A Comparison of Medium-Sized Basis Sets for the Prediction of Geometries, Vibrational Frequencies, Infrared Intensities and Raman Activities for Water

Brian C. Ferrari
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando FL 32816
E-mail: Brian.Ferrari@ucf.edu

Chris J. Bennett
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando FL 32816
Department of Chemistry, University of Central Florida, 4111 Libra Drive, Orlando FL 32816
Florida Space Institute, 12354 Research Parkway, Partnership 1 Building, Suite 214, Orlando FL 32826
Department of Chemistry & Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, GA 30332
E-mail: Christopher.Bennett@ucf.edu

Abstract.
Optimized geometries, vibrational frequencies, as well as infrared intensities and Raman activities were calculated for water (H2O) utilizing popular quantum mechanical approaches. Here, density functional theory (DFT) calculations were performed using the B3LYP (Becke, three-parameter, Lee-Yang-Parr) functional, as well as ab initio calculations using second-order Møller-Plesset (MP2) perturbation theory and coupled-cluster with single, double and perturbative triple excitations [CCSD(T)] levels of theory were used. We assess and benchmark the performance of 69 different atomic orbital basis sets including various popular families of medium-sized basis sets typically of two to four zeta quality and differing levels of augmentation by polar and diffuse functions. The basis sets range from the commonly adopted Pople-style (6-31G & 6-311G), Dunning’s correlation consistent (cc-pV(n+d)Z & aug-cc-pV(n+d)Z, as well as Truhlar’s calendar variations, Jensen’s polarization consistent (pc-n & aug-pc-n), Ahlrichs (def2-...), Sapporo’s and Karlsruhe’s as well as atomic natural orbitals (ANO) such as NASA Ames (ANO), Nese-style, and Roco-style. We also compare several basis sets specifically designed to calculate vibrational and electronic properties, including the Sadlej-pVTZ (and LPol-X families), as well as SNS families of Barone. The results are compared to experimental values where available, or calculations performed with 5 or 6 zeta-level (e.g., cc-pV6Z). The performance of each family of basis sets is discussed in terms of their accuracy (and pitfalls), as well as computational resource scaling and efficiency. The Def2 family performs very well overall, yielding more accurate results with lower runtimes than traditional basis sets. ‘May’ basis sets also provide accurate predictions of vibrational frequencies at significantly lower costs. Raman activities can be accurately calculated using MP2 under harmonic approximation with several ‘spectroscopic’ families performing well.

1. Introduction
Predictions of molecular properties, such as geometries, vibrational frequencies, infrared intensities and Raman activities have become powerful tools for chemists and physicists performing spectroscopic analysis of substances. Particularly involving the characterization...
or quantification of unstable species, or radicals which are produced \textit{in situ} during experiments, where experimental values are often not available [1, 2]. Commonly employed quantum mechanical approaches to calculation these properties include density functional theory (DFT) and \textit{ab initio} methods, of which extensive studies have been done investigating accuracy of their predictions [3–11]. However, only a few studies exist systematically comparing the performance of a limited number of differing atomic orbital basis sets to accurately calculate these properties [12]. Moreover, careful selection from the plethora of basis set choices now easily accessible (for example through the EMSL database, https://bse.pnl.gov/bse/portal) can lead to optimization of the approach implemented based upon computational resources and required accuracy levels [12–14]. Since computational resources are often constrained, Pople-style basis sets are often employed since they are often not demanding, and widely available in commercial packages, however, they are often out-performed by other basis sets at similar computational cost. On the other hand, correlation consistent basis sets are widely available and - although computationally heavy - they do systematically converge on accurate results when either 5/6 zeta basis sets are used, or complete basis set extrapolation schemes are incorporated (where each zeta, $\zeta$, typically approximately doubles the number of atomic orbitals). Since computational costs for these calculations scale with $N^4$ to $N^7$, the runtimes for even medium-sized molecules under 5/6 zeta become unfeasible. However, these large calculations don’t necessarily produce the most accurate predictions. The results from several basis set families can be extrapolated to complete basis set (CBS) limit via a series of calculations performed at lower zeta level [15–17]. This study compares the results of several medium-sized basis sets across three levels of theory, in order to identify which basis sets can be reliably utilized to predict molecular properties (such as geometries, frequencies, IR intensities and Raman activities) maximizing the precision/accuracy for a given set of computational resources. Here, we only focus on preliminary results for water, but the findings are in line with those observed for other molecules we have investigated. Both DFT and \textit{ab initio} calculations were performed, using Becke’s three parameter exchange functional [18] along with Lee-Yang-Parr’s correlation functional [19] (B3LYP), second-order Møller-Plesset (MP2) perturbation theory [20] and coupled-cluster with single, double and perturbative triple excitations [CCSD(T)] [21, 22]. For each level of theory 69 different atomic orbital basis sets were used varying from Pople-style (6-31G & 6-311G) [23], Dunning’s correlation consistent (cc-pV(n+d)Z & aug-cc-pV(n+d)Z) [24], as well as Truhlar’s calendar variations [25–30], Jensen’s polarization consistent (pc-n & aug-pc-n) [31–33], Alrich’s (def2-...) [34], Sapporo’s and Karlsruhe’s, as well as atomic natural orbitals (ANOs) such as NASA Ames (ANOn), Neese-style [35], and Roos-style. We also compare several basis sets specifically designed for calculating vibrational and electronic properties, including the Sadlej-pVTZ (and LPol-X families) [36,37], as well as SNS & NO7 families of Barone [38,39] and a new generation of Thakkar DZ basis sets, NLO-1 [40]. Calculated values are compared to both experimental values [41–43] where available, as well as various convergences found either running large zeta calculations or through extrapolation. Though many medium-sized basis sets may converge on the solutions provided by extrapolation schemes and high-zeta results, this does not guarantee, and in fact rarely results in the prediction of molecular properties being accurately calculated due to inherent flaws in the level of theory taken. Therefore, minor improvements at the cost of significantly larger runtimes are rarely worthwhile. It is of note that this study does not account or attempt to correct for errors arising from inadequacies in the levels of theory described here, although it is possible to do so (e.g., by correcting for the truncation of the theoretical approaches, core-correlation effects, or relativistic effects which need to be taken into account to reproduce frequencies to within 1 cm$^{-1}$). However, due to fortuitous cancellation of errors, the CBS limits, or cc-pV6Z results are often within 5-10 cm$^{-1}$ of the corrected values [44]. Here, we report on which basis sets can be reliably utilized to predict molecular properties (such as geometries, frequencies, IR intensities and Raman activities) maximizing the precision/accuracy
for a given set of computational resources. The need for high zeta basis sets, or extrapolated results can often be avoided, if the basis sets are more carefully considered for the property being evaluated, and more accurate results reliably obtained at a lower computational cost. We briefly address the benefits of using anharmonic corrections to determine vibrational frequencies versus the use of scaling factors for a few select cases.

2. Methods
Calculations were done using the General Atomic and Molecular Electronic Structure System, GAMESS(US) [45–47], using the 14 FEB 2018 (R1) version of Games(US). All calculations were run using spherical harmonics; Pople-style basis sets were typically run using cartesian functionals unless they incorporated t-type polarization functions, where spherical harmonics were used. The segmented versions of polarization consistent basis sets were used, it should also be noted that within the Def2 family QZVPD is equivalent to QVZPPD and that QZVP is equivalent to QVZPP. Tight convergence criteria were used throughout (e.g., integral cutoff $10^{-12}$, primitive cutoff $10^{-25}$, $10^{-7}$ SCF convergence, gradient convergence $10^{-6}$ for geometries). Hessian and Raman calculations were run with semi-numerical (B3LYP and MP2) or fully-numerical gradients [CCSD(T)] under the double harmonic approximation with projected frequencies and two displacements in each cartesian direction during force field computations. The frozen-core approximation was used for both MP2 and CCSD(T) calculations. B3LYP calculations were run using an “army grade” pruned grid (JANS=2) with roughly 71,000 grid points and 155 radial points per atom. In general calculations were run on four processors, with only a few exceptions. Larger basis sets such cc-pV5Z, cc-pV6Z and Pcss-4 were run with fully numerical gradients for all theory methods, and required calculations to be run on a single processor. For a few special cases, the anharmonic frequencies were calculated using Vibrational self-consistent field (VSCF) [48]. These calculations were run with third order mode couplings, ten grid points, with second order degenerate perturbation theory corrections. Extrapolated complete basis set (CBS) limits were obtained using the following two-point linear extrapolation formula generated by Halkier et al [49].

$$E_{X,Y} = E_X^{\text{cosh}} \frac{X^3 - E_Y^{\text{cosh}} Y^3}{X^3 - Y^3}$$

(1)

Where X and Y are the basis set's zeta value, with $Y = X - 1$, CBS limits were found using the highest zeta basis sets calculated per basis “family”. Error analysis was done by taking the average error between the calculated values and either experimental data or CBS limits.

3. Results and Discussion

| Optimization | Experimental [42] | Bond Length (Å) | Bond Angle (°) |
|--------------|------------------|----------------|--------------|
|              |                  | B3LYP | MP2 | CCSD(T) | B3LYP | MP2 | CCSD(T) |
| 1 cc-pV6Z    | 0.9606(67)       | 0.9582(68) | 0.9582(69) | 105.120 | 104.335 | 104.422 |
| 2 cc-pV5Z    | 0.9605(63)       | 0.9580(63) | 0.9580(61) | 105.091 | 104.286 | 104.373 |
| 3 Aug-cc-pVQZ| 0.9608(55)       | 0.9584(55) | 0.9584(55) | 105.110 | 104.260 | 104.365 |
| 4 Jul-cc-pVQZ| 0.9608(53)       | 0.9584(52) | 0.9584(52) | 105.109 | 104.263 | 104.360 |
| 5 Jan-cc-pVQZ| 0.9608(50)       | 0.9584(48) | 0.9584(49) | 105.114 | 104.209 | 104.301 |
| 6 May-cc-pVQZ| 0.9608(48)       | 0.9584(45) | 0.9584(45) | 105.122 | 104.219 | 104.306 |
| 7 cc-pVQZ    | 0.9605(45)       | 0.9577(43) | 0.9579(43) | 104.877 | 104.018 | 104.116 |
| 8 Aug-cc-pVTZ| 0.9621(41)       | 0.9614(39) | 0.9616(39) | 105.082 | 104.109 | 104.180 |
| 9 Jul-cc-pVTZ| 0.9620(39)       | 0.9614(37) | 0.9616(37) | 105.100 | 104.125 | 104.188 |
| 10 Jun-cc-pVTZ| 0.9620(35)      | 0.9609(34) | 0.9611(35) | 105.110 | 104.063 | 104.118 |
| 11 May-cc-pVTZ| 0.9620(32)      | 0.9603(31) | 0.9604(31) | 105.072 | 103.893 | 103.920 |
| 12 cc-pVTZ   | 0.9616(29)       | 0.9594(27) | 0.9594(29) | 104.517 | 103.518 | 103.582 |
| 13 Aug-cc-pVDZ| 0.9651(21)      | 0.9650(20) | 0.9650(21) | 104.741 | 103.873 | 103.937 |
| 14 Jul-cc-pVDZ| 0.9650(16)      | 0.9653(14) | 0.9658(14) | 105.035 | 104.232 | 104.256 |
| 15 Jun-cc-pVDZ| 0.9686(8)       | 0.9665(8) | 0.9672(8) | 104.045 | 102.835 | 102.798 |
| 16 May-cc-pVDZ| 0.9686(10)      | 0.9665(8) | 0.9672(9) | 104.045 | 102.835 | 102.798 |
Infrared Intensities and Raman Intensities in $\AA^2$amu$^{-1}$

| Exp [s] | Infrared Intensities | Raman Intensities |
|---------|----------------------|-------------------|
| 3       | 2.94                 | 92 s              |
|         |                      |                   |
| #       | H2MP                 | MP2               |
| 3       | 4.63(57)             | 6.09(57)          |
| 4       | 5.20(52)             | 6.68(51)          |
| 5       | 1.51(55)             | 6.80(54)          |
| 6       | 5.16(49)             | 6.90(48)          |
| 7       | 3.89(45)             | 6.55(43)          |
| 8       | 4.59(41)             | 5.56(39)          |
| 9       | 5.86(36)             | 7.44(33)          |
| 10      | 9.91(39)             | 7.61(37)          |
| 11      | 6.05(34)             | 6.05(34)          |
| 12      | 3.16(29)             | 5.74(29)          |
| 13      | 4.00(19)             | 4.14(19)          |
| 14      | 1.15(8)              | 1.62(8)           |
| 15      | 6.33(14)             | 9.13(13)          |
| 16      | 11.75(9)             | 16.28(9)          |
| 17      | 2.84(14)             | 6.56(15)          |
| 18      | 4.45(37)             | 6.42(34)          |
| 19      | 2.72(26)             | 5.12(26)          |
| 20      | 7.23(22)             | 9.84(19)          |
| 21      | 2.06(17)             | 3.60(15)          |
| 22      | 9.48(11)             | 12.96(7)          |
| 23      | 3.41(17)             | 5.82(17)          |
| 24      | 6.34(5)              | 10.08(3)          |
| 25      | 4.64(60)             | 6.09(60)          |
| 26      | 4.66(40)             | 5.60(38)          |
| 27      | 5.76(28)             | 9.48(28)          |
| 28      | 7.35(2)              | 13.81(3)          |

**Table 3: Infrared Intensities in $km \cdot mol^{-1}$ and Raman Activities in $\AA^2$amu$^{-1}$**

These results were taken from Tew et al. (2007).

An error in the calculation of this vibrational frequency caused this particular result to be invalid.

Removed due to convergence issues.
XXX IUPAP Conference on Computational Physics

IOP Publishing

IOP Conf. Series: Journal of Physics: Conf. Series 1290 (2019) 012013
doi:10.1088/1742-6596/1290/1/012013

3.1. Calculations Removed from Study

After careful consideration, certain combinations of basis set and theory methods were dropped from the study due to issues discussed in this section. The version of Gamess(US) used for this study, does not support basis sets with high angular momentum i-functions (5/6 zeta basis sets) to be utilized for predicting IR intensities or Raman activities. Zeta 5/6 Hessian calculations for CCSD(T) were too computationally intensive and did not provide significant improvements on molecular property predictions to justify their extreme runtimes. Sadlej-LpolX-fl was also

Table 4: VSCF Calculations for Anharmonic Frequencies in cm⁻¹ and Infrared Intensities in Km · mol⁻¹

| VSCF                  | Exp [50] | ω₁   | ω₂   | ω₃   | IR₁ | IR₂ | IR₃ |
|-----------------------|----------|------|------|------|-----|-----|-----|
| B3LYP/cc-pV6Z         | 3639.4   | 1546.2 | 3713.5 | 4.39 | 76.68 | 61.01 |
| B3LYP/cc-pV5Z         | 3639.8   | 1546.9 | 3713.7 | 4.17 | 76.16 | 58.07 |
| B3LYP/Aug-cc-pVQZ     | 3637.1   | 1545.8 | 3711.3 | 3.92 | 77.46 | 61.74 |
| B3LYP/May-cc-pVTZ     | 3635.2   | 1542.4 | 3710.7 | 5.61 | 71.47 | 59.03 |
| B3LYP/aug-pc3         | 3638.5   | 1545.5 | 3712.9 | 3.92 | 77.63 | 61.76 |
| B3LYP/Def2-QZVDP      | 3639.5   | 1545.8 | 3713.8 | 3.90 | 77.46 | 61.57 |
| MP2/cc-pV5Z           | 3680.3   | 1553.0 | 3779.6 | 6.53 | 72.86 | 74.81 |
| MP2/Aug-cc-pVQZ       | 3671.5   | 1550.1 | 3771.2 | 5.32 | 74.24 | 76.47 |
| MP2/May-cc-pVTZ       | 3677.3   | 1559.5 | 3775.3 | 8.49 | 65.02 | 71.19 |
| MP2/aug-pc3           | 3669.3   | 1551.0 | 3769.4 | 5.33 | 74.05 | 76.98 |
| MP2/Def2-QZVDP        | 3679.6   | 1552.5 | 3779.6 | 5.30 | 73.16 | 75.67 |
| SCS-MP2/Def2-QZVDP    | 3673.7   | 1569.2 | 3768.0 | 5.29 | 73.31 | 76.07 |
| CCSD(T)/Def2-QZVDP    | 3664.6   | 1564.7 | 3746.8 | 13.03 | 93.92 | 79.31 |
| CCSD(T)/Def2-TZVDP    | 3650.2   | 1569.5 | 3731.9 | 13.11 | 93.55 | 78.27 |
| CCSD(2)T/Def2-QZVDP   | 3760.9   | 1571.1 | 3752.2 | 3.64 | 73.08 | 58.12 |
| CCSD(2)T/Def2-TZVDP   | 3656.4   | 1571.0 | 3737.2 | 3.04 | 70.65 | 54.36 |

6
dropped from most calculations (MP2 within optimization and hessian calculations, across all theories within Raman calculations) due to convergence issues.

3.2. Molecular Geometries

The floating integer values within Figures 1-4 throughout the following sections are the zeta values for the basis sets. As expected, CCSD(T) performs best for calculating molecular geometries yielding consistent results with few errors. B3LYP converges faster than MP2 and CCSD(T), with the latter providing only slightly more accurate bond lengths and angles, for water.

![Figure 1: Unsigned Error of CCSD(T) Bond Angle Calculations Compared to Experimental Data](image)

It is notable to mention that the Pople-style basis sets provide reasonably accurate geometries, and it requires considerably high zeta basis sets to converge on the accurate geometry; particularly for the bond angle of water. While bond length converges quickly, for all levels of theory studied here, the bond angle converges more slowly, as shown for CCSD(T) in Figure 1. In general, most basis families show linear trend lines of slower convergence as either diffuseness or zeta values increase; making DFT with pople basis sets particularly well-suited for optimizing the geometries of large molecules. The Def2 family performs well in both runtime and accuracy, similarly both the PCseg and APCseg yield high accuracy predictions with lower runtimes, particularly when compared to their CC counterparts.

3.3. Vibrational Frequencies

When comparing vibrational frequencies to experimental data CCSD(T) calculations yielded the most accurate results, as expected. Interestingly, the NO7 and SNS basis set families yield results within ten wavenumbers (here, we consider this a reasonable threshold for the characterization or discrimination of most molecular species) with minimal runtimes [11].
Figure 2: Mean Unsigned Error of CCSD(T) Frequency Calculations Compared to Experimental Data

The Def2 basis family consistently outperforms both correlation consistent and polarization consistent basis sets, at significantly reduced computational expense. Both the augmented and non-augmented SPK- basis sets also perform very well in calculating frequencies, producing similar accuracy and runtimes as Def2 sets. While, for B3LYP, the commonly used Pople family, converges above the 10 $cm^{-1}$ threshold with similar runtimes. It is noteworthy to mention that fortuitous cancellation of errors often plays a large role here, particularly with smaller basis sets. However, if these errors consistently offset a deficiency in the level of theory encountered at the basis set limit, or that is not overcome unless much larger basis sets are used, this may prove useful. As an example, for the prediction of vibrational frequencies, the ‘may’ calendar basis sets consistently out-perform even their fully augmented counterparts at only a minor additional computational expense compared to the CCn basis at all levels of theory studied. B3LYP is known to over-estimate bond-lengths, and under-estimate harmonic frequencies. Often, a scaling factor is applied to reduce the errors in predicting anharmonic values from calculations performed under the harmonic approximation [51–53]. In Table 4 we can see clearly that the anharmonic frequencies for B3YLP are all below the experimental values. Again in Table 4, the anharmonic frequencies are shown for a few select levels of theory and basis sets. Here, we can see that B3LYP consistently underestimates the experimental frequencies, and that even the may-cc-pVTZ basis set values lie within 5 $cm^{-1}$ of the cc-pV6Z values. Similarly, we see for MP2 that the def2, may and pc families provide excellent agreement with the cc-pV5Z values. For each level of theory, the def2-QZVPD basis set approaches the accuracy of cc-pV5Z/cc-pV6Z much faster than most other basis sets of similar quality. Since anharmonic calculations are computationally expensive, CCSD(T) calculations were only performed using two of the def2 basis sets, where it could be seen that although an improvement over MP2 is observed, there is still not overall agreement with the experimental harmonic frequencies. This is likely because MP2 and CCSD(T) do not perform as well when used to calculate energetics at geometries far from the equilibrium position. To emphasize this point, calculations were additionally ran using the SCS-MP2 and CCSD(2)$_T$ levels of theory, which in each case show a marked improvement in the agreement with experimental anharmonic frequencies. All of the levels of theory considered here, struggle to reproduce the $\nu_2/\omega_2$ frequency. We note that the CCSD(2)$_T$/def2-QVZPD level of theory only has a MUE of $\approx 13$ $cm^{-1}$.
3.4. Infrared Intensities
From Figure 3, we can see that though the Pople and other small basis sets can outperform larger basis sets at predicting some infrared intensities, however, caution is urged since this is not always the case. Both the B3LYP and MP2 levels of theory have significant errors when calculating infrared intensities under the harmonic approximation, even when extrapolated to their CBS limits. The B3LYP theory level convergence line appears to lie just above the ten percent error margin, some basis sets - particularly small basis sets, such as Pople - deviate from this producing more accurate results. In some cases where small basis sets are utilized, the errors arising actually compensates for the deficiencies in B3LYP, producing seemingly more accurate results. However, these results may occur sporadically, and caution is urged using these results, which could easily be taken out of context [54].

![Infrared Intensities B3LYP](image)

Figure 3: Mean Unsigned Error of B3LYP Infrared Intensity Calculations Compared to Experimental Data

Both the harmonic and anharmonic calculation of the infrared frequencies is subject to the accuracy of the initial Hessian used to project the geometries, as well as the accuracy of the level of theory and basis set used to calculate the energetics and dipole moment. As explained previously, the MP2 and CCSD(T) levels of theory are not as reliable at geometries required for anharmonic, or even harmonic calculations. As a result, infrared intensities are not as dependent upon diffuse and polar functions as Raman activities, but rather how accurately the relative energies are calculated at each geometry, which explains why some basis sets that do not incorporate large polarization and diffuse functions are only necessary to the extent that the energetics and dipole moment are calculated with sufficient accuracy. Table 4 shows that by utilizing SCS-MP2 [55] or CCSD(2)T [56] levels of theory (the former at no additional cost, the latter at twice the cost of regular CCSD(T) calculations), further improvements in the frequencies and IR intensities can be obtained.

3.5. Raman Activities
In contrast to the IR intensities, the Raman activities can be very accurately calculated under the harmonic approximation, in particular when the MP2 level of theory is used (in agreement with previous studies [3, 4]) and are extremely sensitive to the degree of polarization and diffusivity incorporated into the basis sets.
It is clear from Figure 4 that while increasing the zeta-level or overall size of basis set is helpful, those basis sets that incorporate higher polarization levels and more diffuse functions are generally more accurate at reproducing Raman activities. In particular, the spectroscopic basis sets all perform well here (e.g., Sadlej, N07, SNS). The (PPD versions of) Def2 and (aug) Roos basis sets perform very well here also. The NLO-1 basis set also performs fairly well in calculating Raman activities, comparable to many quadruple-zeta basis sets. The dependence on the level of diffusivity is well demonstrated from the correlation consistent and calendar basis sets (cc→jun→jul→aug).

4. Conclusion

From our investigation we have identified various trends for the predictions of water’s molecular properties, of which the most notable is the Def2-n basis set family’s overall performance. Pople-style basis sets tend to be very consistent with their predictions, typically producing minimal changes between each individual basis set’s calculation. Pople basis sets are a reasonable choice for exploration, but there is little systematic improvement within their family, and are not recommended when results require high accuracy or quantitative values. Although they are occasionally somewhat accurate for infrared intensities when used with the B3LYP level of theory, they are not recommended to predict Raman activities. The Aug-CCn family is good for predicting Raman activities, particularly with the MP2 level of theory, however, they are computationally heavy, and outperformed by less computationally heavy options, such as aug-PCn, Def2-PD, and SPK augmented versions, as well as the majority of spectroscopic basis sets in this area. These basis sets have been shown to produce results comparable to six zeta basis sets, with minimal runtimes for the calculation of many optical properties. Another notable trend is the NO7 and SNS basis families success with calculating vibrational frequencies with harmonic approximation methods. Regarding the ANO series of basis sets, the NASA and Neese series offer high accuracy for geometries, as well as vibrational frequencies and infrared intensities, but lack sufficient diffuse functions to be able to accurately determine Raman activities (even the Aug-Neese sets). Only the Roos augmented ANO basis sets were able to accurately reproduce Raman activities, and represent an excellent all-round choice. We noticed that for the calculation of Raman activities basis sets which incorporated large numbers of diffuse functions always provided superior accuracy, however our scope was limited to the water molecule and further
investigations need to be conducted. We are currently increasing the scope of this study to include various other molecular species in the hopes of validating trends found in this study.

References

[1] Bennett C J, Ennis C P and Kaiser R I 2014 The Astrophysical Journal 782 63
[2] Glaser R, Hodgen B, Farrelly D and McKee E 2007 Astrobiology 7 455–470
[3] Zvereva E E, Shagidullin A R and Katsyuba S A 2010 The Journal of Physical Chemistry A 115 63–69
[4] Jiménez-Hoyos C A, Janesko B G and Scuseria G E 2008 Physical Chemistry Chemical Physics 10 6621–6629
[5] Huber K P 2013 Molecular spectra and molecular structure: IV. Constants of diatomic molecules (Springer Science & Business Media)
[6] Helgaker T, Jorgensen P and Olsen J 2014 Molecular electronic-structure theory (John Wiley & Sons)
[7] Halls M D and Schlegel H B 1998 The Journal of chemical physics 109 10587–10593
[8] Halls M D and Schlegel H B 1999 The Journal of chemical physics 111 8819–8824
[9] Scott A P and Radom L 1996 The Journal of Physical Chemistry 100 16502–16513
[10] Buczek A, Kupka T and Broda M A 2011 Journal of molecular modeling 17 2029–2040
[11] Galabov B, Yamaguchi Y, Remington R B and Schaefer H F 2002 The Journal of Physical Chemistry A 106 819–832
[12] Hill J G 2013 International Journal of Quantum Chemistry 113 21–34
[13] Jensen F 2013 Wiley Interdisciplinary Reviews: Computational Molecular Science 3 273–295
[14] Kupka T and Lim C 2007 The Journal of Physical Chemistry A 111 21–34
[15] Helgaker T, Klopper W, Koch H and Noga J 1997 The Journal of chemical physics 106 9639–9646
[16] Truhlar D G 1998 Chemical Physics Letters 294 45–48
[17] Petersson G and Al-Laham M A 1991 The Journal of chemical physics 94 6081–6090
[18] Becke A D 1993 The Journal of chemical physics 98 5648–5652
[19] Lee C, Yang W and Parr R G 1988 Physical review B 37 785
[20] Møller C and Plesset M S 1934 Physical Review 1074–1085
[21] Olson R M, Bentz J L, Kendall R A, Schmidt M W and Gordon M S 2007 Journal of Chemical Theory and Computation 3 1312–1328
[22] Hariharan P C and Pople J A 1973 Theoretica chimica acta 28 213–222
[23] Woon D E and Dunning Jr T H 1993 The Journal of chemical physics 99 3730–3737
[24] Papajak E and Truhlar D G 2010 Journal of chemical theory and computation 7 10–18
[25] Dunning Jr T H 1989 The Journal of chemical physics 90 1007–1023
[26] Woon D E and Dunning Jr T H 1994 The Journal of chemical physics 100 2975–2988
[27] Kendall R A, Dunning Jr T H and Harrison R J 1992 The Journal of chemical physics 96 6796–6806
[28] Papajak E and Truhlar D G 2010 Journal of chemical theory and computation 6 597–601
[29] Papajak E, Zheng J, Xu X, Leverentz H R and Truhlar D G 2011 Journal of chemical theory and computation 7 3027–3034
[30] Jensen F 2001 The Journal of Chemical Physics 115 9113–9125
[31] Jensen F 2002 The Journal of chemical physics 116 7372–7379
[32] Jensen F 2014 Journal of chemical theory and computation 10 1074–1085
[33] Weigend A and Ahlrichs R 2005 Physical Chemistry Chemical Physics 7 3297–3305
[34] Neese F and Valeev E F 2010 Journal of chemical theory and computation 7 33–43
[35] Baranowska A and Sadlej A J 2010 Journal of computational chemistry 31 552–560
[36] Benkova Z, Sadlej A J, Oakes R E and Bell S E 2005 Journal of computational chemistry 26 145–153
[37] Barone V, Cimino P and Stendardo E 2008 Journal of chemical theory and computation 4 751–764
[38] Barone V, Biczysko M and Bloino J 2014 Physical Chemistry Chemical Physics 16 1759–1787
[39] Paschoal D, Costa M F and Dos Santos H F 2014 International Journal of Quantum Chemistry 114 796–804
[40] Shimanouchi T 1977 Journal of physical and chemical reference data 6 993–1102
[41] Hoy A and Bunker P R 1979 Journal of Molecular Spectroscopy 74 1–8
[42] Culot F and Liévri J 1992 Physica Scripta 46 502
[43] Tew D P, Klopper W, Heckert M and Gauss J 2007 The Journal of Physical Chemistry A 111 11242–11248
[44] Gordon M S and Schmidt M W 2005 Advances in electronic structure theory: GAMESS a decade later (Amsterdam: Elsevier) pp 1167–1189
[45] Sayvetz A 1930 The Journal of Chemical Physics 7 383–389
[46] Ishimura K, Acc S N T C, McMurtrie L, Davidson E and Rys J 1993 J. Comput. Chem 14 1347–1363
[47] Krasnoschekov S V, Schutski R S, Craig N C, Sibaev M and Crittenden D L 2018 The Journal of chemical physics 148 084102
[49] Halkier A, Helgaker T, Jørgensen P, Klopper W, Koch H, Olsen J and Wilson A K 1998 *Chemical Physics Letters* **286** 243–252
[50] Johnson B G, Gill P M and Pople J A 1993 *The Journal of chemical physics* **98** 5612–5626
[51] Merrick J P, Moran D and Radom L 2007 *The Journal of Physical Chemistry A* **111** 11683–11700
[52] Andersson M P and Uvdal P 2005 *The Journal of Physical Chemistry A* **109** 2937–2941
[53] Halls M D, Velkovski J and Schlegel H B 2001 *Theoretical Chemistry Accounts* **105** 413–421
[54] Hudson R L, Gerakines P A and Ferrante R F 2018 *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **193** 33–39
[55] Grimme S 2003 *The Journal of chemical physics* **118** 9095–9102
[56] Piecuch P, Wloch M, Gour J R and Kinal A 2006 *Chemical physics letters* **418** 467–474