On the Computation of Dipole Moments: A Recommendation on the Choice of the Basis Set and the Level of Theory

Juan Camilo Zapata and Laura K. McKemmish*

School of Chemistry, University of New South Wales, 2052, Sydney

E-mail: l.mckemmish@unsw.edu.au

Abstract

Together with experimental data, theoretically predicted dipole moments represent a valuable tool for different branches in the chemical and physical sciences. With the diversity of levels of theory and basis sets available, a reliable combination must be carefully chosen in order to achieve accurate predictions. In a recent publication (J. Chem. Theory Comput. 2018, 14, 4, 1969–1981), Hait and Head-Gordon took a first step in this regard by providing recommendations on the best density functionals suitable for these purposes. However, no extensive study has been performed to provide recommendations on the basis set choice. Here, we shed some light into this matter by evaluating the performance of 38 general-purpose basis sets of single up to triple zeta-quality, when coupled with nine different levels of theory, in the computation of dipole moments. The calculations were performed on a data set with 114 small molecules containing second- and third-row elements. We based our analysis in regularised root mean square errors (regularised RMSE), where the difference between the calculated $\mu_{\text{calc}}$ and benchmark $\mu_{\text{bmk}}$ dipole moment values is derived as $(\mu_{\text{calc}}[D] - \mu_{\text{bmk}}[D])/(\max(\mu_{\text{bmk}}[D], 1[D]))$. This procedure ensures relative errors for...
ionic species and absolute errors for species with small dipole moment values. Our results indicate that the best compromise between accuracy and computational efficiency is achieved by performing the computations with an augmented double zeta-quality basis set (i.e. aug-pc-1, aug-pcseg-1, aug-cc-pVDZ) together with a hybrid functional (e.g. ωB97X-V, SOGGA11-X). Augmented triple-zeta basis sets could enhance the accuracy of the computations, but the computational cost of introducing such a basis set is substantial compared with the small improvement provided. These findings also highlight the crucial role that augmentation of the basis set with diffuse functions on both hydrogen and non-hydrogen atoms plays in the computation of dipole moments.

Introduction

Molecular dipole moments play a crucial role for many chemical properties. A single dipole moment arises from the difference in electronegativity between the two atoms that constitute a chemical bond. The three-dimensional nature of molecules allows individual dipole moments to act along different directions, which results in the addition or cancellation of these quantities depending upon the molecular geometry. Therefore, molecular dipole moments can deliver insight into the study of molecular structures. But it is not only in the elucidation of molecular structures that dipole moments play a crucial role. Strong dipole-to-dipole interactions are translated into higher melting and boiling points for a given compound. This is due to the strong intermolecular interactions that make it harder to separate the molecules in the compound. Dipole moments also play a crucial role in vibrational and rotational spectroscopy. A molecule with a permanent dipole moment will absorb radiation from the microwave region of the spectrum leading to rotational transitions, whereas the arising of a dynamic dipole moment by infrared radiation, derives in vibrational transitions for a given molecule.

With the diverse applications of dipole moments, the need of high-quality measurements of these quantities is imperative as it can serve as reference data for further experiments.
State-of-the-art quantum chemistry packages can provide a reliable way to generate these results, as they are capable of calculating dipole moments following a straightforward black-box procedure. However, considering the volume of basis sets and levels of theory available, the choice of a particular model chemistry – i.e. a level of theory and basis set combination – can be overwhelming.

A first step in addressing this issue was recently made by Hait and Head-Gordon, who calculated dipole moments for a data set containing 152 small molecules at the CCSD(T) level of theory, extrapolated to the complete basis set (CBS) limit. These benchmarking results were used to test the performance of 88 popular density functionals from all levels of Jacob’s Ladder. They found remarkable performance delivered by hybrid and double-hybrid functionals, whereas local functionals, in general, were found to deliver lower performance.

As dipole moments highly depend upon the tail region of the wave function, the basis set used in the computations must provide an accurate description of this region. This requirement can be fulfilled by augmenting the basis set with diffuse functions, which are normally s- or p-functions with very small exponents that force them to decay slowly with the distance from the nucleus. This feature allows the basis set to cover a wider range and therefore better describes the region far from the nucleus. An example where this improvement can be observed, is the study conducted by Halkier and co-workers whom demonstrated that the inclusion of diffuse functions in the basis set, results in accurate calculated dipole moments. Furthermore, they showed that diffuse functions provide a faster basis set convergence of the SCF part of the dipole moment in the calculations. More recently, Hickey and co-workers tested the influence of the Dunning’s cc-pV{D,T}Z and aug-cc-pV{D,T}Z bases and the Sadlej pVTZ basis set in the calculation of dipole moments. They found that basis sets of triple zeta-quality augmented with diffuse functions (i.e. aug-cc-pVTZ) deliver the most accurate results, whereas unaugmented basis sets resulted in considerable lower performance. Conway and co-workers used the Dunning’s aug-cc-pCVnZ basis set family to propose a strategy to calculate highly accurate dipole moments. Using a two-point for-
mula, they were able to produce high-quality dipole moments by extrapolating the energies calculated with small basis sets, at a substantial reduction of computational effort.

Though the knowledge of augmenting the basis set with diffuse functions in dipole moment calculations appears to be well established in the community, a comprehensive study to test the performance of different basis set families in these calculations has not been performed. The fact that nearly all general-purpose basis sets available possess an augmented version with diffuse functions present, implies that the basis set family quality will also have an influence in the calculations. Here, we aim to fill this gap by testing the performance of 38 general-purpose basis sets of single up to triple zeta-quality coupled with nine different levels of theory. The calculations are performed across a data set containing 114 molecules from second- and third-row elements.

This paper is organised as follows: Section describes the method used for our analysis. This includes an explanation of the data set, as well as the analysis procedure and computational details followed throughout the study. Section provides the results with a discussion made in terms of basis set zeta-quality. This section finishes with our recommendations on the best basis set and level of theory to be used in the computation of dipole moments. Finally, a summary of our findings and some concluding remarks are presented in Section.

Method

Data Set

We considered 114 molecules containing second- and third-row elements. These molecules were taken from the 152 small molecules studied by Hait and Head-Gordon, including their original geometries. Molecules containing Li, Be, B, Na and Al atoms were removed as some basis sets (pVTZ and ma-n) tested in this study were not available for those elements.

We assessed the performance of 38 general-purpose basis sets, from single up to triple zeta-quality. The choice of these basis sets was representative in order to cover a wide range
of common general-purpose basis sets. The basis sets considered were:

1. the Slater-mimicking minimal basis set STO-6G;<sup>16,17</sup>

2. the Pople-style basis sets 6-31G<sup>18</sup> and 6-31G.*<sup>19,20</sup> The bases 6-31+G<sup>25,26</sup> and 6-311+G<sup>25,27</sup> augmented with diffuse functions on non-hydrogen atoms only. The bases 6-31G*<sup>22,23</sup> and 6-311G*<sup>22,23</sup> augmented with polarisation functions on non-hydrogen atoms only. The bases 6-31+G*<sup>18–22,25–28</sup> and 6-311+G*<sup>22,27</sup> augmented with both diffuse and polarisation functions on non-hydrogen atoms, as well as the bases 6-31++G**<sup>18–22,25,27,28</sup> 6-311++G**<sup>22,27</sup> and 6-311++G(2df,2pd) which include diffuse and polarisation functions in both hydrogen and non-hydrogen atoms;

3. the Jensen polarisation consistent basis sets (pc-n<sup>29–32</sup> and aug-pc-n<sup>29–33</sup>, n = 0, 1 and 2) and their segmented version (pcseg-n<sup>34</sup> and aug-pcseg-n<sup>34</sup>, n = 0, 1 and 2);

4. the Dunning correlation consistent basis sets (cc-pVnZ<sup>35–38</sup> and aug-cc-pVnZ<sup>35–39</sup>, n = D, T);

5. the Ahlrichs-Karlsruhe basis set family def2-n<sup>40</sup>, the ma-n<sup>40,41</sup> bases augmented with diffuse functions on non-hydrogen atoms and the def2-nD bases augmented with diffuse functions on both hydrogen and non-hydrogen atoms<sup>40,42</sup> ( n = SVP, TZVP and TZVPP); and

6. the Sadlej pVTZ basis set<sup>43–45</sup> a triple-zeta basis set developed specifically to calculate dipole moments.

To test the interaction of the basis set and the level of theory, the calculations were performed with nine different levels of theory. We included the benchmark wave function methods Hartree-Fock (HF)<sup>46</sup> and MP2.<sup>47</sup> The density functional methods chosen were based on recommendations of Hait and Head-Gordon<sup>10</sup> namely: M06,<sup>48</sup> SOGGA11-X<sub>V</sub>,<sup>19</sup> ωB97X-V,<sup>50</sup> B2PLYP<sup>51</sup> and DSD-PBEPBE-D3.<sup>52</sup> We also included B3LYP<sup>53,54</sup> for comparison due to its popularity, especially for infrared spectroscopy, and included the dispersion corrected B3LYP-D3<sup>55</sup> (not considered by Hait and Head-Gordon) since dispersion is known to be important in accurate dipole moment predictions.
Analysis Approach

Our paper aims to identify suitable modest basis set and level of theory combinations, that provide high quality dipole moments for a given reasonable calculation time.

Figure 1: Combined Jacob’s Ladder and Pople diagram for *ab initio* quantum chemistry calculations.

Figure 1 presents our main tool to assess the accuracy of our calculations. The *y* axis is basis set size while the *x* axis uses a generalised ‘Level of Theory’ ordering both wave function methods and DFT functionals in order of computational time, thereby merging the traditional Pople diagram\(^{56}\) with Jacob’s Ladder.\(^{57}\) This figure clearly labels with lines the three important properties for determining the accuracy of our computational quantum chemistry calculations - basis set incompleteness error (BSIE), level of theory error (LoTE) and model chemistry error (ModChem error). Each of the three metrics can be calculated from the difference between two of the three calculations shown in red boxes in Figure 1: a modest calculation (i.e. moderate basis set and level of theory) whose accuracy is to be determined, a complete basis set calculation with the same level of theory and the exact solution to the Schrodinger equation.

For practical calculations, basis set and level of theory combinations should be chosen such that the basis set incompleteness error and level of theory error are comparable, with the least time consuming contribution having the lower error; e.g. if choice of level of
theory is more important for timings than basis set choice, then the basis set should be
increased in size until BSIE is less than LoTE or the basis set contribution to calculation
time dominates the level of theory contribution. The model chemistry error is related to the
basis set incompleteness and level of theory errors by the triangular inequality and is thus
at maximum the sum of these two component errors, but can go as low as zero if the basis
set incompleteness and level of theory errors cancel one another.

To calculate these three important properties, we based our analysis in regularised root
mean square errors (regularised RMSE) comparing calculated dipole moments ($\mu_{\text{calc}}[D]$) with
the benchmark values ($\mu_{\text{bmk}}[D]$) produced by Hait and Head-Gordon. These benchmark
values were changed accordingly with the property to be determined, i.e. BSIE, LoTE or
ModChem Error. The difference between the calculated and benchmark values was reg-
ularised as $(\mu_{\text{calc}} - \mu_{\text{bmk}})/(\max(\mu_{\text{bmk}}, 1[D]))$ following the same procedure that Hait and
Head-Gordon implemented in their calculations. This choice meant relative errors were cal-
culated for molecules with large dipole molecules but absolute errors were calculated for
molecules with small dipole moments.

**Benchmark values**

Though a true complete basis set calculation is impossible, a near CBS result suitable for
comparisons against double and triple zeta basis set results can be easily obtained through
very large basis sets or extrapolation from large basis sets. Similarly, CCSD(T) results are
a suitable substitute for exact solution to the Schrodinger Equation when considering the
accuracy of DFT calculations.

Our benchmark values were obtained from Hait and Head-Gordon except for B3LYP-
D3. For HF and MP2, they calculated these results from aug-cc-pV{Q,5}Z or aug-cc-
pV{T,Q}Z basis set extrapolation, whereas, for DFT, aug-pc-4 results were used without
extrapolation. Following the same procedure, our B3LYP-D3 benchmark values were calcu-
lated using aug-pc-4.
Computational Details

All computations were performed using the computational quantum chemistry package Q-Chem 5.2. The package Q-Chem 5.2 already includes most of the basis sets tested in this study, except for the ma-n family \(n = \text{SVP, TZVP, TZVPP}\) and the Sadlej pVTZ basis set. These basis sets were defined manually using the general basis command in the input file. The ma-n basis sets were downloaded from the Truhlar research group website at the University of Minnesota (https://comp.chem.umn.edu/basissets/basis.cgi) and the pVTZ basis was downloaded from the Basis Set Exchange website. 59–61

The workflow for the computations was carried out using the ACCDB collection created by Peverati and Morgante. 62 This collection is based on a Snakemake 63 flow containing optimised molecular geometries from different databases as well as a set of computational tools useful for file and data handling. The ACCDB collection is supported with a Q-Chem input file template that was modified, when needed, for some calculations. The templates for the input files can be found in SI documentation.

Results and Discussion

Performance at the CBS Limit

Our focus here on basis sets necessitates a reduction in the data set size from 152 to 114 molecules after excluding molecules containing Li, Be, B, Na and Al. In this section, we test whether this significantly affects our result.

The regularised RMSE values for the reduced data set used in the present study are visually presented in figure 2, along with the regularised RMSE values found by Hait and Head-Gordon using the 152 molecules data set. Figure 2 shows that the reduction of the data set does not lead to significant deviations from the overall behaviour delivered by the
Figure 2: Performance of the chosen levels of theory in the computation of dipole moments. The dark-blue and lilac bar plots correspond to the reduced data set (used in the present study) and the 152 molecules data set, respectively. The grey bar plot corresponds to the reduced data set analysed with regular RMSE instead of regularised RMSE. Apart from B3LYP-D3 (produced here), all data were collected from the data set benchmark study performed by Hait and Head-Gordon. The $x$ axis in the figure is organised as follows: Hartree-Fock, hybrid functionals, MP2 and double-hybrid functionals.

Though dispersion is known to play a crucial role in dipole moment predictions, B3LYP-D3 did not provide any improvement in the computations and was slightly outperformed by the B3LYP functional.

The grey bar plot in the figure presents the performance of all levels of theory in the reduced data set when no regularisation to the RMSE values is applied. Though the explicit numbers diverge between the data sets and the figures of merit used to perform the analysis (i.e. regularised vs non-regularised RMSE), the overall performance of the different levels of theory remains fairly comparable, indicating that the choice of regularised or non-regularised RMSE values will not have a significant influence in the subsequent analysis.
Finite Basis Set Performance

Based on the diversity of basis sets assessed in the present study, we decided to classify the basis sets in five different groups according with their augmentation with diffuse and polarisation functions:

1. *No Diffuse or Polarisation:* Basis sets with neither diffuse nor polarisation functions, i.e. STO-6G, 6-31G, 6-311G, pc-0 and pcseg-0;
2. *Diffuse only, no Polarisation:* Basis sets augmented with diffuse functions only, i.e. 6-31+G, 6-311+G, aug-pc-0 and aug-pcseg-0;
3. *Polarisation, no Diffuse:* Basis sets augmented with polarisation functions only, i.e. 6-31G*, 6-311G*, pc-{1,2}, pcseg-{1,2}, cc-pV{D,T}Z and def2-{SVP,TZVP,TZVPP};
4. *Polarisation & Diffuse on non-Hydrogen Atoms:* Basis sets augmented with polarisation functions and diffuse function on non-hydrogen atoms only, i.e. 6-31+G*, 6-311+G*, ma-{SVP,TZVP,TZVPP}; and
5. *Polarisation & Diffuse on Hydrogen and non-Hydrogen Atoms:* Basis sets augmented with polarisation functions and diffuse functions on both hydrogen and non-hydrogen atoms; i.e. 6-31++G**, 6-311++G**, 6-311++G(2df,2pd), aug-pc-{1,2}, aug-pcseg-{1,2}, aug-cc-pV{D,T}Z, def2-{SVP,TZVP,TZVPP}D and pVTZ

Tables 1 and 2 present the regularised RMSE for the model chemistry (ModChem Error) and basis set incompleteness errors (BSIE) of all basis sets, across the different levels of theory. Figure 3 provides a visual representation of this data including only the *Polarisation, no Diffuse* and *Polarisation & Diffuse on Hydrogen and non-Hydrogen Atoms* basis set classes. Each row in the figure corresponds to the zeta-quality of the basis set tested. The columns, on the other hand, indicate the different computational quantum chemistry errors: the first column corresponds to the model chemistry error (ModChem Error), whereas the second row corresponds to the basis set incompleteness error (BSIE). The bar plots underneath the data points represent to the level of theory error (LoTE) (dark-blue bar plot presented in figure 2).
Table 1: Regularised RMSE values for all basis sets across the different levels of theory tested in this study. The notation for the augmentation of the basis sets is given at the bottom of the table. The values correspond to the basis set incompleteness error (ModChem Error).

| Basis | CBS Limit | M06-2X | SOGGA11-X | MP2 | B2PLYP | DSD-PBEPBE-D3 | Average |
|-------|-----------|--------|-----------|-----|--------|---------------|---------|
| CBS   | 0.485 | 0.455 | 0.525 | 0.425 | 0.405 | 0.425 | 0.455 |
| M06-2X | 0.413 | 0.435 | 0.445 | 0.435 | 0.445 | 0.435 | 0.429 |
| SOGGA11-X | 0.425 | 0.425 | 0.445 | 0.425 | 0.435 | 0.425 | 0.429 |
| MP2 | 0.435 | 0.435 | 0.445 | 0.435 | 0.435 | 0.435 | 0.435 |
| B2PLYP | 0.445 | 0.445 | 0.445 | 0.445 | 0.445 | 0.445 | 0.445 |
| DSD-PBEPBE-D3 | 0.435 | 0.435 | 0.435 | 0.435 | 0.435 | 0.435 | 0.435 |
| Average | 0.429 | 0.429 | 0.445 | 0.429 | 0.435 | 0.429 | 0.435 |

### Table 2: Regularised RMSE values for all basis sets across the different levels of theory tested in this study. The notation for the augmentation of the basis sets is given at the bottom of the table. The values correspond to the basis set incompleteness error (BSI Error).

| BSIE | Augmentation | Level of Theory | CBS Limit | M06-2X | SOGGA11-X | MP2 | B2PLYP | DSD-PBEPBE-D3 | Average |
|------|--------------|----------------|-----------|--------|-----------|-----|--------|---------------|---------|
| CBS   | 0.195 | 0.075 | 0.025 | 0.150 | 0.175 | 0.025 | 0.150 | 0.175 | 0.025 |
| M06-2X | 0.450 | 0.400 | 0.350 | 0.400 | 0.350 | 0.400 | 0.350 | 0.400 | 0.350 |
| SOGGA11-X | 0.350 | 0.300 | 0.250 | 0.300 | 0.250 | 0.300 | 0.250 | 0.300 | 0.250 |
| MP2 | 0.300 | 0.300 | 0.250 | 0.300 | 0.250 | 0.300 | 0.250 | 0.300 | 0.250 |
| B2PLYP | 0.350 | 0.350 | 0.300 | 0.350 | 0.300 | 0.350 | 0.300 | 0.350 | 0.300 |
| DSD-PBEPBE-D3 | 0.350 | 0.350 | 0.300 | 0.350 | 0.300 | 0.350 | 0.300 | 0.350 | 0.300 |
| Average | 0.350 | 0.350 | 0.300 | 0.350 | 0.300 | 0.350 | 0.300 | 0.350 | 0.300 |

Columns notation:
1  **Augmented basis on both hydrogen and non-hydrogen atoms**
2  **Augmented basis on non-hydrogen atoms only**
3  **Regular basis**

### Notes:
- BSIE values are given for all basis sets across the different levels of theory tested in this study.
- The notation for the augmentation of the basis sets is given at the bottom of the table.
- The values correspond to the basis set incompleteness error (BSI Error).
Figure 3: Performance of different basis sets when coupled with the nine levels of theory tested. Each row in the figure corresponds to the zeta-quality of the basis set (double and triple), and the columns correspond to different computational quantum chemistry errors (model chemistry and basis set incompleteness errors, respectively). The bar plots underneath the data points correspond to the level of theory error represented as the dark-blue bar plot in figure 2.

It is clear that for most double and triple zeta basis sets, the basis set incompleteness error is larger than the level of theory error, demonstrating that for practical calculations of dipole moment, the basis set choice is at least as important as choice of density functional approximation. In most cases, the model chemistry error and basis set incompleteness error are very similar; only for some hybrid functionals with triple zeta basis sets does the level
of theory error dominate over the basis set incompleteness error in determining the overall model chemistry error. There is no significant evidence of large scale error cancellation between basis set incompleteness error and level of theory error. Even for the oft-utilised B3LYP/6-31G*, there is very minimal error cancellation and the basis set incompleteness error dominates: model chemistry error is 0.163, basis set incompleteness error is 0.167 and the level of theory (i.e. DFT) error is 0.055.

Figure 3 also shows that the performance of a given basis set is significantly enhanced by augmentation with diffuse functions on both hydrogen and non-hydrogen atoms. These bases led to the lowest regularised RMSE values when compared with their versions without diffuse functions, or with diffuse functions on non-hydrogen atoms only (see tables 1 and 2). In fact, when diffuse functions are only included on non-hydrogen atoms, the performance of the basis set is considerably worse. This is attributed to the fact that hydrogen atoms are often placed in the outer region of molecules, which must be accurately described to predict reliable dipole moments.

Overall, the double hybrid functionals perform noticeably poorer with finite basis sets than the hybrid functionals, despite the MP2 finite basis set results being superior to the MP2 infinite basis set results. The reasons for this result are unclear to the authors, but notable in terms of directing practical recommendation for calculation types especially as double hybrids are much more computationally expensive than hybrid functionals.

Though our results show that the performance of double and triple-zeta basis sets are overall clearly superior to single-zeta basis sets, the choice of exactly which double or triple-zeta basis set is used also substantially affects the quality of a dipole moment calculation. In fact, based on our results, the choice of the basis set is far more important than the choice of DFT functional for practical dipole moment calculations where very large basis sets are unsuitable.

We therefore consider in turn the best basis set choice of single, double and triple-zeta quality; the zeta quality is of course a good proxy for calculation time.
Single-zeta basis sets: Tables 1 and 2 clearly show that the single-zeta basis sets introduce basis set incompleteness and model chemistry errors far in excess of the level of theory error and thus should be avoid in dipole moment computations. Augmentation with diffuse functions on both hydrogen and non-hydrogen atoms does not improve the performance, i.e. aug-pc-0 and aug-pcseg-0 do not outperform other single-zeta basis sets. This demonstrates the crucial role that the basis set zeta-quality plays in electronic structure calculations. Although the computation of dipole moments depends upon an accurate description of the tail region of the wave function, and therefore augmenting the basis set with diffuse functions seems reasonable, the basis set used needs to provide an accurate electronic representation first, in order to deliver reliable results.

Double Zeta-Quality Basis Sets: The Jensen pc-1 and pcseg-1, the Ahlrichs-Karlsruhe def2-SVP and Dunning cc-pVDZ bases have very good performance, with regularised RMSE between 0.152 and 0.203 on average. The augmented Jensen and Dunning double-zeta basis sets, i.e. aug-pc-1 and aug-pcseg-1 and aug-cc-pVDZ, perform even better, with BSIE and ModChem errors approaching and even sometimes dropping below the infinite basis set level of theory error (LoTE) for hybrid functionals. The def2-SVPD basis has an slightly superior performance than the unaugmented def2-SVP, whereas ma-SVP – augmented only on non-hydrogen atoms – performs considerably worse. Notably, the Pople-style basis sets of double zeta-quality account for the worst performance of all double zeta-quality basis sets tested. Neither the augmentation with diffuse (6-31+G) nor polarisation (6-31G*) functions on non-hydrogen atoms only results in any improvement in the results. The 6-31+G* and 6-31++G** bases set deliver a significant improvement, but are still outperformed by the Jensen and Dunning basis sets.

The low-cost electronic structure approach PBEh-3c/def2-mSVP\cite{2008JCP...129.1743S} is known to provide fast accurate general chemical properties by means of error cancellation between BSI and level of theory errors. The regularised RMSE for this approach, when used for the calculation of dipole moments, is 0.164 (second row in table 1) which makes it comparable with the
average performance of the Jensen, Dunning and Ahlrichs-Karlsruhe unaugmented double-zeta basis sets. The model also outperforms most of the widely used B3LYP/6-31G and B3LYP/6-31+G models and its performance is comparable to the B3LYP/6-31G* model. However, when double-zeta bases augmented with diffuse functions on both hydrogen and non-hydrogen atoms are used, the PBEh-3c/def2-mSVP model is significantly outperformed.

**Triple Zeta-Quality Basis Sets:** Once again, the Jensen, Dunning and Ahlrichs-Karlsruhe basis sets account for the best overall performance with augmented functions reducing basis set incompleteness errors by about 20 to 40% relative to unaugmented functions. The Sadlej pVTZ basis set also has very strong performance, with the performance of the unaugmented Ahlrichs-Karlsruhe basis sets (def2-TZVP and def2-TZVPP) being similar. Finally, most of the Pople-style 6-311G bases, and its further augmentations, have the worst overall performance and do not improve overall to the double-zeta Pople 6-31G family. This poor performance can be attributed to the fact that Pople’s 6-311G family basis functions are too tight and the extra basis set flexibility acts to give flexibility to the description of the core and not valence electrons.\(^{65}\) The 6-311++G(2df,2pd) basis, with additional diffuse and polarisation functions on both hydrogen and non-hydrogen atoms, provides a significant improvement in the calculations, with regularised RMSE values similar to the unaugmented Jensen, Dunning and Ahlrichs-Karlsruhe bases, but it is still outperformed by the augmented version of the aforementioned basis sets.

**Spin-Polarised and Non-Spin-Polarised Molecules:**

Open-shell (spin-polarised) molecules often represent a bigger computational challenge than closed-shell molecules (non-spin-polarised).\(^{66}\) This requirement leads, in some cases, to different recommendations regarding level of theory and basis set choice. Here, we have divided the data set into spin-polarised and non-spin-polarised molecules to see the influence of the basis set choice in these two groups. Figure\(^{4}\) presents the average performance of the different basis set families when considering spin-polarised and non-spin-polarised molecules only.
The average performance was taken over double and triple-zeta quality bases only, as our results show that they provide the strongest performance in the computations.

Figure 4: Average performance of the different basis set families when considering only spin-polarised and non-spin-polarised molecules. The average performance across the reduced data set is also presented.

Apart from the Pople family, the figure shows that a better performance for a given basis set will be found when the computations involve non-spin-polarised molecules. However, the trend across the different groups of molecules remains fairly unchanged, indicating that there is no preference of basis set choice when considering spin-polarised and non-spin-polarised molecules. This implies that the recommendations given from the 114 molecules analysis can also be generalised to these two groups of molecules.

**Recommendations for Users**

Our results clearly show that Pople-style basis sets, e.g. 6-31G*, are substantially out-performed by Jensen, Dunning and Ahlrichs-Karlsruhe basis sets when calculating dipole moments and should not be considered among the options for these calculations. The widely used B3LYP/6-31G* model could be replaced with the PBEh-3c/def2-mSVP model as a much faster method with similar performance.

For a very high accuracy calculation for small molecules, double hybrid functional calculations should be paired with a quadrupole basis set or CBS-extrapolation, following the
recommendations of Hait and Head-Gordon.\textsuperscript{10}

For larger molecules or for studies with a large number of target molecules, we recommend a hybrid functional paired with the aug-pcseg-1 or aug-pc-1 basis sets. From the results of Hait and Head-Gordon,\textsuperscript{10} $\omega$B97X-V and SOGGA11-X are good choices since larger basis set calculations (e.g. aug-pcseg-2) can reliably be used to benchmark selected values if desired. The Jensen basis sets are recommended over the Dunning basis sets as the former were specifically optimised for DFT calculations and come in segmented and unsegmented versions to be suitable for different integral evaluation packages.

We justify our recommendations with figure\textsuperscript{5} which compares the performance of all basis sets tested when coupled with the best two hybrid functionals, $\omega$B97X-V and SOGGA11-X, in terms of approximate timings and errors. It is clear and aligns with our expectation that the triple-zeta basis sets (augmented and unaugmented) are much more time-consuming than the augmented double-zeta basis sets. Yet, compared to the augmented double zeta basis sets, the unaugmented triple zeta basis sets perform worse and the augmented triple zeta basis sets only slightly better, justifying our earlier recommendations.

It is worth noting that the Sadlej pVTZ basis, specifically optimised for dipole moments, can provide an outstanding performance in the calculation of dipole moments. However, one must conduct a careful selection of the hybrid functional chosen to perform the calculations as, in some cases, the coupling could result in a more time-consuming calculation (e.g. $\omega$B97X-V/pVTZ in figure\textsuperscript{5}) or derive in less accurate results (e.g. M06/pVTZ in tables \textsuperscript{1} and \textsuperscript{2}). This erratic performance can perhaps be attributed to the fact the original basis set design in 1988-1992 was for high level correlated methods as DFT functionals were not yet widely available. Furthermore, the Sadlej basis set is not available for some elements. Therefore, we are hesitant to recommend this basis set currently, though we support the overall practice of designing basis sets for specific applications.
Figure 5: Performance for all basis sets at the ωB97X-V and SOGGA11-X levels of theory (first and second row in the figure, respectively). The dotted line in each plot represents the respective level of theory error. The first column corresponds to the model chemistry error (ModChem Error) and the second column to the basis set incompleteness error (BSIE). The size of the data points in the figure represents the computational time required to perform the calculations with the most time-consuming molecule (CH$_3$O) from the data set with Q-Chem 5.2. Note that calculation times are very sensitive to program and platform, and are timings that should only be considered as indicative.

Concluding Remarks

To provide an effective yet reliable way to computationally predict dipole moments, here we have tested the performance of 38 general-purpose basis sets, of single up to triple zeta quality, in the computation of dipole moments. The basis sets were coupled with nine different levels of theory, belonging to hybrid and double-hybrid density functionals, and
wave function-based methods.

Our results show that the best compromise between accuracy and computational efficiency is achieved by coupling a hybrid functional (e.g. \(\omega\text{B97X-V} \) or \(\text{SOGGA11-X} \)) with an augmented double zeta-quality basis from the Jensen (aug-pc-1 or aug-pcseg-1) or Dunning (aug-cc-pVDZ) families. One could obtain a slightly superior performance by using an augmented triple zeta-quality basis set, but the computational cost makes it impractical.

Hait and Head-Gordon\(^{10}\) suggested that double-hybrid functionals provide the best way to predict dipole moments. However, our results show that it is only the case when the computations involve CBS limit results; double hybrid functionals were outperformed by hybrid functionals when they both were coupled with double and triple zeta-quality basis sets.

The triple-zeta Sadlej pVTZ basis can also provide reliable and fast results. However, probably because it was designed for correlated methods, we found it has some erratic performance with certain hybrid functionals, e.g. lower performance or significantly larger calculation times. Nevertheless, basis sets designed for specific properties have significant promise in enabling high performance for reduced cost.

Single zeta-quality and Pople-style basis sets should not be implemented in calculations involving dipole moments. These basis sets presented the worst regularised RMSE values across all different levels of theory.

Finally, our results highlight the relevant role played by diffuse functions in the computation of dipole moments. The augmentation of the basis sets with diffuse functions on both hydrogen and non-hydrogen atoms results in an overall enhanced performance when compared with the unaugmented basis sets.
Acknowledgement

This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS enabled capability supported by the Australian Government.

The authors declare no conflicts of interest.

Supporting Information Available

The following files are available free of charge.

The data underlying this paper are available in the paper itself and in its online supplementary material. These include the following files and folders:

- si_documentation.pdf - explanation of the files and workflow within the SI documentation. This PDF also contains the input file sample used to perform the calculations.
- dipole_analysis.xlsx - data analysis for the different method/basis set dipole moment calculations.
- comp.times.xlsx - timings for each method/basis set combination.
- NSP-dataset.xlsx - data analysis for non-spin-polarised molecules only.
- SP-data.xlsx - data analysis for spin-polarised molecules only.
- Basis_LoT_Performance - folder containing csv files with the relevant information from each output file.
- geometries_dipolemoment - folder containing all reference geometries used throughout the study.
References

(1) Sidgwick, N. V. Dipole Moment and Molecular Structure. *Chemical Reviews* **1936**, *19*, 183–194.

(2) Davies, M. *Some Electrical and Optical Aspects of Molecular Behaviour*; Elsevier Science, 1965; Chapter 3, pp 41–65.

(3) Piela, L. *Ideas of Quantum Chemistry*; Elsevier, 2007; Chapter 12, pp 615–680.

(4) Moldoveanu, S. C.; David, V. *Essentials in Modern HPLC Separations*; Elsevier, 2013; Chapter 4, pp 115–143.

(5) Siesler, H. W. *Encyclopedia of Spectroscopy and Spectrometry*, 3rd ed.; Elsevier Inc., 2016; pp 30–39.

(6) Engel, T.; Reid, P. *Physical Chemistry*, 3rd ed.; Pearson, 2013; Chapter 18, pp 457–479.

(7) Engel, T.; Reid, P. *Physical Chemistry*, 3rd ed.; Pearson, 2013; Chapter 19, pp 483–509.

(8) McQuarrie, D. A.; Simon, J. D. *Physical Chemistry: A Molecular Approach*, 1st ed.; University Science Books, 1997; Chapter 5, pp 157–179.

(9) McQuarrie, D. A.; Simon, J. D. *Physical Chemistry: A Molecular Approach*, 1st ed.; University Science Books, 1997; Chapter 13, pp 495–537.

(10) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *Journal of Chemical Theory and Computation* **2018**, *14*, 1969–1981.

(11) Nagy, B.; Jensen, F. In *Reviews in Computational Chemistry*; Parrill, A. L., Lipkowitz, K. B., Eds.; Reviews in Computational Chemistry; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017; Chapter 3, pp 121–126.
(12) Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P. Basis-set convergence of the molecular electric dipole moment. *Journal of Chemical Physics* **1999**, *111*, 4424–4430.

(13) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; Wiley, 2013; Chapter 5, pp 203–204.

(14) Hickey, A. L.; Rowley, C. N. Benchmarking quantum chemical methods for the calculation of molecular dipole moments and polarizabilities. *Journal of Physical Chemistry A* **2014**, *118*, 3678–3687.

(15) Conway, E. K.; Gordon, I. E.; Polyansky, O. L.; Tennyson, J. Use of the complete basis set limit for computing highly accurate ab initio dipole moments. *Journal of Chemical Physics* **2020**, *152*.

(16) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *The Journal of Chemical Physics* **1969**, *51*, 2657–2664.

(17) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *The Journal of Chemical Physics* **1970**, *52*, 2769–2773.

(18) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules. *The Journal of Chemical Physics* **1971**, *54*, 720–723.

(19) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **1972**, *56*, 2257–2261.

(20) Dill, J. D.; Pople, J. A. Self-consistent molecular orbital methods. XV. Extended
Gaussian-type basis sets for lithium, beryllium, and boron. *The Journal of Chemical Physics* **1975**, *62*, 2921–2923.

(21) Binkley, J. S.; Pople, J. A. Self-consistent molecular orbital methods. XIX. Split-valence Gaussian-type basis sets for beryllium. *The Journal of Chemical Physics* **1977**, *66*, 879–880.

(22) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics* **1982**, *77*, 3654–3665.

(23) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980**, *72*, 650–654.

(24) McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z =11–18. *The Journal of Chemical Physics* **1980**, *72*, 5639–5648.

(25) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *Journal of Computational Chemistry* **1983**, *4*, 294–301.

(26) Gill, P. M.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. The performance of the Becke-Lee-Yang-Parr (B-LYP) density functional theory with various basis sets. *Chemical Physics Letters* **1992**, *197*, 499–505.

(27) Spitznagel, G. W.; Clark, T.; von Ragué Schleyer, P.; Hehre, W. J. An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl. *Journal of Computational Chemistry* **1987**, *8*, 1109–1116.
(28) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* 1973, 28, 213–222.

(29) Jensen, F. Polarization consistent basis sets: Principles. *The Journal of Chemical Physics* 2001, 115, 9113–9125.

(30) Jensen, F. Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit. *The Journal of Chemical Physics* 2002, 116, 7372–7379.

(31) Jensen, F.; Helgaker, T. Polarization consistent basis sets. V. The elements Si–Cl. *Journal of Chemical Physics* 2004, 121, 3463–3470.

(32) Jensen, F. Polarization Consistent Basis Sets. 4: The Elements He, Li, Be, B, Ne, Na, Mg, Al, and Ar †. *The Journal of Physical Chemistry A* 2007, 111, 11198–11204.

(33) Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *The Journal of Chemical Physics* 2002, 117, 9234–9240.

(34) Jensen, F. Unifying General and Segmented Contracted Basis Sets. Segmented Polarization Consistent Basis Sets. *Journal of Chemical Theory and Computation* 2014, 10, 1074–1085.

(35) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* 1989, 90, 1007–1023.

(36) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *The Journal of Chemical Physics* 1993, 98, 1358–1371.

(37) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *The Journal of Chemical Physics* 1994, 100, 2975–2988.
(38) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theoretical Chemistry Accounts* **2011**, *128*, 69–82.

(39) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992**, *96*, 6796–6806.

(40) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297.

(41) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theoretical Chemistry Accounts* **2011**, *128*, 295–305.

(42) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *Journal of Chemical Physics* **2010**, *133*.

(43) Sadlej, A. J. Medium-size polarized basis sets for high-level correlated calculations of molecular electric properties. *Collection of Czechoslovak Chemical Communications* **1988**, *53*, 1995–2016.

(44) Sadlej, A. J. Medium-size polarized basis sets for high-level-correlated calculations of molecular electric properties. II. Second-row atoms: Si through Cl. *Theoretica Chimica Acta* **1991**, *79*, 123–140.

(45) Sadlej, A. J.; Urban, M. Medium-size polarized basis sets for high-level-correlated calculations of molecular electric properties. III. Alkali (Li, Na, K, Rb) and alkaline-earth (Be, Mg, Ca, Sr) atoms. *Journal of Molecular Structure: THEOCHEM* **1991**, *283*, 147–171.
(46) Fock, V. Naeherungsmethode zur Losung des quantenmechanischen Mehrkoerperproblems. *Zeitschrift fuer Physik* **1930**, *61*, 126–148.

(47) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618-622.

(48) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

(49) Peverati, R.; Truhlar, D. G. Communication: A global hybrid generalized gradient approximation to the exchange-correlation functional that satisfies the second-order density-gradient constraint and has broad applicability in chemistry. *The Journal of Chemical Physics* **2011**, *135*, 191102.

(50) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Physical Chemistry Chemical Physics* **2014**, *16*, 9904.

(51) Zhang, Y.; Xu, X.; Goddard, W. A. Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics. *Proceedings of the National Academy of Sciences* **2009**, *106*, 4963–4968.

(52) Kozuch, S.; Martin, J. M. L. Spin-component-scaled double hybrids: An extensive search for the best fifth-rung functionals blending DFT and perturbation theory. *Journal of Computational Chemistry* **2013**, *34*, n/a–n/a.

(53) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics* **1993**, *98*, 1372–1377.
(54) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, *98*, 11623–11627.

(55) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(56) Pople, J. A. Two-dimensional chart of quantum chemistry. *The Journal of Chemical Physics* **1965**, *43*, 229–230.

(57) Goerigk, L.; Mehta, N. A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User. *Australian Journal of Chemistry* **2019**.

(58) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics* **2015**, *113*, 184–215.

(59) Feller, D. The role of databases in support of computational chemistry calculations. *Journal of Computational Chemistry* **1996**, *17*, 1571–1586.

(60) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *Journal of Chemical Information and Modeling* **2007**, *47*, 1045–1052.

(61) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *Journal of Chemical Information and Modeling* **2019**, *59*, 4814–4820.

(62) Morgante, P.; Peverati, R. ACCDB: A collection of chemistry databases for broad computational purposes. *Journal of Computational Chemistry* **2019**, *40*, 839–848.

(63) Köster, J.; Rahmann, S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **2012**, *28*, 2520–2522.
(64) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *Journal of Chemical Physics* 2015, **143**, 0–19.

(65) Cox, C. S.; Zapata, J. C.; McKemmish, L. K. Mixed Ramp-Gaussian Basis Sets for Core-Dependent Properties: STO-RG and STO-R2G for Li-Ne. *Australian Journal of Chemistry* 2020.

(66) Bally, T.; Borden, W. T. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley, 1999; Vol. 13; Chapter 1, pp 1–97.