

# Neural Networks vs Gaussian Process Regression for Representing Potential Energy Surfaces: a Comparative Study of Fit Quality and Vibrational Spectrum Accuracy

Aditya Kamath<sup>1</sup>, Rodrigo A. Vargas-Hernández<sup>2</sup>, Roman V. Krems<sup>2</sup>, Tucker Carrington Jr.<sup>3</sup>, and Sergei Manzhos<sup>1,a</sup>

<sup>1</sup>Department of Mechanical Engineering, National University of Singapore, Block EA #07-08, 9 Engineering Drive 1, Singapore 117576

<sup>2</sup>Department of Chemistry, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada

<sup>3</sup>Department of Chemistry, Queen's University, Chernoff Hall, 90 Bader Lane, Kingston, ON, Canada K7L 3N6

## Abstract

For molecules with more than three atoms, it is difficult to fit or interpolate a potential energy surface (PES) from a small number of (usually *ab initio*) energies at points. Many methods have been proposed in recent decades each claiming a set of advantages. Unfortunately, there are few comparative studies. In this paper, we compare neural networks (NN) with Gaussian process (GP) regression. We re-fit an accurate PES of formaldehyde and compare PES errors on the entire point set used to solve the vibrational Schrödinger equation, i.e. the only error that matters in quantum dynamics calculations. We also compare the vibrational spectra computed on the underlying reference PES and the NN and GP potential surfaces. The NN and GP surfaces are constructed with exactly the same points and the corresponding spectra are computed with the same points and the same basis. The GP fitting error is lower and the GP spectrum is more accurate. The best NN fits to 625/1250/2500 symmetry unique potential energy points have global PES root mean square errors (RMSE) of 6.53/2.54/0.86 cm<sup>-1</sup>, whereas the best GP surfaces have RMSE values of 3.87/1.13/0.62 cm<sup>-1</sup>, respectively. When fitting 625 symmetry unique points, the error the first 100 vibrational levels is only 0.06 cm<sup>-1</sup> with the best GP fit, whereas the spectrum on the best NN PES

---

<sup>a</sup> Author to whom correspondence should be addressed. E-mail: mpemanzh@nus.edu.sg; tel.: +65 6516 4605 ; fax: +65 6779 1459.

has an error of  $0.22 \text{ cm}^{-1}$ , with respect to the spectrum computed on the reference PES. This error is reduced to about  $0.01 \text{ cm}^{-1}$  when fitting 2,500 points with either NN or GP. We also find that the GP surface produces a relatively accurate spectrum when obtained based on as few as 313 points.

## Introduction

Most calculations of observables such as reaction rates or vibrational spectra require a potential energy surface (PES). The PES is a consequence of the Born-Oppenheimer approximation. To construct a PES, one must either interpolate between potential energy points in a multi-dimensional configuration space or find a suitable analytical representation of the global PES. There are many common fitting and interpolation methods and it is impossible to cite them all. Some representative references are Refs. 1-5. To compute a spectrum of spectroscopically useful accuracy (e.g. errors smaller than  $1 \text{ cm}^{-1}$ ), it is necessary to obtain a very accurate (errors on the order of  $1 \text{ cm}^{-1}$ ) representation of the PES. For molecules with more than three atoms this is difficult due to a large number of ab initio points required and the complexity of multi-dimensional PES.<sup>6-7</sup> It can be argued that, for complex molecules, methods based on machine learning designed for fitting multi-dimensional functions are particularly appealing due to their flexibility and potentially favorable dimensionality scaling. In particular, representing PES by artificial neural networks (NN) has become popular in recent years. An alternative class of methods for constructing a PES is based on Bayesian inference, leading to a non-parametric description of potential energy points as draws from probability distributions. One example of such methods is kriging, a kernel interpolation method representing the potential surface with a Gaussian process (GP) determined by covariances between potential energy points. Although GP regression does not provide an analytical PES function, there is evidence that GP makes it possible to do dynamics calculations using only a small number of potential energy points.<sup>6-10</sup> In this paper, we compare GP regression and neural network (NN) fitting methods to compute the vibrational spectrum of formaldehyde.

All of the common PES fitting methods have advantages and disadvantages. Naturally, scientists tend to prefer methods they developed or [methods](#) they are familiar with. In previous papers, we pursued NN-based approaches because (i) NN theorems guarantee the universal approximator quality of a single-hidden layer NN.<sup>11-12</sup> (ii) an NN can be used to produce sum-of-

product PESs which facilitate the use of quantum dynamics methods,<sup>13-16</sup> (iii) NNs can be easily used as building blocks in conjunction with other approaches such as HDMR (High Dimensional Model Representation) or the multimode expansion<sup>17-20</sup> and (iv) NNs easily allow the realization dimensionality reduction<sup>21-22</sup> which is advantageous when data are very sparse.<sup>21,23</sup> Very importantly, NNs have an important advantage that there are user-friendly and readily available packages in environments such as Matlab, R, Octave, and Python. In most cases, one can get a good NN fit by changing a single parameter – the number of neurons. In principle, one can also select the shape of the neuron.<sup>24</sup> But in practice this is not required, as sigmoid neurons provide the best fit quality for most systems (and for all systems we have studied).<sup>25</sup> An exponential neuron is required to obtain the sum-of-product form.<sup>13-14</sup> In Refs. 26 and 13, it was shown for the first time that NNs can provide spectroscopically accurate potentials, and potentials in sum-of-product form with a reasonably small number of terms, respectively. NNs also have disadvantages: they extrapolate poorly (which is true for most black-box methods) and require non-linear fitting, which is costlier than linear fitting. Neither of these is severe. At most several hours are needed with the largest NNs we used, including those in this paper, when using modern computers (including multicore desktop workstations).

Unfortunately, there are very few truly comparative studies of fitting methods. One such comparative study is Ref. 27. There, we compared an NN committee fit,<sup>25</sup> a interpolated moving least squares (IMLS) with permutationally invariant polynomials (PIP) fit, and a combined PIP-NN fit of the nine-dimensional PES of methane (fit in the ten-dimensional space of interatomic distances). State-of-the-art *ab initio* points were used to fit the PES, and the PESs fitted with different methods were used to compute accurate vibrational levels. The root mean square error (RMSE) of the PES representation for energies up to 20,000  $\text{cm}^{-1}$  above the minimum with NN/PIP-NN/PIP-IMLS was 2.47/1.58/0.85  $\text{cm}^{-1}$ . These are test-set errors, where a *small* subset of points was selected as the test set from the total set of about 100,000 *ab initio* energies. With PIP-NN, the test set was 5% of the total dataset, with PIP-IMLS, 25%, and with NN fits, 20%. The quality of the levels was remarkably similar among the PESs fit with the three methods: the difference in the equilibrium CH bond length was about 0.00001 Å and differences in fundamental frequencies were on the order of 1  $\text{cm}^{-1}$ . Moreover, the RMSE of the first 36 transitions with respect to experimental values was 4.24  $\text{cm}^{-1}$  with PIP-IMLS and 3.55/3.82  $\text{cm}^{-1}$  with NN committees with 3/26 NNs (the last two numbers were not included in Ref. 27). The RMSE of the

spectra computed with all three methods differ by less than  $0.5 \text{ cm}^{-1}$ . NN fitting therefore appears to work well. *We note, however, that neither the size of the test set, nor the size of the total ab initio dataset were sufficient to densely sample the nine-dimensional space.* This is almost always the case when a PES for a molecule with more than three atoms is fitted. Because it is now possible to compute vibrational levels of molecules with more than three atoms,<sup>28-32</sup> it is important to test the most promising methods for fitting or interpolating surfaces.

In this paper, we point out that it is possible to compute very accurate levels with a quantum dynamics method that uses PES points with the same distribution as the fitting set and identical to those in the test set. If instead, as is common, one uses a direct product Gauss quadrature, then the PES is evaluated at points far from equilibrium and in regions in which there are few or no fitting (or test) points. Due to disparity of the fitting and evaluation distributions, the RMSE of a PES fit at test points provides an incomplete picture.

Recently, Gaussian processes (GP) regression has been proposed as an efficient PES representation tool.<sup>6-10,33,34</sup> Like NNs, GPs have the advantages of generality and ease of use. There are GP software packages for Matlab, R, Octave, and Python. GP regression is a non-parametric interpolation method; in the version presented here, it is trained by finding maximum likelihood estimates of a small number of hyper-parameters. To obtain and use a GP representation of a surface, one selects an appropriate correlation function. It was recognized in Ref. 7 that the best function is system dependent; that is also our experience (see below). It was also noted in Ref. 7 that a specific selection of fitting points may be required to get a good fit; this is not the case with NNs, where any random point selection with a given distribution results in similar errors. For the formaldehyde PES, we show here that a random selection of GP fitting points is sufficient to yield an accurate PES. In Ref. 7, it was suggested that the dimensionality scaling of GP is so good that “a 6D PES can be obtained with only 60 quantum chemistry calculations”. It was also stated that “Artificial neural networks are expected to require many more ab initio points”. Here, we directly test these propositions.

In this paper, we compare NN and GP for the purpose of obtaining a PES accurate enough that vibrational levels computed on it have errors of less than  $1 \text{ cm}^{-1}$ . Errors on GP PESs reported previously are rather large. For example, Ref. 7 reports an RMSE of about  $500 \text{ cm}^{-1}$  with 2,400 fitting points for a four-atom molecule ( $\text{N}_4$ ) in an energy range of about  $35,000 \text{ cm}^{-1}$ . In Ref. 6, a test RMSE of about  $17/7 \text{ cm}^{-1}$  (fit/prediction) was obtained with a GP PES for a triatomic molecule

with 3,741 data points – a rather dense set in 3D. Reducing the number of fitting points to 1,000 increases the error to about 24/35  $\text{cm}^{-1}$ . To put these numbers in perspective, in Ref. 26 we obtained a median test error of 0.5  $\text{cm}^{-1}$  (fit RMSE 1  $\text{cm}^{-1}$ ) with 1,500 points in 3D (for  $\text{H}_2\text{O}$ ) in a range 20,000  $\text{cm}^{-1}$ . In 6D, for  $\text{H}_2\text{O}_2$  (using 5,000 points up to 15,000  $\text{cm}^{-1}$ ) and for  $\text{H}_2\text{CO}$  (using 2,500 points up to 17,000  $\text{cm}^{-1}$ ), we obtained median test errors on the order of 1  $\text{cm}^{-1}$  (fit RMSE 2  $\text{cm}^{-1}$  for both molecules). The GP fits cited above were mostly employed with classical dynamics and 3D quantum reaction dynamics calculations, where the accuracy requirements are not as strict as for vibrational calculations. The 6D NN fits of Ref. 26 were done with a two-stage NN to avoid holes in the PES, which further complicates comparisons. In summary, with the available numbers, it is hard to assess the utility of GP PESs for vibrational calculations.

In this paper, we make a *direct* comparison by using GP and NN to re-fit a reference PES for  $\text{H}_2\text{CO}$ . We compare vibrational spectra computed on the NN and GP PESs. Levels are computed with the method of Manzhos and Carrington that uses a space-fixed Cartesian kinetic energy operator and Gaussian basis functions.<sup>35</sup> It allows us to compare directly levels computed on the two fitted PESs with those computed on the reference PES from which they are obtained. We confirm that the GP PES has somewhat smaller errors than the NN PES obtained using the same set of fit points and the same set of 120,000 test points; enough points are chosen to make both PESs spectroscopically *accurate*.

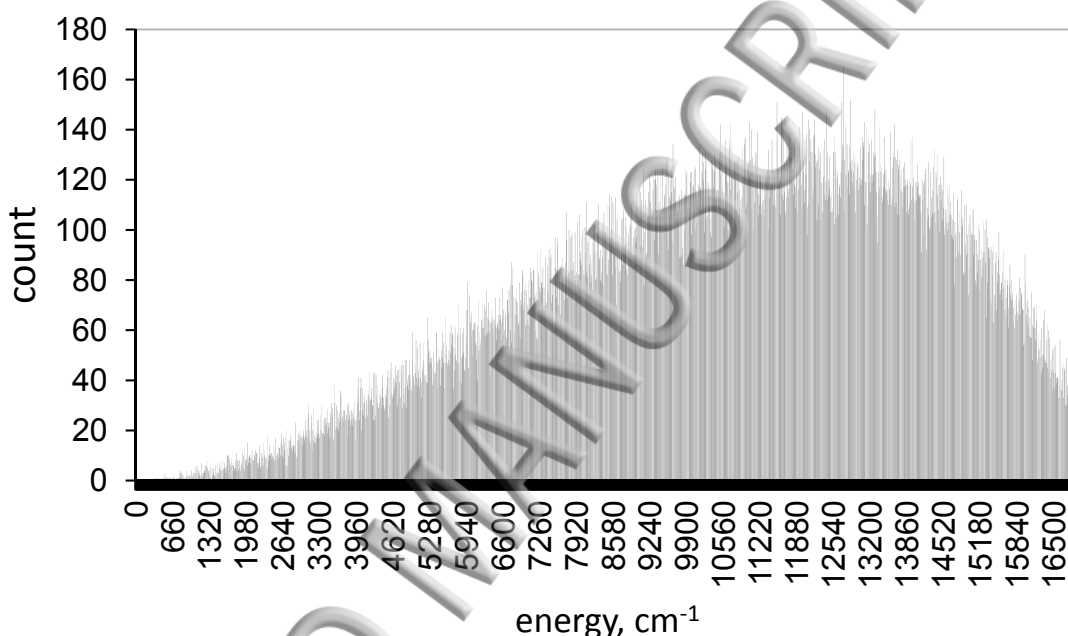
## Methods

The GP and NN PESs of this paper are obtained by re-fitting the PES of Ref. 36. We note that although GP regression is not a fitting technique, we will in the following use the words “fit” and “fit error” when discussing both the NN and GP PES. The vibrational spectrum computed on this reference PES is close to the experimental spectrum.<sup>35</sup> Both the GP and the NN PESs are constructed from a set of points in bond coordinates  $\boldsymbol{x} = (\text{CO}, \text{CH}_1, \text{CH}_2, \angle\text{OCH}_1, \angle\text{OCH}_2, \angle\text{H}_1\text{COH}_2)$ , where  $\angle\text{H}_1\text{COH}_2$  is the dihedral angle between the OCH planes. The points are a pseudo-random Sobol sequence<sup>37</sup> and we accept points if

$$\frac{V_{\max} - V(\boldsymbol{x}) + \Delta}{V_{\max} + \Delta} > \text{rand} \quad , \quad (1)$$



where  $rand$  is a (uniformly distributed) random number in  $[0, 1]$ . We use  $V_{max} = 17,000 \text{ cm}^{-1}$  and  $\Delta = 500 \text{ cm}^{-1}$ . The coordinate ranges are determined by  $\mathbf{x}_{min} = (1.03, 0.84, 0.84, 83, 83, 105)$ ,  $\mathbf{x}_{max} = (1.50, 1.69, 1.69, 162, 162, 255)$ , where bond lengths are in  $\text{\AA}$  and angles in degrees. We and others have shown previously that with this point selection scheme it is possible to obtain an accurate PES and accurate vibrational spectra.<sup>26,35</sup> In this way, a data set of 120,000 points was produced distributed over energy values as shown in Fig. 1.



**Figure 1.** Distribution of energy values of the PES dataset.

We will refer to the potential energy points used to “train” the NN or GP models for constructing the PES as the “training set” or the “fitting points”. The training sets are small subsets of the 120,000-point set, and PES errors are computed on the entire set. The training sets are drawn randomly using a uniform distribution, i.e. they maintain the distribution of Eq. (1). As the full dataset is sufficient to produce an accurate vibrational spectrum, the errors we report characterize the region of the PES required to compute the lowest vibrational states. In both the NN and GP fits, we augment both the training and test sets with symmetrically equivalent points obtained by flipping the dihedral angle around the planar equilibrium geometry. The numbers of training points we report is the number of symmetry unique points (only they must be computed *ab initio*). In

In NN and GP fits, we average PES values at symmetry unique points and image points. This slightly decreases the error of the PESs and of the vibrational levels. We report NN results obtained by averaging over 10 NNs, each with a different set of initial parameters and fitting different (randomly drawn) sets of points. This is referred to as a committee of 10. The averaging removes the random component of the fitting error.<sup>25</sup> In the GP case, we also averaged over 10 fits. The different GP fits differ by the choice of the training points.

For the description of NN and GP regression methods to fitting potential energy surfaces, see Refs. 25 and 6, 7 respectively. Here, we only provide details needed for reproducibility. The NN fits with single-hidden layer NNs with sigmoid neurons and linear output neurons are done using Matlab's Neural Network Toolbox using the *newfit* and *train* functions, on data scaled to [0, 1], as is customary. That is, we use a simple NN setup readily available in publically available software. Training was done for 2,000 epochs with the Levenberg-Marquardt algorithm. The GP fits are done in Python using the *gaussian\_process* function from module *sklearn*. An initial set of hyper-parameters is optimized by maximizing the *log* likelihood function using optimizer *fmin\_cobyla* from module *scipy* to find the best combination of hyper-parameters providing the estimates of the correlations between the potential energy points in the training set. The data are also scaled in the GP fits. Different kernel functions were tried. We find that the "squared exponential" kernel function yielded the best results. We thus used different software packages for NN and GP. Although both Matlab and Python have NN and GP capabilities, the errors we obtained with GP in Matlab were much larger than the GP errors in Python, and the errors we obtained with NN using Python modules were much higher than the errors obtained with Matlab. We used the best software for each method.

Vibrational spectra were computed with the method of Manzhos and Carrington that uses the space-fixed Cartesian kinetic energy operator and Gaussian basis functions.<sup>35</sup> We shall refer to this as the SFGB (Space-Fixed Gaussian Basis) method. The SFGB method allows us to compute the vibrational spectrum using (only) the 120,000 points from which the fitting points are selected. We used 40,000 Gaussian basis functions centered on the first 40,000 points. The widths of all basis functions were the same and are chosen as described in Ref. 35. We computed the first 100 vibrational levels. Spectra computed on the NN and GP PESs were compared to a reference spectrum computed on the PES of Ref. 36 with a variational method. The description of the

variational method is in Ref. 35. In this paper, we use more quadrature points and basis functions than in Ref. 35.

**Table 1.** RMSE (test errors computed on 120,000 points) of the PES obtained with the NN and GP methods for different numbers of (symmetry unique) fitting points  $N_{pts}$ , with single NN/GP fits and with a committee of 10 fits (<10 NN/GP>). The NN RMSE values separated by “/” are for 100/150/250 neurons per NN for 2,500 points, 70/100/150 neurons per NN for 1,250 points 50/75/100 neurons for 625 points, and 20/30/40/50 neurons for 313 points. The values are in  $\text{cm}^{-1}$ .

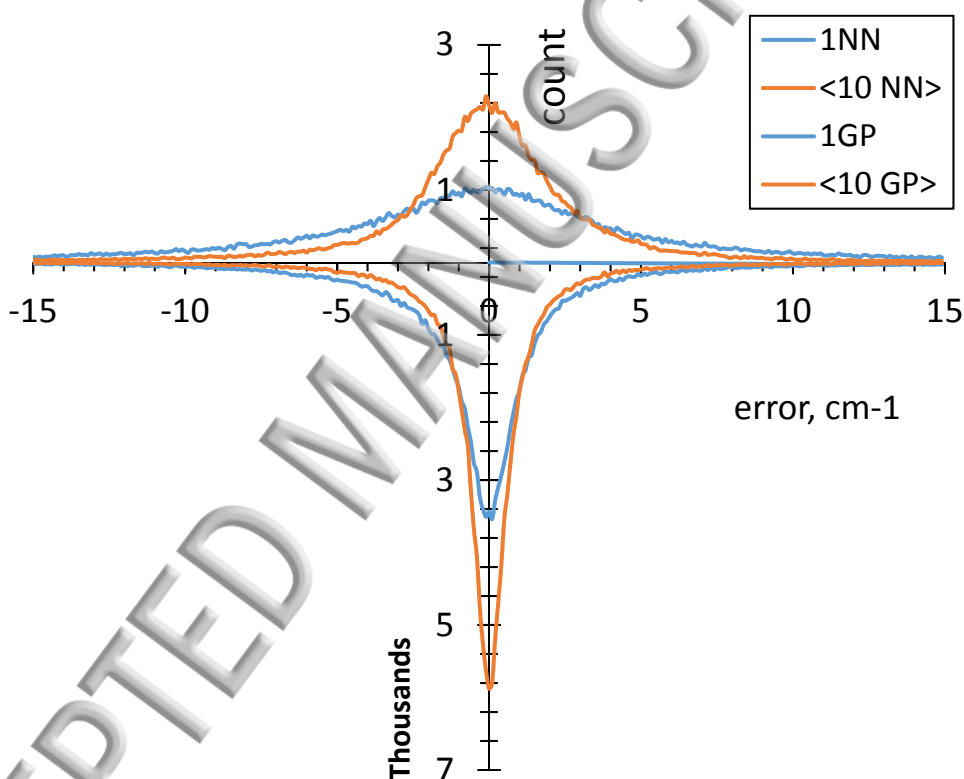
No. of fit points $N_{pts}$	NN		GP	
	1 NN	<10 NN>	1 GP	<10 GP>
313	198.00/103.93/87.77	119.11/53.97/43.90	29.09	17.18
625	21.12/12.91/12.03	13.36/7.52/6.53	5.98	3.87
1,250	9.29/5.74/4.38	5.74/3.36/2.54	2.17	1.13
2,500	4.59/2.43/1.12	2.27/1.23/0.86	1.08	0.62

## Results

The errors of the PESs obtained with the NN and GP methods, when fitting with different numbers of points, are given in Table 1. These errors are computed by averaging over all the points used to solve the Schrödinger equation. It is errors at these points, and *only* errors at these points that determine the accuracy of the energy levels. When fitting with NN or GP regression, the RMSE changes from fit to fit (due to random selection of fitting points and, in the case of NN, random parameter initialization). For GP, the RMSE of different fits differ by only about  $\pm 10\%$  or on the order of  $1 \text{ cm}^{-1}$ . For NN, the difference between errors from fit to fit is about 30%. When  $N_{pts} = 313$ , the fit-to-fit RMSE variation is larger; this and the large magnitude of RMSE values indicate that 313 points are not sufficient. The numbers in the second and fourth columns are the RMSE for the first of the 10 fits. The numbers in the third and fifth columns are average errors. Averaging improves the NN errors somewhat more when the number of points is smaller. The numbers of neurons  $N$  for all NN fits in Table 1 are such that all parameters are well determined (the number of parameters is smaller than the number of data). The number of parameters is



$(d+2) \times N+1$  with  $d=6$  the dimensionality and  $N$  the number of neurons; the number of fitted data is  $N_{\text{pts}} \times 2$  (due to symmetry equivalents). A high-quality PES can be obtained with  $N_{\text{pts}} = 625$  or more symmetry unique points. The distributions of the errors for  $N_{\text{pts}} = 625$  are shown in Fig. 2. They are qualitatively similar for other values of  $N_{\text{pts}}$ . It is clear from Table 1 and Fig. 2 that GP is able to achieve a smaller error with each of the fitting point sets. NN requires more fitting data to achieve a similar error. Furthermore, this error can be substantially reduced (by up to a factor of two) by using a committee of NN or GP fits. With 2,500 points, NN fits are practically as good as GP fits.



**Figure 2.** Distributions of the errors obtained with 100-neuron NN (top) and GP (bottom) fits to 650 points, for single-NN/GP fits as well as for committees of 10 fits.

A comparison of level errors on the GP and NN PESs is given in Table 2. All PESs are obtained from 625 symmetry unique points. The levels are calculated with the SFGB method. The errors in the top part of the table are computed with respect to levels calculated on the reference PES by using a quadrature-based variational method that uses a large basis, a direct product Gauss

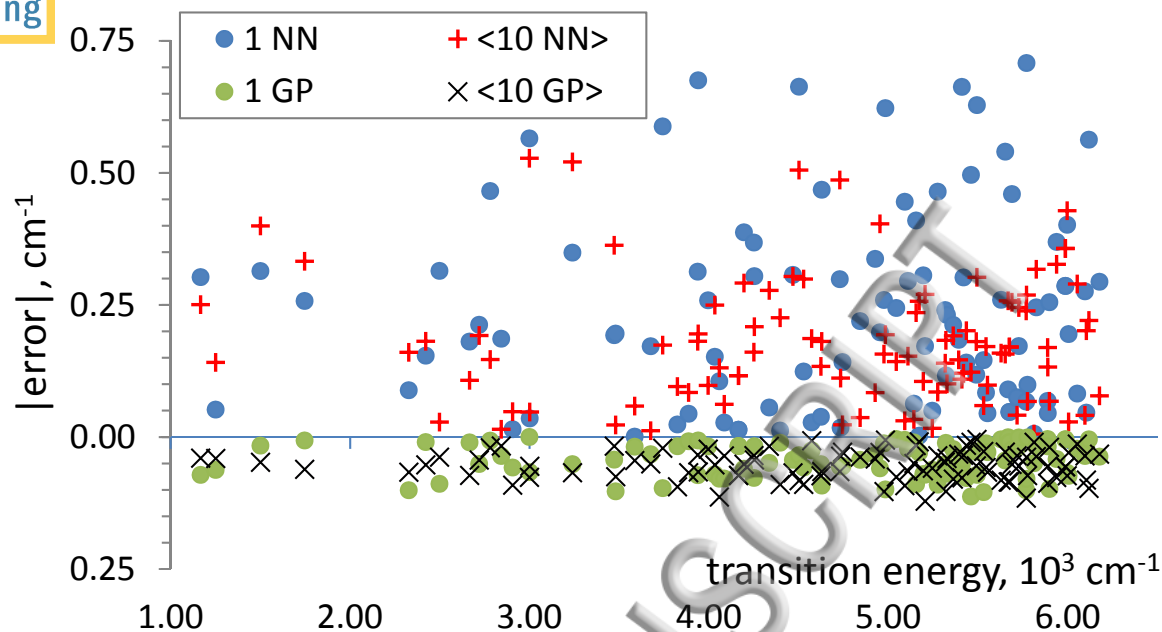
quadrature grid, and the Lanczos algorithm.<sup>26</sup> The errors in the bottom part of the table are computed with respect to levels calculated on the reference PES by using the SFGB method. The errors in columns 3-6 in the top part of the table conflate errors in the SFGB method and errors in the fitting of the PESs. In the bottom part of the table, the errors are due entirely to errors in the fitting of the PESs. A list of all 100 levels and errors, for different  $N_{\text{pts}}$ , is given in Supplementary Material. Absolute errors in differences between levels and the corresponding ZPEs are shown in Fig. 3. Similar plots for other  $N_{\text{pts}}$  are given in Supporting Material. The GP fits significantly outperform the NN fits. When the same method is used to compute the spectra on both the reference PES and the fitted (to 625 points) PES, the RMSE of the 100 lowest vibrational levels is only  $0.06 \text{ cm}^{-1}$  on the 10-member GP PES. The corresponding ZPE error is  $0.1 \text{ cm}^{-1}$ . With a single GP, the errors are even smaller. Therefore, although the RMSE is reduced by averaging, the levels are not improved (they are slightly improved when using more points, see Supplementary Material). This is also seen in Fig. 3. The level errors are higher with NN: a 10 NN committee has an error for the lowest 100 levels of  $0.22 \text{ cm}^{-1}$  and a ZPE error of  $0.07 \text{ cm}^{-1}$ . The NN committee much improves the spectrum quality over a single NN.

Although with 625 symmetry unique points the GP is certainly superior, all of the NN and GP errors are very low. We have confirmed that, as could be surmised from the PES errors, with 2,500 symmetry unique points, an NN PES is at least as accurate as the GP PES with 625 points. The RMSE for the lowest 100 levels is  $0.024/0.017 \text{ cm}^{-1}$  with 1/10 NNs and  $0.005/0.003 \text{ cm}^{-1}$  with 1/10 GPs; see the Supplementary Material. Comparing the top and bottom parts of the table reveals that the errors introduced by GP and NN PES fitting methods are much smaller than the difference between the SFGB and the variational method of computing the spectrum, which is itself small, on the order of  $1 \text{ cm}^{-1}$  RMSE. With GP regression, it is possible to maintain an accurate spectrum when fitting as few as 313 points: with 1/10 GP, we obtain errors of  $1.97/1.57 \text{ cm}^{-1}$  for 100 levels vs the variational calculation and  $1.25/0.32 \text{ cm}^{-1}$  vs the reference PES. Much larger errors are obtained when fitting 313 points with NNs, although an NN committee is able to produce a relatively accurate PES:  $10.3/1.87 \text{ cm}^{-1}$  vs the variational calculation and  $10.5/1.22 \text{ cm}^{-1}$  vs the exact PES are obtained with 1/10 NNs, see Supplementary Material.

Table 2. The RMSE (in  $\text{cm}^{-1}$ ) of the first 50 and 100 vibrational levels of  $\text{H}_2\text{CO}$  with respect to reference levels.  $N_{\text{pts}} = 625$  for both the NN and GP surfaces. For all levels except the ZPE, errors of differences from ZPEs are presented.

	Ref. PES	NN		GP	
		1 NN	<10 NN>	1 GP	<10 GP>
RMSE with respect to variational spectrum on the Ref. PES					
ZPE error	-0.23	-0.41	-0.17	-0.16	-0.13
rmse 100 levels	0.82	0.82	0.83	0.82	0.81
rmse 50 levels	0.47	0.43	0.47	0.46	0.45
RMSE with respect to SFGB spectrum on the Ref. PES					
ZPE error		-0.18	0.07	0.07	0.10
rmse 100 levels		0.30	0.22	0.05	0.06
rmse 50 levels		0.21	0.16	0.04	0.04

Why is it that with only 625 symmetry unique points it is possible to make PESs on which levels have sub- $\text{cm}^{-1}$  errors, whereas in a previous NN fit for formaldehyde<sup>26</sup> more than 2,000 points were required to achieve similar level errors? In Ref 26 a much better (hence more points) PES was required because the levels were computed using a quadrature grid most of whose points are in regions where the density of fitting/test points is low. To account for this disparity of the fitting point and evaluation point distributions, it is necessary to use more fitting points. None of the numbers in columns 3-6 are affected by this disparity because the fitting points are a subset of points used by the SFGB method, and *all* points used by the SFGB method are included in the test RMSE calculation. When this is the case, level errors are much smaller than PES RMSEs.



**Figure 3.** Absolute errors on transition frequencies computed on PESs fitted to 625 points with different methods (NN: top half, GP: bottom half) vs the spectrum computed on the reference PES. Similar plots for PESs obtained with 313 and 2500 training points are shown in Supplementary Material.

### Discussion and conclusions

We have presented the first comparison of the neural network and Gaussian process regression approaches for representing a potential energy surface. The two methods were used under exactly the same conditions to build six-dimensional PESs of formaldehyde, from a reference PES. Vibrational spectra were computed on the NN and GP PESs with a method that uses the same point distribution as that used to fit the PES. The points used to fit the PESs are a subset of those used to compute the vibrational levels, and *all* points used to compute the vibrational spectrum are included in the RMSE calculation of the PES fits.

The accuracy of the NN and GP PESs is assessed in two ways: 1) we computed the RMSE on the entire set of 120,000 points used to compute the spectrum; 2) we compared energy levels calculated on the NN and GP PESs with levels computed on the reference PES. We find that GP outperforms NN for formaldehyde. Fitting only 625 symmetry-unique points with GP, a global RMSE of about  $6 \text{ cm}^{-1}$  can be achieved and reduced to about  $4 \text{ cm}^{-1}$  by using a committee of 10

GP fits. NN fits to the same data result in rmse values of about 12 and 7  $\text{cm}^{-1}$  for a single NN and a committee of 10 NNs, respectively. Increasing the number of points to 1,250 decreases the NN RMSE to about 4 and 2  $\text{cm}^{-1}$  with a single NN and a committee of 10 NNs, respectively.

The relative performance of the two methods with respect to the quality of the spectrum reflects that of the PES RMSEs. The error (rmse) of the lowest 100 vibrational levels of  $\text{H}_2\text{CO}$  on a GP PES fitted to 625 points is only 0.05  $\text{cm}^{-1}$ . Little improvement is observed with a committee of 10 GP fits. In the NN case, the error for a single NN is 0.3  $\text{cm}^{-1}$  and the error for the committee of 10 is 0.2  $\text{cm}^{-1}$ . Spectra accurate to 0.01  $\text{cm}^{-1}$  can be obtained by fitting 2,500 points. When fitting 625 or more points, both fitting methods thus resulted in a PES on which the spectrum is quite accurate and both spectra deviate by a similar amount from the reference spectrum obtained with a variational method, by about 1.6/0.9  $\text{cm}^{-1}$  RMSE for the first 100/50 levels. With GP regression, it is possible to maintain an accurate spectrum when fitting as few as 313 points, with spectral errors on the order of 1  $\text{cm}^{-1}$ . When the sampling density is this low, the quality of the spectrum on NN-fitted PESs is degraded: errors are on the order of 10  $\text{cm}^{-1}$  for single-NN fits, although a committee of NN is able to reduce the error toward 1  $\text{cm}^{-1}$ .

The GP regression and NN methods are conceptually different approaches. Whereas NNs fit a complex multi-dimensional function with superpositions of simple functions, GP regression can be viewed as a weighted average of the training points, with the weights determined by covariances between the *ab initio* points. It is thus instructive to discuss some of the most important relative advantages and disadvantages of the two methods.

- (i) *Cost of evaluation.* The NN fits are generally faster to evaluate. For a single-hidden-layer NN, the cost of evaluation at point  $\mathbf{x}$  is dominated by the matrix-vector multiplication  $\mathbf{W}\mathbf{x}$  with  $\mathbf{W}$  of size  $N \times d$ <sup>25</sup>, where  $N$  is the number of neurons and  $d$  is the dimensionality of the configuration space. To evaluate a PES produced by GP regression at point  $\mathbf{x}$ , one must evaluate a product  $\mathbf{k}^*\mathbf{K}^{-1}\mathbf{y}$ , where  $\mathbf{K}$  is the covariance matrix of size  $N_{\text{pts}} \times N_{\text{pts}}$ ,  $\mathbf{k}^*$  is the vector of covariances between  $\mathbf{x}$  and all training points and  $\mathbf{y}$  is the vector of the training points.<sup>6,7</sup> Thus, unless an NN representation requires a large number of neurons and/or multiple layers, the evaluation of a surface by GP regression will be slower. Moreover, the difference in the evaluation costs is expected to increase with the dimensionality of the system. When it is important to minimize the number of *ab initio* points, GP regression is a better choice. Nonetheless, NN provide an easy and general way to quickly build a good PES. The



committee fits can be used to remove the random component of the error but require more data. They are recommended when *ab initio* calculations are relatively cheap.

- (ii) *Overfitting*. A major problem of black box fitting methods is overfitting.<sup>38</sup> Both an NN fit and GP regression become more accurate when trained with more points. When fitting a PES with a NN, it is important to avoid overfitting. As the number or distribution of training points change, one needs to reevaluate the choice of the optimal number of neurons, and the layer structure of the NN if one needs to use a multi-layer NN. Overfitting is generally not a concern in GP regression. Once the optimal kernel function is chosen, the GP model of a PES can be trained with the same choice of the correlation function for any number and distribution of training points. The GP regression with the same kernel function is guaranteed to produce more accurate results when trained by more *ab initio* points. The comparative study performed here shows that GP is able to produce good potentials when using fewer data.
- (iii) *Ease of fitting*. When using GP regression, the input requires the choice of a simple mathematical function to represent covariances. For a fixed number of points, the errors of the ensuing PES may change several-fold for different choices of the covariance function. In some cases,  $\mathbf{K}$  may be ill-conditioned, as is the case for rapidly varying data (see, for example, Ref. 9, where we used GP regression to describe resonant scattering of molecules). This should not happen for PESs, which are generally smooth functions of coordinates. It is necessary to ensure that the *ab initio* points in the training set are sufficiently far apart in order to avoid singularities in  $\mathbf{K}^{-1}$ . Proximity or overlap of points is not a problem for NN fits. In contrast, with NN, one must choose an appropriate number of neurons that achieves a good fit without overfitting.

We have observed that levels computed on PESs obtained from only 625 symmetry unique points have  $\text{sub-cm}^{-1}$  errors, despite the fact that the fitting RMSE is on the order of  $5 \text{ cm}^{-1}$ . The total number of points required when computing vibrational levels can be reduced by not using a direct product grid which ineluctably has points where the PES fitting error is large. Vibrational calculations with variational, quadrature based, methods, typically use large numbers of points distributed in areas of the configuration space where there are no or few fit or test errors; this means that the effective PES error may not be reflected in the “posted” RMSE. It is useful to use fitting points and evaluation points chosen from the same distribution, as is done here.

We emphasize that the conclusions reached here are for the part of the PES that determines the bound states. More analysis of the relative performance of GP regression and NN fits is needed for reactive systems, where the relevant configuration space is larger, and for more complex molecular systems with more degrees of freedom. We also note that the potential energy sampling scheme used in this study was not optimized for a specific PES or observable, and one can expect better performance of both methods with an optimized sampling of *ab initio* points. This is confirmed by our upcoming study,<sup>39</sup> where we show that accurate reaction dynamics results can be obtained with GP models of PES obtained with only 37 *ab initio* points for a 3D reaction system and 290 *ab initio* points for a 6D reaction system..

### Acknowledgements

This work was supported by the Ministry of Education of Singapore and the Natural Sciences and Engineering Research Council of Canada. We thank Xiao-Gang Wang for providing the reference spectrum and Daniel Koch for technical assistance.

### Supplementary material

Tables of vibrational spectra computed on the reference PES and with NN and GP fits using different numbers of points, as well as plots of absolute errors on transition frequencies on NN and GP fitted PES vs. the reference PES.

### References

- <sup>1</sup> J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley, A. J. C. Varandas, *Molecular Potential Energy Functions*, Wiley, Chichester, England, 1984.
- <sup>2</sup> T. Hollebeek, T. -S. Ho, H. Rabitz, *Annu. Rev. Phys. Chem.* **50**, 537 (1999).
- <sup>3</sup> B. J. Braams, J. M. Bowman, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- <sup>4</sup> M. A. Collins, *Theor. Chem. Acc.* **108**, 313 (2002).
- <sup>5</sup> C. M. Handley, P. L. A. Popelier, *J. Phys. Chem. A.* **114**, 3371 (2010).
- <sup>6</sup> B. Kolb, P. Marshall, B. Zhao, B. Jiang, and H. Guo, *J. Phys. Chem. A.* **121**, 2552 (2017).
- <sup>7</sup> J. Cui and R. V. Krems, *J. Phys. B: At. Mol. Opt. Phys.* **49**, 224001 (2016).

- <sup>8</sup>J. Cui, Z. Li, and R. V. Krems, J. Chem. Phys. **143**, 154101 (2015).
- <sup>9</sup>J. Cui and R. V. Krems, Phys. Rev. Lett. **115**, 073202 (2015).
- <sup>10</sup>J. Alborzpour, D. P. Tew and S. Habershon, J. Chem. Phys. **145**, 174112 (2016).
- <sup>11</sup>K. Hornik, M. Stinchcombe, H. White, Neural Netw. **2**, 359 (1989).
- <sup>12</sup>K. Hornik, Neural Netw. **4**, 251 (1991).
- <sup>13</sup>S. Manzhos, T. Carrington Jr, J. Chem. Phys. **125**, 194105 (2006).
- <sup>14</sup>E. Pradesh and A. Brown, Phys. Chem. Chem. Phys. (2017), DOI: 10.1039/C7CP04010E
- <sup>15</sup>M. H. Beck, A. Jäckle, G. A. Worth, H.-D. Meyer Phys. Rep. **324**, 1 (2000)
- <sup>16</sup>A. Leclerc and T. Carrington Jr. J. Chem. Phys. **140**, 174111 (2014).
- <sup>17</sup>S. Manzhos, K. Nakai, K. Yamashita, Chem. Phys. Lett. **493**, 229 (2010).
- <sup>18</sup>S. Manzhos, K. Yamashita, T. Carrington, Comput. Phys. Comm., **180**, 2002 (2009).
- <sup>19</sup>S. Manzhos, T. Carrington, J. Chem. Phys. **129**, 224104 (2008).
- <sup>20</sup>S. Manzhos, T. Carrington, J. Chem. Phys. **127**, 014103 (2007).
- <sup>21</sup>S. Manzhos, K. Yamashita, Surf. Sci. **604**, 555 (2010).
- <sup>22</sup>G. E. Hinton, R. R. Salakhutdinov, Science **313**, 504 (2006).
- <sup>23</sup>S. Manzhos, K. Yamashita, T. Carrington: Model Reduction and Data Analysis, series Lecture Notes in Computational Science and Engineering, vol. **75**, pp. 133-149, Springer Verlag, (2011).
- <sup>24</sup>A. N. Gorban, Appl. Math. Lett. **11**, 45 (1998).
- <sup>25</sup>S. Manzhos, R. Dawes, T. Carrington, Int. J. Quantum Chem. **115**, 1012 (2015).
- <sup>26</sup>S. Manzhos, X. Wang, R. Dawes and T. Carrington Jr. J. Phys. Chem. A **110**, 5295 (2006).
- <sup>27</sup>M. Majumder, S. E. Hegger, R. Dawes, S. Manzhos, X.-G. Wang, T. Carrington Jr., J. Li, H. Guo, Mol. Phys. **113**, 1823 (2015).
- <sup>28</sup>T. Carrington, J. Chem. Phys. **146**, 120902 (2017).
- <sup>29</sup>J. M. Bowman, S. Carter, and X. Huang, Int. Rev. Phys. Chem. **22**, 533 (2003).
- <sup>30</sup>G. Avila and T. Carrington Jr, J. Chem. Phys. **134**, 054126 (2011).
- <sup>31</sup>X. Wang and T. Carrington Jr, J. Chem. Phys. **138**, 104106 (2013).
- <sup>32</sup>NITROGEN, Numerical and Iterative Techniques for Rovibronic Energies with General Internal Coordinates, a program by P. B. Changala, <http://www.colorado.edu/nitrogen>
- <sup>33</sup>C. M. Handley, G. I. Hawe, D. B. Kell, P. L. A. Popelier, Phys. Chem. Chem. Phys. **11**, 6365 (2009).
- <sup>34</sup>A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

- <sup>35</sup> S. Manzhos, T. Carrington, J. Chem. Phys. **145**, 224110 (2016).
- <sup>36</sup> S. Carter, N. C. Handy, J. Demaison, Mol. Phys. **90**, 729 (1997).
- <sup>37</sup> I. M. Sobol, USSR Comput. Math. Math. Phys. **7**, 86 (1967).
- <sup>38</sup> N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, J. Mach Lear. Res. **15**, 1929 (2014).
- <sup>39</sup> R. A. Vargas-Hernandez, J. Guan, D. H. Zhang and R. V. Krems, to be published

ACCEPTED MANUSCRIPT

