C E, Swaminathan V N, Charell T, Zinth W and Kohler B 2007 *Science* **315** 625–9
[5] Rauer C, Nogueira J J, Marquetand P and González L 2016 *J. Am. Chem. Soc.* **138** 15911–16
[6] Ahmad I, Ahmed S, Anwar Z, Sheraz M A and Sikorski M 2016 *Int. J. Photoenergy* **2016** 1–19
[7] Häse F, Valleau S, Pyzer-Knapp E and Aspuru-Guzik A 2016 *Chem. Sci.* **7** 5139–47
[8] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 *Nature* **559** 547–55
[9] Liu J and Prezhdo O V 2015 *J. Phys. Chem. Lett.* **6** 4463–9
[10] Domcke W and Sobolewski A L 2013 *Nat. Chem.* **5** 257–8
[11] Liu F, Du L, Zhang D and Gao J 2017 *Sci. Rep.* **7** 1–12
[12] Mai S, Marquetand P and González L 2018 *WIREs Comput. Mol. Sci.* **8** e1370
[13] Mai S, Marquetand P and González L 2016 *J. Phys. Chem. Lett.* **7** 1978–83
[14] Doltsinis N L 2006 *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches* (Jülich, Germany: John von Neuman Institut for Computing) vol 31
[15] Subotnik J E, Jain A, Landry B, Petit A, Ouyang W and Bellonzi N 2016 *Annu. Rev. Phys. Chem.* **67** 387–417
[16] Blank T B and Brown S D 1994 *J. Chemometr.* **8** 391–407
[17] Blank T B, Brown S D, Calhoun A W and Doren D J 1995 *J. Chem. Phys.* **103** 4129–37
[18] Brown D F R, Gibbs M N and Clary D C 1996 *J. Chem. Phys.* **105** 7597–604
[19] Tafeit E, Estelberger W, Horejsi R, Moeller R, Oettl K, Vrecko K and Reibnegger G 1996 *J. Mol. Graphics Modell.* **14** 12–18
[20] No K T, Chang B H, Kim S Y, Jhon M S and Scheraga H A 1997 *Chem. Phys. Lett.* **271** 152–6
[21] Richings G W and Habershon S 2017 *Chem. Phys. Lett.* **683** 228–33
[22] Hu D, Xie Y, Li X, Li L and Lan Z 2018 *J. Phys. Chem. Lett.* **9** 2725–32
[23] Dral P O, Barbatti M and Thiel W 2018 *J. Phys. Chem. Lett.* **9** 5660–3
[24] Chen W K, Liu X Y, Fang W H, Dral P O and Cui G 2018 *J. Phys. Chem. Lett.* **9** 6702–8
[25] Williams D M G and Eisfeld W 2018 *J. Chem. Phys.* **149** 204106
[26] Xie C, Zhu X, Yarkony D R and Guo H 2018 *J. Chem. Phys.* **149** 144107
[27] Richings G W and Habershon S 2018 *J. Chem. Phys.* **148** 134116
[28] Guan Y, Zhang D H, Guo H and Yarkony D R 2019 *Phys. Chem. Chem. Phys.* **21** 14205–13
[29] Westermayr J, Gastegger M, Menger M F S J, Mai S, González L and Marquetand P 2019 *Chem. Sci.* **10** 8100–7
[30] Guan Y, Guo H and Yarkony D R 2020 *J. Chem. Theory Comput.* **16** 302–13
[31] Krems R V 2019 *Phys. Chem. Chem. Phys.* **21** 13392–13410
[32] Richings G W, Robertson C and Habershon S 2019 *J. Chem. Theory Comput.* **15** 857–70
[33] Polyak I, Richings G W, Habershon S and Knowles P J 2019 *J. Chem. Phys.* **150** 041101
[34] Guan Y, Guo H and Yarkony D R 2019 *J. Chem. Phys.* **150** 214101
[35] Wang Y, Xie C, Guo H and Yarkony D R 2019 *J. Phys. Chem.* A **123** 5231–41
[36] Westermayr J, Gastegger M and Marquetand P 2020 *J. Phys. Chem. Lett.* **11** 3828–34
[37] Behler J 2015 *Int. J. Quantum Chem.* **115** 1032–50
[38] Rupp M 2015 *Int. J. Quantum Chem.* **115** 1058–73
[39] Li Z, Kermode J R and De Vita A 2015 *Phys. Rev. Lett.* **114** 096405
[40] Gastegger M, Behler J and Marquetand P 2017 *Chem. Sci.* **8** 6924–35
[41] Botu V, Batra R, Chapman J and Ramprasad R 2017 *J. Phys. Chem.* C **121** 511–22
[42] Artrith N, Urban A and Ceder G 2017 *Phys. Rev.* B **96** 014112
[43] Behler J 2017 *Angew. Chem. Int. Edit.* **56** 12828–40
[44] Chmiela S, Tkatchenko A, Sauceda H E, Poltavsky I, Schütt K T and Müller K R 2017 *Sci. Adv.* **3** e1603015
[45] Schütt K T, Gastegger M, Tkatchenko A, Müller K R and Maurer R J 2019 *Nat. Commun.* **10** 5024
[46] Behler J, Reuter K and Scheffler M 2008 *Phys. Rev.* B **77** 115421
[47] Carbogno C, Behler J, Reuter K and Groß A 2010 *Phys. Rev.* B **81** 035410
[48] Barbatti M, Aquino A J A and Lischka H 2006 *Mol. Phys.* **104** 1053–60
[49] Tapavicza E, Tavernelli I and Rothlisberger U 2007 *Phys. Rev. Lett.* **98** 023001
[50] Tavernelli I, Tapavicza E and Rothlisberger U 2009 *J. Chem. Phys.* **130** 124107
[51] Tavernelli I, Tapavicza E and Rothlisberger U 2009 *J. Mol. Struct.: Theor.* **914** 22–9
[52] Hudock H R, Levine B G, Thompson A L, Satzger H, Townsend D, Gador N, Ullrich S, Stolow A and Martínez T J 2007 *J. Phys. Chem.* A **111** 8500–8
[53] Hollas D, Šištík L, Hohenstein E G, Martínez T J and Slavíček P 2018 *J. Chem. Theory Comput.* 14 339–350
[54] Mai S, Richter M, Ruckenbauer M, Oppel M, Marquetand P and González L 2018 Sharc2.0: Surface hopping including arbitrary couplings – program package for non-adiabatic dynamics (sharc-md.org)
[55] Wigner E 1932 *Phys. Rev.* **40** 749–50
[56] Tennyson J 2016 *J. Chem. Phys.* **145** 120901
[57] Gastegger M and Marquetand P 2018 arxiv:1812.07676 [physics.chem-ph]
[58] Cortes C, Jackel L D, Solla S A, Vapnik V and Denker J S 1994 Learning curves: Asymptotic values and rate of convergence ed Cowan J D, Tesauro G and Alspector J *Advances in Neural Information Processing Systems 6* (San Mateo, CA: Morgan Kaufmann Publishers) pp 327–34
[59] Müller K R, Finke M, Murata N, Schulten K and Amari S 1996 *Neural Comput.* **8** 1085–106
[60] von Lilienfeld O A 2018 *Angew. Chem. Int. Edit.* **57** 4164–9
[61] Christensen A S, Faber F A and von Lilienfeld O A 2019 *J. Chem. Phys.* **150** 064105
[62] Christensen A, Faber F, Huang B, Bratholm L, Tkatchenko A, Müller K and Lilienfeld O A 2017 *QML: A Python toolkit for quantum machine learning* (https://github.com/qmlcode/qml)
[63] Gastegger M and Marquetand P 2015 *J. Chem. Theory Comput.* **11** 2187–98
[64] Chmiela S, Sauceda H E, Müller K R and Tkatchenko A 2018 *Nat. Commun.* **9** 3887

[65] Schütt K T, Sauceda H E, Kindermans P J, Tkatchenko A and Müller K R 2018 *J. Chem. Phys.* **148** 241722
[66] Faber F A, Christensen A S, Huang B and von Lilienfeld O A 2018 *J. Chem. Phys.* **148** 241717
[67] Christensen A S, Bratholm L A, Faber F A and von Lilienfeld O A 2020 *J. Chem. Phys.* **152** 044107
[68] van der Walt S, Colbert S C and Varoquaux G 2011 *Comput. Sci. Eng.* **13** 22–30
[69] Al-Rfou R *et al* (The Theano Development Team) 2016 arXiv:abs/1605.02688
[70] Kingma D P and Ba J 2014 arXiv:1412.6980 [cs.LG]
[71] Cheng H T, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong L, Jain V, Liu X and Shah H 2016 arXiv:1606.07792