Physics-Guided Curve Fitting for Potential-Energy Functions of Diatomic Molecules

Karl K. Irikura

Chemical Sciences Division, National Institute of Standards and Technology,

Gaithersburg, MD 20899 USA

ABSTRACT

When computing the potential-energy curve of a diatomic molecule for predictive spectroscopy, high-level calculations are usually desired. The best calculations are expensive, so few points are usually available. The points are fitted to a continuous function, such as a polynomial. Ro-vibrational energy levels are then computed using the fitted function, and spectroscopic constants extracted. However, there may be problems with overfitting, with inadequate flexibility of the fitting function, or with dependence of results upon the choice of fitting function. More fundamentally, the fitting function is selected using aesthetics or convenience, instead of physics. Here we suggest using a lower-level, high-resolution ab initio potential as a guide. Instead of fitting the sparse, high-level data directly, the energy differences between the high-level points and the guiding potential are fitted. The results are improved even with an inexpensive guiding potential. This simple strategy involves little additional effort and can be recommended for routine use. It is similar to some interpolation strategies in the literature of polyatomic molecules. When the guiding potential extends beyond the high-level data, extrapolations are also improved.

INTRODUCTION

Polyatomic potential-energy functions are in demand for use in molecular dynamics and related applications. The high dimensionality ($D \geq 3$) requires that ab initio points be sparsely placed, making interpolation an important topic. The high dimensionality also makes analytical derivatives cost-effective, because each gradient computation provides D times as much information as one energy calculation. Fitting and interpolating multidimensional potential energy surfaces is an area of active study; only a few papers are cited here.[1-8] The strategy by Fu et al.[9] is most closely related to the present suggestion. The present methods were developed independently.

A calculation on a diatomic molecule requires less computer time than a calculation on a polyatomic molecule. Its potential is only one-dimensional, so fewer calculations are needed to fill a grid. Unlike polyatomics, energy gradients are not needed because the application is usually eigenstate computation, not molecular dynamics. For these reasons, one might suppose that typical practice is to compute a large number of points on a dense grid. However, there is a contrary desire to use the best possible theoretical method, which in practice is defined as a method for which only a few points are computationally affordable. Thus, interpolation requires attention even for diatomics. Interpolation error may dominate the errors in vibrational eigenvalues.[10] The choice of grid points may also require study for each problem.[11]

STRATEGY

A continuous function is generally needed to convert discrete data into energy eigenvalues. Typical practice is to approximate the continuous function $E(R)$ as

$$E(R)=f_M\left(R;\{(R_i,E_i)\}\right), \quad (1)$$

where $f_M(R)$ is a continuous model function, such as a spline or a polynomial. The model function is fitted to the ab initio data, which is a small set of pairs of distances and energies, (R_i, E_i). Polynomials remain a popular choice of model function.[12] Sometimes the independent variable is transformed to obtain desired behavior. For example, a polynomial in $(1/R)$ correctly approaches a constant value as $R \to \infty$, while an ordinary polynomial diverges. A polynomial in $\exp(-R)$ behaves correctly as $R \to \infty$ while remaining finite at $R=0$ (the united-atom limit).

In the present strategy, we include a high-resolution potential energy function computed using an inexpensive, lower-level theoretical method. This provides the physics used to guide the interpolation of the high-level data. The low-level data are converted to continuous form, $g(R)$, using spline interpolation. The low-level grid should be dense, so that interpolations are short and the method of interpolation does not matter. We replace eq. (1) by

$$E(R)=f_M\left(R;\{(R_i,E_i)\}\right)+g(R), \quad (2)$$

where $g(R)$ is the continuous "guiding potential" derived from the low-level data. This reduces to eq. (1) in the case $g(R) = 0$. As in eq. (1), the parameters of $f_M(R)$ are fitted to the high-level data. However, they are only required to reproduce the *difference* in energy between the high-level and low-level data, which usually displays weaker structure and smaller magnitude than the high-level data alone.

Note that the result of eq. (2) is not a "dual-level" potential energy curve because the low-level method does not contribute directly to the energy. $E(R)$ is the high-level potential with interpolation (and possibly extrapolation). The low-level guiding potential, $g(R)$, although lacking adjustable parameters, contributes in the same way as the choice of model function, $f_M(R)$.

COMPUTATIONAL DETAILS

Vibrational eigenvalues are computed using the Fourier-grid Hamiltonian (FGH) method[13] with 501 grid points. FGH has periodic boundary conditions, so a padding interval, equal in width to the data set, is added to the left and right sides of the potential to suppress periodic artifacts. Each padding interval is filled with a constant (flat) potential defined by its nearest computed value. To test for periodicity trouble, the unsigned sum of wavefunction amplitudes at the left and right edges is divided by the maximum amplitude. If that ratio exceeds 10^{-6}, the padding interval is increased. The high-resolution, low-level potential is interpolated using cubic Akima splines to yield the function $g(R)$. Spectroscopic constants are determined using the minimum number of energy levels.

Sometimes a fitted potential has unphysical downward turns. In such cases, before computing energy levels, the fitted potential is truncated to retain only the correct minimum and its basin. Full isotopic masses (carrying all electrons) are used in computing energy levels.

For F_2, the high-resolution grid has 221 points from R = (1.1 to 2.2) Å, a spacing of 0.005 Å. The coarse, irregular grid consists of the nine points with R = 1.28, 1.30, 1.32, 1.36, 1.42, 1.48, 1.54, 1.58 and 1.63 Å, which are close to the classical turning points (on the high-level potential) for v = 0, 1, 2, and 3. Low-level methods are HF/6-31G*, B3LYP/6-31G*, fcMP2/cc-pVDZ and fcCCSD/cc-pVDZ, all spin-restricted, where the prefix "fc" indicates that core electrons are left uncorrelated ("frozen"). The high-level method is RHF-CCSDT/aug-cc-pwCVTZ with all electrons correlated. We are testing procedures, so a high-resolution grid is also computed at the high level to provide access to "correct" values. All calculations are done using Gaussian09[14,15] except for CCSDT, which is done using the CFOUR programs.[16,17]

For CO, the high-resolution grid has 201 points from R = (0.8 to 1.8) Å, a spacing of 0.005 Å. Guiding potentials are computed at: CASPT3(10,8)/aug-cc-pwCVTZ (third-order multireference perturbation theory[18]), icMRCI(10,8)+Q /aug-cc-pwCV5Z (internally contracted multireference singles and doubles configuration interaction[19] with Davidson correction for additional correlation), and UHF-fcCCSD/aug-cc-pVQZ. The CCSD calculations are done using Gaussian09 and the multireference calculations are done using MOLPRO.[15,19,20]

For C_2, the high-level data from Boschen et al.[21] span a wide range, from 0.9 to 20 Å. Here, only distances up to 6 Å are considered. The high-resolution grid is from 0.9 to 6.0 Å in steps of 0.005 Å (1021 points). Inexpensive guiding potentials are CASSCF(8,8)/cc-pVTZ and icMRCI(8,8)+Q/aug-cc-pwCVTZ using three-state-averaged orbitals. The X state is identified as the lower of the $\Lambda = 0$ states in the CASSCF, and as the icMRCI state dominated by that CASSCF reference. Abrams and Sherrill obtained a correctly shaped potential from RHF-fcFCI (full configuration interaction).[22] However, this is costly (about 100 times more than the MRCI and 10^4 times more than the CASSCF), so for this method we use a superposition of three grids: from 0.9 to 1.9 (ΔR = 0.005 Å), from 1.9 to 3.2 (ΔR = 0.02 Å), and from 3.2 to 6.0 (ΔR = 0.1 Å) for a total of 294 points. We solve for the three lowest states (of irrep A_g in D_{2h}), to get beyond the curve crossing near 1.7 Å while obtaining values for X $^{1}\Sigma_{g}^{+¿}$¿. The X state in the FCI calculations is identified by requiring smoothness in the energy and in the components of the quadrupole moment, although it is not always clear. These calculations, including the FCI,[23] are done using MOLPRO. Data analysis, including splining and polynomial fitting, is done in Python using the standard Scipy library.

RESULTS AND DISCUSSION

***Testing on F_2*.** Difluorine has only a formal single bond, but is an anomalous molecule with strong dynamical electron correlation.[24,25] Thus, the qualitative difference between a low-level and a high-level method may be atypically large and challenging for our procedure. To determine the distance resolution needed to exceed wavenumber precision, a grid of 2201 points (from R = 1.1 to 2.1 Å; resolution = 0.0005 Å) is computed at the fcCCSD/cc-pVDZ level. The ground vibrational level, E_0, is computed along with the first four intervals: E_0^1, E_1^2, E_2^3 and E_3^4. The numerical values are affected by the choice of spline interpolation method. The difference between cubic and linear splines is computed for each of the energy quantities as a function of grid resolution. The value of E_0 is the most sensitive to the grid resolution, by far. A grid resolution of 0.0075 Å is needed to converge E_0 to a precision of 1 cm^{-1}. A resolution of 0.035 Å is adequate for the four energy intervals. Throughout the present report, a resolution of 0.005 Å is used for "high-resolution" grids.

From a high-resolution, high-level dataset we obtain reference, "correct" values for the vibrational energy levels (using the Akima spline). These "correct" values are E_0 = 457.57 cm^{-1} and the successive vibrational intervals E_0^1 = 898.08 cm^{-1}, E_1^2 = 874.22 cm^{-1}, E_2^3 = 849.85 cm^{-1}, and E_3^4 = 824.93 cm^{-1}. The "correct" values do not represent experimental data except to the extent that the high-level method represents reality. However, they are the appropriate benchmarks for testing the fitting procedures. Note that the v = 4 level depends upon parts of the potential beyond the sparse high-level data (see Computational Details), and will therefore reflect the accuracy of extrapolation.

Figure 1 illustrates the difference in structure between the high-level, sparse data and their differences from each of the four guiding potentials. The task of the model function, $f_M(R)$, is to fit the points in each plot. For some of the guiding potentials, the range of the energy differences is larger than that of the high-level data themselves. However, the differences are more nearly linear and might be more accurately fitted by the model function, $f_M(R)$.

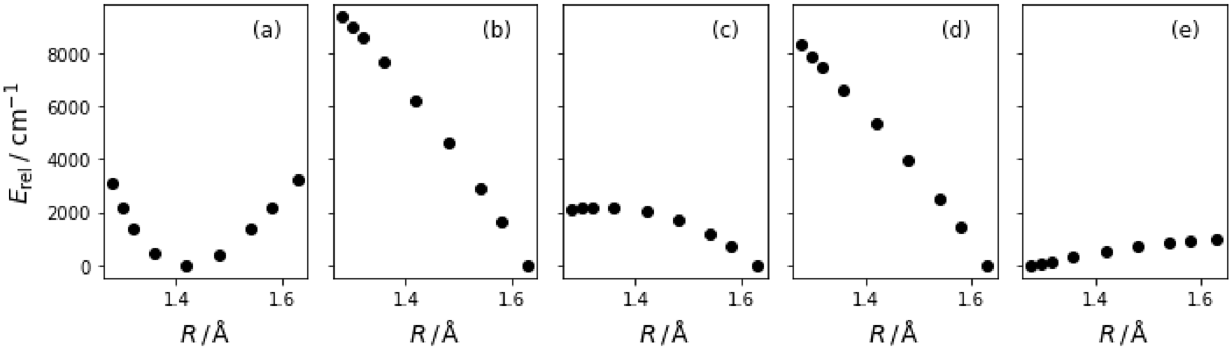

Figure 1. (a) Sparse high-level (RHF-CCSDT/aug-cc-pwCVTZ) energies of F_2; their differences from the guiding potentials: (b) HF/6-31G*, (c) B3LYP/6-31G*, (d) fcMP2/cc-pVDZ, (e) fcCCSD/cc-pVDZ. The plots have the same vertical scale to facilitate comparison.

To explore the helpfulness of a guiding potential, we start with an unusually simple model function, a linear polynomial of degree 1. Fitting it to the points in Fig. 1a, as described by eq. (1), or by eq. (2), using the four guiding potentials, results in the potential curves shown in Figure 2. The curves obtained via eq. (2) are qualitatively correct, while that from eq. (1) is of course useless. The quantitative performance of these curves is shown in Table 1. Performance improves along with the quality of the guiding potential. The best guiding potential used here, fcCCSD/cc-pVDZ, reproduces the vibrational fundamental within 1.1 %. The level $v = 4$ relies upon parts of the potential that must be extrapolated beyond the sparse, high-level data used in the fitting. Despite extrapolation, the errors in the 4-3 interval are no worse than expected from the errors in the lower intervals.

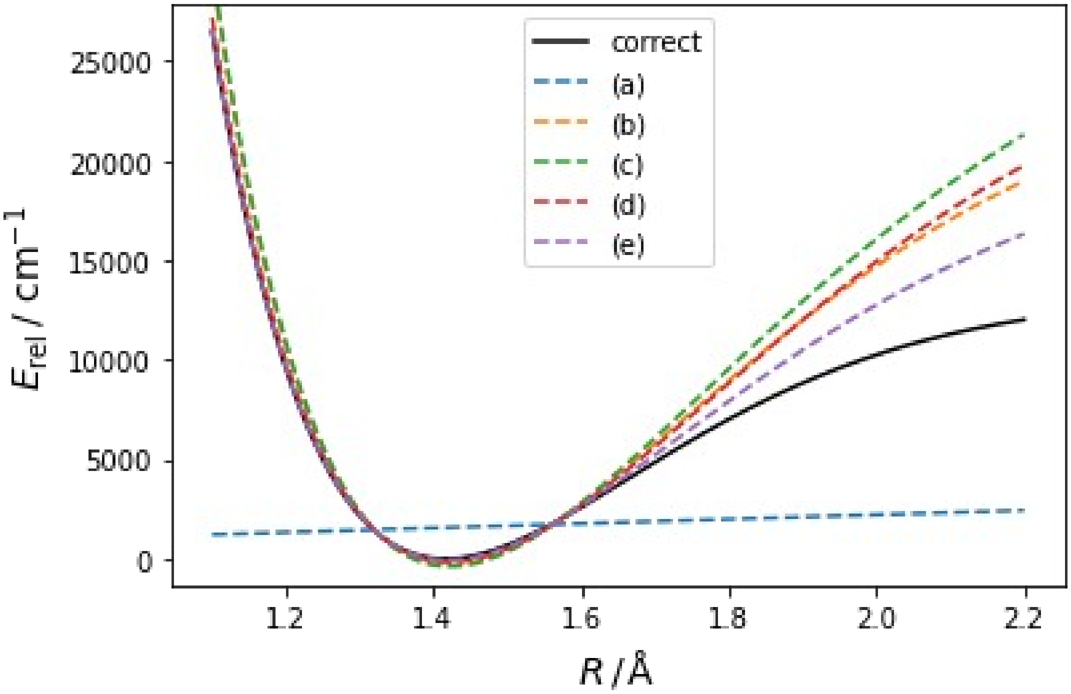

Figure 2. Potential energy curves of F_2 fitted using a linear function as f_M(). The solid black line is the "correct" high-level curve. (a) Direct fit using eq. (1); guided fitting using eq. (2) with guiding potentials $g(R)$: (b) HF/6-31G*, (c) B3LYP/6-31G*, (d) fcMP2/cc-pVDZ, (e) fcCCSD/cc-pVDZ.

Table 1. Errors in vibrational quantities from fitted curves, relative to the "correct" values.[a]

$f_M(R)$	$g(R)$	E_0 / cm^{-1}	E_0^1 /cm^{-1}	E_1^2 /cm^{-1}	E_2^3 /cm^{-1}	E_3^4 /cm^{-1}
"correct" values		*457.57*	*898.08*	*874.22*	*849.85*	*824.93*
linear	HF/6-31G*	34.23	73.59	81.45	89.67	98.33
linear	B3LYP/6-31G*	42.87	92.72	103.22	113.75	124.45
linear	fcMP2/cc-pVDZ	14.06	35.07	46.01	57.42	69.34
linear	fcCCSD/cc-pVDZ	2.81	9.82	16.70	24.02	31.82
quadratic	0	-38.42	-59.78	-35.92	-11.55	13.37
quadratic[b]	HF/6-31G*	2.27	3.65	1.81	-0.80	-4.35
quadratic[b]	B3LYP/6-31G*	1.25	2.45	1.64	-0.60	-4.52

quadratic	fcMP2/cc-pVDZ	-6.73	-9.86	-4.04	1.92	8.01
quadratic	fcCCSD/cc-pVDZ	-5.25	-7.77	-3.24	1.57	6.65
cubic[b]	0	10.46	19.37	10.00	-6.67	-37.50
cubic[b]	HF/6-31G*	-0.39	-0.44	-0.02	0.14	0.04
cubic[b]	B3LYP/6-31G*	-0.21	0.20	0.62	-0.08	-2.05
cubic[b]	fcMP2/cc-pVDZ	0.11	0.65	0.70	-0.23	-2.46
cubic[b]	fcCCSD/cc-pVDZ	0.01	0.31	0.39	-0.11	-1.41
quartic	0	1.62	1.57	-0.45	-0.39	2.65
quartic[b]	HF/6-31G*	-0.30	-0.27	0.08	0.09	-0.29
quartic	B3LYP/6-31G*	-0.57	-0.55	0.18	0.16	-0.58
quartic[b]	fcMP2/cc-pVDZ	-0.37	-0.35	0.10	0.09	-0.50
quartic[b]	fcCCSD/cc-pVDZ	-0.25	-0.24	0.07	0.06	-0.35
quintic[b]	0	-0.07	-0.12	0.03	-0.02	-0.63
quintic	HF/6-31G*	0.05	0.09	-0.02	0.01	0.44
quintic	B3LYP/6-31G*	-0.04	-0.03	0.03	0.05	0.50
quintic	fcMP2/cc-pVDZ	0.02	0.04	-0.01	0.01	0.30
quintic	fcCCSD/cc-pVDZ	0.02	0.03	-0.01	0.01	0.20
sextic	0	-0.01	-0.01	0.00	0.01	0.13
sextic[b]	HF/6-31G*	0.01	0.01	-0.01	-0.01	-0.12
sextic	B3LYP/6-31G*	-0.04	-0.03	0.03	0.05	0.49
sextic[b]	fcMP2/cc-pVDZ	-0.00	-0.00	0.00	0.00	0.01
sextic[b]	fcCCSD/cc-pVDZ	0.00	0.00	-0.00	0.00	-0.01
degree 8	0	0.01	0.00	0.00	0.01	0.05
degree 8	HF/6-31G*	0.00	0.00	0.00	0.01	0.06
degree 8[b]	B3LYP/6-31G*	-0.00	0.00	-0.01	-0.10	-0.84
degree 8	fcMP2/cc-pVDZ	0.00	0.00	-0.00	0.00	0.00
degree 8	fcCCSD/cc-pVDZ	0.00	0.00	-0.00	0.01	0.05

[a] The highest interval requires the $v=4$ level, whose turning points lie beyond the range of the high-level data and therefore rely upon extrapolation of the fitted potential.

[b] Fitted curve truncated to eliminate turnover.

Although a linear model function will seldom be the best choice, there are at least two situations in which it could be: (1) when only two data points are available, presumably from an extremely costly method, and (2) when more than two data points are available but they contain noise that must be mitigated by an averaging process such as least-squares fitting. Situation #2 is especially relevant for Monte-Carlo methods of electronic structure. Note that the energy minimum, E_e, and its location, R_e, can also be determined from only two points by using eq. (2).

Next we try a quadratic polynomial of degree 2 for $f_M()$. The resulting fitted curves are shown in the Supporting Information. Compared with Figure 1, the agreement with the reference potentials is improved. The quantitative performance, shown in Table 1, is markedly better than for the linear model function. Errors for the extrapolated interval (E_3^4) are different from expected by extrapolating the errors in the lower intervals. Surprisingly, the best performance is no longer from the best guiding potentials. Of course, eq. (1) yields only a harmonic potential with this model function.

A quadratic function is typically used when only three points are available and a harmonic result is acceptable. In that case, one chooses the three points of lowest energy, to estimate the curvature at the energy minimum. If we do that here, the results from eq. (1) are improved. For example, the error in the fundamental transition is reduced to 16.50 cm^{-1}. However, that still exceeds the errors from using eq. (2).

A cubic polynomial of order 3 is the simplest polynomial that can provide anharmonic results via eq. (1). The resulting fitted curves are shown in the Supporting Information. The cubic function plunges downward as the bond length increases. Thus, when used without a guiding potential it is meaningful only near the minimum, as is well known.[26] Including the entire range of Fig. 3 leads to nonsensical results from eq. (1). Instead, the cubic values in Table 1 for $g(R) = 0$ are obtained by truncating the fitted potential at its turnover point, i.e., where it begins its downward plunge. Similar behavior, but milder, is seen in the curves obtained using eq. (2). The errors from guided fitting are smaller with the cubic model function than with the quadratic.

A quartic polynomial (order 4) does not necessarily plunge downward and is the lowest-order polynomial that is generally useful for obtaining anharmonic energy levels via eq. (1). The corresponding fitted curves are plotted in the Supporting Information and the quantitative errors are listed in Table 1. For the guided fitting, results are noticeably better (than with the cubic function) only for the extrapolated interval. The results are markedly improved for the direct fitting of eq. (1), although it remains inferior to guided fitting.

A quintic polynomial (order 5) also displays a severe downturn in the curve from eq. (1) (see Supporting Information). After the usual truncation, the resulting energy errors are in Table 1. The errors are now quite small for all fitted curves. The results from guided fitting are still better than from eq. (1), but only marginally.

A sextic polynomial (order 6) is often used for fitting a discrete potential that includes a modest number of points. Quantitative results are listed in Table 1 and the curves are shown in the Supporting Information. The errors are even smaller than with the quintic model function. The direct fit of eq. (1) is marginally better than the fit guided by B3LYP, and marginally worse than fits guided by MP2 or CCSD.

The high-resolution data set has nine points, so the highest-order polynomial that can be fitted is of order 8. Since this makes full use of the data, it could be argued that it does not “waste” any data, although there is no room for noise and there is a risk of overfitting. The fitted curves are shown in the Supporting Information. The results are comparable to those from the sextic polynomial. Overall, it appears mildly advantageous, especially for the extrapolated interval, to choose a model function that does not lead to a strong downturn in the fitted potential.

Application to CO. Powell and Dawes have published sparse potential energy curves for carbon monoxide computed using a quantum Monte Carlo (QMC) method.[27] To obtain spectroscopic constants, they fitted selected points from their data to a Morse model potential. The points were selected to yield spectroscopic constants close to the experimental values. The authors had difficulty with the curve-fitting, noting that the Morse potential cannot fit the data over their full range. Also note that a Morse potential truncates all anharmonicity at the first constant ($\omega_e x_e$).[26] Here, we restrict our attention to the most expensive calculations, Table 1 in ref. [27], and to the four QMC points that are relevant for the lower vibrational levels. These points lie at R/Å = 0.9, 1.0, 1.1 and 1.3. (The next point is at 2.6 Å.) 0.9 Å is close to the inner classical turning point for v = 24, for which the outer turning point is near 1.655 Å.[28]

Figure 3 shows the sparse QMC data and their differences with the guiding potentials. Unlike Fig. 1, the vertical scales are different within Fig. 3. (Otherwise, the structure within Fig. 3bcd would not be visible.) The differences between QMC and the multireference methods are nearly linear.

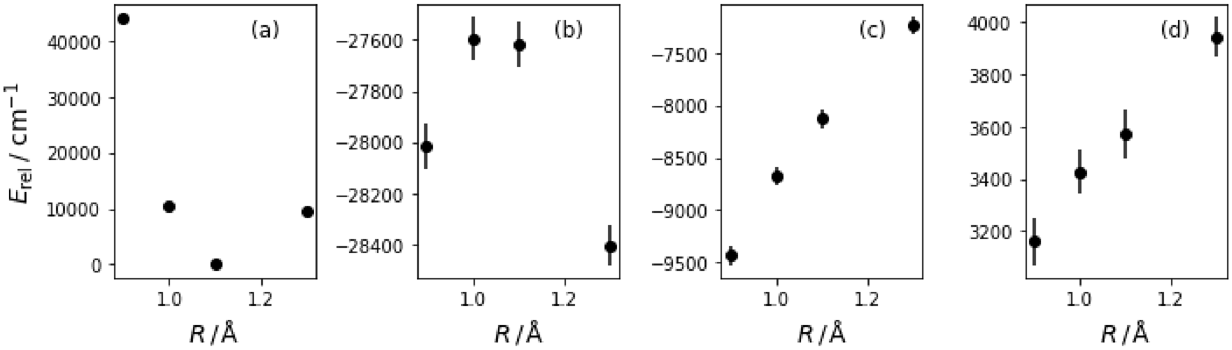

Figure 3. (a) Sparse high-level (QMC) energies[27] of CO; their differences from the guiding potentials: (b) UHF-fcCCSD/aug-cc-pwCVQZ, (c) CASPT3/aug-cc-pwCVTZ, (d) icMRCI(10,8)+Q/aug-cc-pwCV5Z. Vertical scales differ.

The uncertainties for the data points are not all the same, so the model function must be fitted using a weighted procedure. We choose each weight to be the inverse square of its stated uncertainty. Starting with $f_M()$ as a linear function, we obtain spectroscopic constants as shown in Table 2. (Potential energy curves are shown in the Supporting Information.) Experimental values are shown in the first row of Table 2, but we can only expect to match them, at best, to the extent that the QMC data are accurate.

Table 2. Spectroscopic constants (cm^{-1}) for CO from guided fitting of QMC data from Powell and Dawes.[27] Experimental values are in the first row.[29]

$f_M(R)$	$g(R)$	ω_e	$\omega_e x_e$	$\omega_e y_e$	B_e	α_e	D_e
experimental values		*2169.8*	*13.29*	*0.0104*	*1.932*	*0.0175*	*6.12e-6*

linear	fc-uCCSD	2218.9	12.15	0.0212	1.938	0.0164	5.92e-6
linear	CASPT3	2197.2	12.77	0.0089	1.937	0.0169	6.03e-6
linear	icMRCI+Q	2185.9	13.17	0.0114	1.937	0.0173	6.09e-6
quadratic	0	2665.9	0	0	1.836	-0.0076	3.48e-6
quadratic	fc-uCCSD	2182.0	13.38	0.0077	1.938	0.0176	6.11e-6
quadratic	CASPT3	2183.3	13.24	0.0030	1.937	0.0174	6.10e-6
quadratic	icMRCI+Q	2183.3	13.26	0.0104	1.937	0.0174	6.10e-6
cubic [a]	0	2401.0	27.48	-4.64	1.975	0.0191	5.37e-6
cubic	fc-uCCSD	2181.6	12.09	0.0385	1.935	0.0168	6.09e-6
cubic	CASPT3	2183.0	12.28	0.0267	1.935	0.0168	6.08e-6
cubic	icMRCI+Q	2183.0	12.33	0.0327	1.935	0.0169	6.08e-6

[a] Truncated at the turnover point (near 1.35 Å).

In this case, we do not know the "correct" results because we do not have a high-resolution grid of QMC data. Although it is only a single example, from the F_2 test we expect that higher-order polynomials will give more accurate results. (This expectation is supported by the increasing coincidence among the alternatively guided potentials as the polynomial degree is increased—see Supporting Information.) Greater precision is associated with a smaller variation of results from different guiding potentials. Considering the values of ω_e, the spreads are 33 cm^{-1} from a linear model function, 1.3 cm^{-1} from a quadratic, and 1.4 cm^{-1} from a cubic. For $\omega_e x_e$, the spreads are 1.02, 0.14 and 0.24 cm^{-1} from linear, quadratic and cubic, respectively. The results appear reasonably converged with the quadratic model function. Since the highest-level guiding potential is from icMRCI+Q, and it shows small and near-linear differences (Fig. 3d), we might favor the results from that guiding potential and the quadratic model function. Note that a cubic function fits the four data points exactly, which means that the statistical noise in the data will not be canceled at all (by least-squares averaging).

In diffusion QMC calculations, the uncertainty of the energy decreases as $T^{-1/2}$, where T is the computer time. For example, spending 4 times as much computer time will decrease the energy uncertainties by half. However, it may be better, instead, to have 4 times as many data points.

The uncertainties associated with the target quantities, such as spectroscopic constants, are important but are seldom reported in theoretical work. The uncertainties should reflect variations in guiding potential, in model function, in the selection of data points,[10] and uncertainty in the data themselves. Note that noise, or uncertainty, in energy calculations is not restricted to stochastic methods. Basis-set extrapolation carries uncertainty as well. Quantitative uncertainty analysis is beyond the scope of this report. However, straightforward Monte Carlo propagation (2000 samples) of the reported[27] energy uncertainties, using the icMRCI+Q guiding potential and quadratic model function, yields standard deviations (in cm^{-1}) of 6.6, 0.21, 0.0026, 0.001, 0.0002 and 0.03e-6 for the constants ω_e, $\omega_e x_e$, $\omega_e y_e$, B_e, α_e and D_e, respectively.

We should note that the Morse model function, although popular for a long time, is not derived from physical laws.[30] Many alternatives have been invented;[31,32] one's choice is essentially arbitrary. In guided fitting, although the choice of theoretical guiding potential is partly arbitrary, it can be improved systematically. Moreover, the guiding potential is based upon the electronic structure of the specific molecule under study. That is, it is based upon physics rather than draftsmanship (French curves and flat splines). The draftsmanship part is the residual model function, $f_M(R)$.

Application to C_2. Dicarbon is well-known for its challenging electronic structure.[21,22] Here, we focus on the spectroscopic constants as determined by fitting the high-level CEEIS (correlation energy extrapolation by intrinsic scaling) data by Boschen et al.[21] The authors reported difficulty fitting the data to a continuous form, which they ascribed to an avoided crossing near 1.7 Å. In the end they used a cubic spline. Here we re-analyze their 43 data points (up to R = 6 Å) using guided fitting. Abrams and Sherrill[22] reported correctly-shaped curves at the fcFCI/6-31G* level, so we use it here as a guiding potential. We also use guiding potentials from CASSCF/cc-pVTZ and icMRCI+Q/aug-cc-pwCVTZ (active core), which are far less expensive than the FCI.

First, we note that direct analysis of the high-level data from Boschen et al., using eq. (1), gives results that depend upon the method of interpolation used to smooth the potential. Table 3 lists the first six vibrational intervals and the ground (zero-point) energy relative to the (interpolated) minimum. The variability with splining method, an arbitrary choice, is shown in the last column. The high variability indicates that the data are still too sparse to define the vibrational levels to wavenumber precision. Alternatively, a physically derived guiding function may reduce the demands placed upon the model function, thus reducing the variability (i.e., uncertainty) from the arbitrary choice of splining method.

Table 3. Vibrational intervals (cm^{-1}) of C_2 as computed from high-level data by Boschen et al.[21] using different splining methods.

interval	Boschen[a]	Akima	linear	quadratic	cubic	variab[b]
ground state	919.5	940.8	1024.3	919.5	919.4	104.9
1-0	1830.4	1816.3	1845.8	1830.4	1830.5	29.5
2-1	1806.1	1804.2	1804.4	1806.3	1806.3	2.1

3-2	1774.5	1781.3	1769.2	1774.7	1774.8	12.1
4-3	1741.5	1738.2	1742.8	1742.3	1742.3	4.6
5-4	1710.1	1710.6	1707.4	1711.7	1711.4	4.2
6-5	1676.8	1683.0	1687.2	1678.2	1678.6	9.0

[a] Cubic-spline results from ref. [21]. [b] Range of values across the four splining methods applied here.

Figure 4 shows the differences between the (relatively) sparse CEEIS data and three choices of guiding potential. As in Fig. 3, the vertical scales are different to show the structure in the plots. The structure is more complicated than in the previous examples, suggesting that fitting will be more difficult. However, Boschen's data are more dense than usual in such work, making splines a reasonable approach.

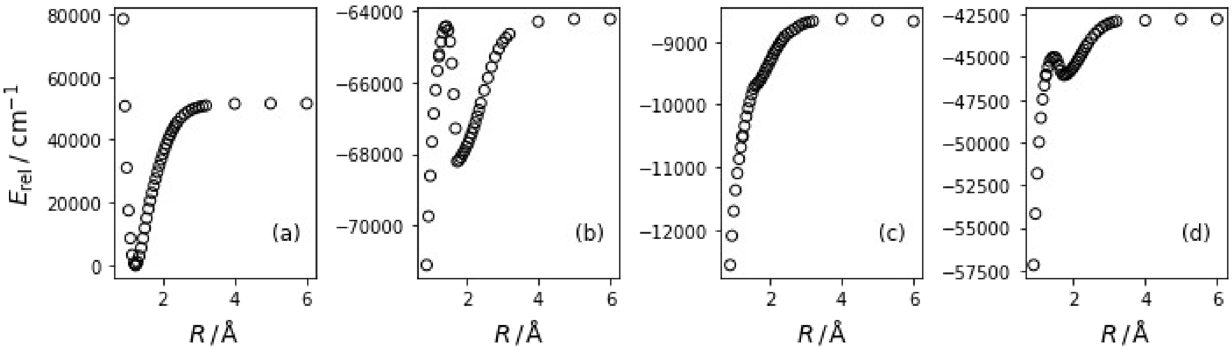

Figure 4. (a) High-level (CEEIS) energies[21] of C_2; their differences from the guiding potentials: (b) CASSCF(8,8)/cc-pVTZ, (c) icMRCI(8,8)+Q/aug-cc-pwCVTZ, (d) fcFCI/6-31G*. Vertical scales differ.

For each of the three guiding potentials as $g(R)$, we apply the four splining methods used in Table 3 as $f_M(R)$. All results are listed in the Supporting Information. The variabilities and the results from the Akima spline are shown in Table 4. The most consistent results, with little dependence upon the choice of splining method, are obtained when guided by the icMRCI+Q potential. However, all the guiding potentials lead to marked improvement over the unguided results in Table 3, and the results from the different guiding potentials (Table 4) are in mutual agreement.

Table 4. Vibrational levels (cm^{-1}) of C_2, and their standard deviations across splining methods, as obtained by guided interpolation of the data by Boschen et al.[21]

interval	CASSCF[a]	MRCI+Q[b]	FCI[c]	CASSCF[a]	MRCI+Q[b]	FCI[c]
	value[d]	value[d]	value[d]	variab[e]	variab[e]	variab[e]
ground state	918.1	919.3	917.4	5.6	0.7	11.3

1-0	1831.5	1830.5	1832.1	2.0	0.0	3.3
2-1	1806.6	1806.7	1806.6	0.9	0.7	0.6
3-2	1774.5	1774.7	1774.2	1.0	0.8	0.6
4-3	1742.5	1742.2	1742.8	0.4	0.5	0.6
5-4	1711.4	1711.4	1711.4	0.3	0.1	0.2
6-5	1678.5	1678.7	1678.4	1.8	0.2	1.6

[a] $g(R)$ = CASSCF(8,8)/cc-pVTZ. [b] $g(R)$ = icMRCI(8,8)+Q/aug-cc-pwCVTZ. [c] $g(R)$ = fcFCI/6-31G*. [d] Using Akima spline. [e] Range in values across results from linear, quadratic, cubic, and Akima splines.

CONCLUSIONS

For a given set of parametric fitting functions, such as polynomials or splines of different order, computed spectroscopic quantities are more consistent, i.e., have smaller uncertainties, when a physics-based guiding function is used as a reference, instead of directly fitting high-level ab initio data. The best results appear to be obtained when the difference between the high-level data and the guiding function is a linear function of bond length. The technique may be especially advantageous for noisy data, as from stochastic electronic-structure methods, because a low-order polynomial can more easily fit the data while averaging the noise. Meaningful spectroscopic information can be obtained from as few as two data points.

DATA AVAILABILITY STATEMENT

The data that support this study are found within the article and the supplementary material.

REFERENCES

[1] M. A. Collins and J. Ischtwan, J. Chem. Phys. **93**, 4938-45, (1990).
[2] J. Ischtwan and M. A. Collins, J. Chem. Phys. **100**, 8080-8, (1994).
[3] K. A. Nguyen, I. Rossi, and D. G. Truhlar, J. Chem. Phys. **103**, 5522-30, (1995).
[4] T. Hollebeek, T. S. Ho, and H. Rabitz, Ann. Rev. Phys. Chem. **50**, 537-70, (1999).
[5] B. Strickler and M. Gruebele, Chem. Phys. Lett. **349**, 137-45, (2001).
[6] G. G. Maisuradze and D. L. Thompson, J. Phys. Chem. A **107**, 7118-24, (2003).
[7] S. Lorenz, A. Gross, and M. Scheffler, Chem. Phys. Lett. **395**, 210-5, (2004).
[8] A. Nandi, C. Qu, P. L. Houston, R. Conte, and J. M. Bowman, J. Chem. Phys. **154**, 051102, (2021).
[9] B. Fu, X. Xu, and D. H. Zhang, J. Chem. Phys. **129**, 011103, (2008).
[10] F. E. Penotti, Computers & Chemistry **21**, 363-7, (1997).
[11] R. J. Le Roy, J. Quant. Spectrosc. Rad. Transf. **186**, 167-78, (2017).
[12] S. R. Battey, D. H. Bross, K. A. Peterson, T. D. Persinger, R. A. VanGundy, and M. C. Heaven, J. Chem. Phys. **152**, 094302, (2020).
[13] C. C. Marston and G. G. Balint-Kurti, J. Chem. Phys. **91**, 3571-6, (1989).
[14] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, Gaussian 09, D.01 (Gaussian, Inc., Wallingford, CT, 2013).

[15] Certain commercial materials and equipment are identified in this paper in order to specify procedures completely. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.
[16] J. F. Stanton, J. Gauss, L. Cheng, M. E. Harding, D. A. Matthews, and P. G. Szalay, CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package, vers. 2.1 (2019); http://www.cfour.de/.
[17] D. A. Matthews, L. Cheng, M. E. Harding, F. Lipparini, S. Stopkowicz, T.-C. Jagau, P. G. Szalay, J. Gauss, and J. F. Stanton, J. Chem. Phys. **152**, 214108, (2020).
[18] H. J. Werner, Mol. Phys. **89**, 645-61, (1996).
[19] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803-14, (1988).
[20] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, WIREs Computational Molecular Science **2**, 242-53, (2012).
[21] J. S. Boschen, D. Theis, K. Ruedenberg, and T. L. Windus, Theor. Chem. Acc. **133**, 1425, (2013).
[22] M. L. Abrams and C. D. Sherrill, J. Chem. Phys. **121**, 9211-9, (2004).
[23] P. J. Knowles and N. C. Handy, Comput. Phys. Commun. **54**, 75-83, (1989).
[24] J. J. M. Wiener, J. S. Murray, M. E. Grice, and P. Politzer, Mol. Phys. **90**, 425-30, (1997).
[25] B. Csontos, B. Nagy, J. Csontos, and M. Kállay, J. Phys. Chem. A **117**, 5518-28, (2013).
[26] G. Herzberg, *Spectra of Diatomic Molecules*, 2nd (corrected) ed. (van Nostrand Reinhold, New York, 1989).
[27] A. D. Powell and R. Dawes, J. Chem. Phys. **145**, 224308, (2016).
[28] T. I. Velichko and S. N. Mikhailenko, Opt. Spectrosc. **118**, 6-10, (2015).
[29] A. Le Floch, Mol. Phys. **72**, 133-44, (1991).
[30] P. M. Morse, Phys. Rev. **34**, 57-64, (1929).
[31] R. J. Le Roy, J. Quant. Spectrosc. Rad. Transf. **186**, 179-96, (2017).
[32] R. Jaquet, Interpolation and fitting of potential energy surfaces: Concepts, recipes and applications in *Potential Energy Surfaces*, edited by A. F. Sax; Lecture Notes in Chemistry vol. 71 (Springer, Berlin, Heidelberg, 1999), pp. 97-175.