ELSEVIER

Journal of Molecular Structure (Theochem) 624 (2003) 239–249

THEO
CHEM

www.elsevier.com/locate/theochem

On the determination of bond lengths by ab initio methods: estimation of errors and some improvements

J. Sánchez Márquez, M. Fernández Núñez*

Departamento de Química-Física de la Universidad de Cádiz, Facultad de Ciencias, Apartado 40, 11510 Puerto Real, Cadiz, Spain

Received 8 September 2002; revised 29 October 2002; accepted 2 December 2002

Abstract

This work presents a very simple statistical analysis of errors in the ab initio determination of molecular geometries. This analysis has allowed us to separate the errors in systematic and random components and to realize that differences between experimental data and theoretical calculations are larger than we have initially supposed. It does not seem easy to reduce those differences beyond certain limit, but our analysis has brought us to some procedures to improve the ab initio methods to calculate bond lengths, by reducing the systematic error correlating the calculated data to the experimental ones, and by reducing the random error by mixing the results of different standard procedures.

Keywords: Molecular geometry; Bond length; Hartree–Fock; Moeller–Plesset; Density functional theory

1. Introduction

The molecular geometry is one of the molecular properties that has been calculated earlier by quantum mechanics methods [1] and wide data bases with molecular geometries calculated by this route exist, from long time ago [2] to more recent studies [8,10]. The comparison with experimental data of the first calculations accomplished with Hartree–Fock methods and minimal bases prove, as it is well known, that errors were several 10^{-2} Å for bond distances and several degrees for bond angles. It would be supposed that, today, using huge bases and introducing electronic correlation into calculations, the situation should be different. We have encountered that; on the contrary, though the results go better, the improvement does not seem comparable to the sophistication of the modern calculations. The situation can bring us even to doubt if the theoretical and experimental distances are actually the same parameters, or if they are related but not identical magnitudes, as is underlined in Refs. [3,4]. The aim of this work, that continues our preliminary paper [5] lies on the hypothesis that a statistical point of view, raising the problem in quantitative

---

* Corresponding author. Tel.: +34-956-016-328; fax: +34-956-016-471.

E-mail address: manuel.fernandez@uca.es (M. Fernández Núñez).

0166-1280/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0166-1280(02)00793-5

terms could bring us to some conclusions, and procedures to improve the theoretical calculation of molecular geometries.

2. A very simple statistical analysis

We have compared the experimental R_e values and calculated equilibrium bond lengths of 45 molecules: H_2, C_2, N_2, O_2, LiH, BeH, BH, CH, NH, OH, FH, LiF, BeO, BeF, BO, BF, CO, CF, NO, NF, CF_2, CN_2, CO_2, C_3, H_2O, OF_2, O_3, SiF_2, SO_2, SeO_2, Cl_2, ClO, ClF, NS, PN, PO, P_2, SF, SO, S_2, H_2S, HSi, SiO, CS, HCl. The theoretical calculations has been performed with GAUSSIAN98 [11] and the molecules chosen among those ones with well known experimental values [6–9]. In order to avoid biased results, *the sample includes very different types of bonds.*

We have employed two statistical parameters to analyze the differences between results obtained by different calculation procedures. The first one is the mean deviation (MD):

$$\mathrm{MD}(\bar{d}) = \frac{\sum_{i=1}^{n} (y_i - \tilde{y}_i)}{n} \qquad (1)$$

where $\{y_i\}$ and $\{\tilde{y}_i\}$ are internuclear distance distributions corresponding to the selected molecules. The first distribution is the experimental distances (R_e) and the second ($\{\tilde{y}_i\}$) is the calculated ones. This parameter goes to zero when all errors are *random*. Its difference from zero estimates the *systematic* error of each method.

The second parameter that we have used has been the mean square deviation (MSD):

$$\mathrm{MSD}(\sigma) = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}{n}} \qquad (2)$$

This parameter gives a measure of the difference between the sets $\{y_i\}$ and $\{\tilde{y}_i\}$, with independence of the sign of the individual differences. In fact the MSD estimates the *total* error of the method (as sum of systematic and random error). Consequently, the comparison of the MD with the MSD should permit to estimate what part of the error is systematic and what part is random.

3. Results

We have tested all the combinations of procedures Hartree–Fock, Moeller–Plesset (MP2), and several density functionals (B3LYP, B3PW91, B3P86) [12–14], with the basis: STO-3G, STO-6G, 3-21G, 6-31G, 6-311G, 6-311G(d,p), 6-311G(2d,p), 6-311 + G(2d,p), 6-311G(3df,3pd), 6-311++G(3df,3pd) [15,16].

Our work has required almost 2500 independent minimizations of molecular geometries, and it is not a simple task to show our results in words. However, they can be easily explained by means of some graphics. We will represent the errors (MSD or MD) against the procedures and basis sets, ordered from minor to longer calculation effort. This way we get an unexpected result: As we can see in Figs. 1 and 2, the calculated distances *do not improve significantly* beyond the base 6-311G(d,p), and this fact is almost independent from the procedure used to carry out the calculations (HF, MP2 or DFT).

We can observe in Fig. 2 (MD) that, as it is well known from long time ago, when we approach to the HF limit the calculated distances are shorter than the experimental ones, they have a systematical error of 0.02 Å for any basis of quality 6-311G(d,p) or better. In addition, it shows that when MP2 calculations approach to 'MP2 limit', the resulting distances are *longer* than the experimental ones(opposite behavior to the observed for Hartree–Fock). From Fig. 2 it can be seen too that for any of the checked procedures, the basis sets corresponding to a larger systematic error in distance calculations are 3-21G, 6-31G and 6-311G.

In order to check the independence of results on the selected set of molecules, we have verified that the first half part of the data lead to equivalent results as the second one. Results are shown in Figs. 3–6, and prove that the dependence is not relevant.

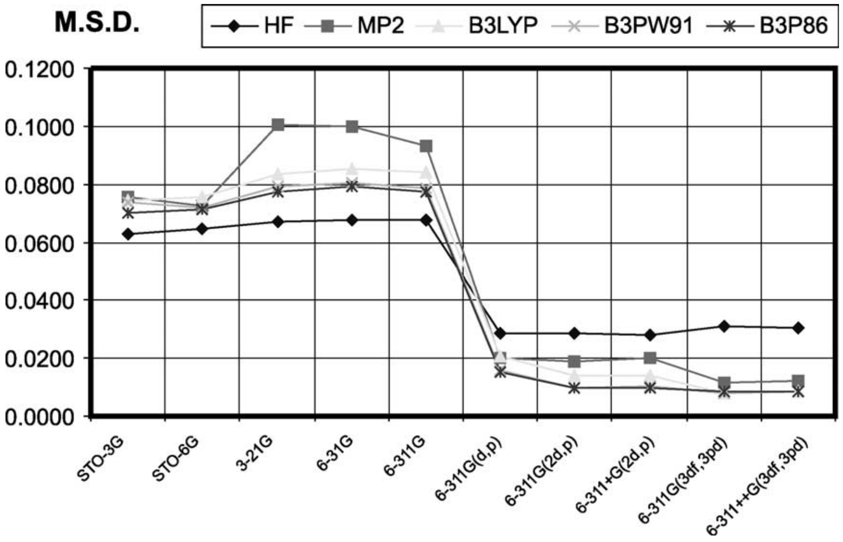

Fig. 1. Mean square deviation (Å) of calculated bond lengths (data abscissas from minor to longer calculation effort). The behaviour of all the procedures is similar. The importance of including polarization in the basis functions is clear, but the addition of more basis functions to 6-311G(d,p) does not improve the results.

4. Improvement of ab initio bond lengths calculation by reduction of its systematic error

The study of the correlation between theoretical distances and experimental ones, permits to realize that for almost every procedures—and especially in some of them—the results can be improved by using an adequate correlation function $R_e \approx f(\tilde{R}_e)$, where $\tilde{R}_e$ is the calculated value and R_e is the experimental one. That is not surprisingly, because this way we can eliminate a substantial part of the systematic error in the determination of the distances.

The meaning of the correlation function $R_e \approx f(\tilde{R}_e)$ can be justified by the following argument. The position of the minimum of an energy function can be estimated by substituting the exact function by a parabolic approximation:

$$E(R) = a_0 + a_1 \cdot R + a_2 \cdot R^2 \tag{3}$$

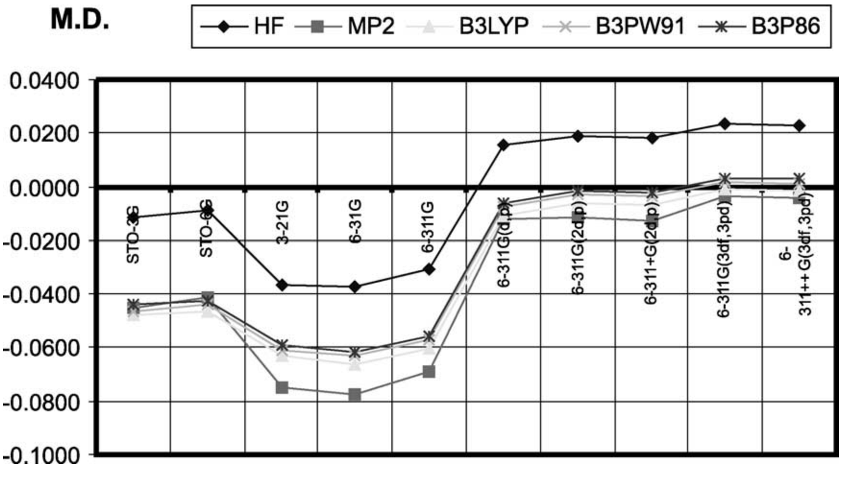

Fig. 2. Mean deviation (Å) of calculated bond length. A positive deviation indicates a calculated distance minor than the experimental one. Figure shows that the systematic errors increase slowly beyond the 6-311G basis set for HF, and decrease for post-HF methods.

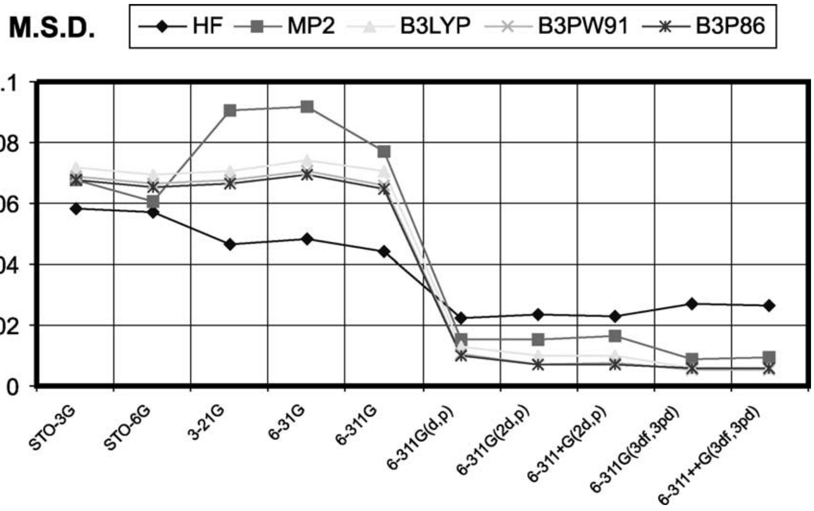

Fig. 3. Mean square deviation (Å). For the molecules: CH, NH, OH, FH, BeO, BeF, BO, BF, CO, CF, NO, NF, SO_2, SeO_2, S_2, H_2S, HSi, SiO, CS, HCl.

When we apply this approximation to the energy calculated with some theoretical method:

$$\tilde{E}(R) = \tilde{a}_0 + \tilde{a}_1 \cdot R + \tilde{a}_2 \cdot R^2 \tag{4}$$

and compare the conditions of minimum:

$$\left(\frac{\partial E}{\partial R}\right)_{R_e} = a_1 + 2 \cdot a_2 \cdot R_e = 0 \tag{5}$$

and

$$\left(\frac{\partial \tilde{E}}{\partial R}\right)_{\tilde{R}_e} = \tilde{a}_1 + 2 \cdot \tilde{a}_2 \cdot \tilde{R}_e = 0 \tag{6}$$

we obtain:

$$a_1 - \tilde{a}_1 + 2 \cdot (a_2 R_e - \tilde{a}_2 \tilde{R}_e) = 0 \tag{7}$$

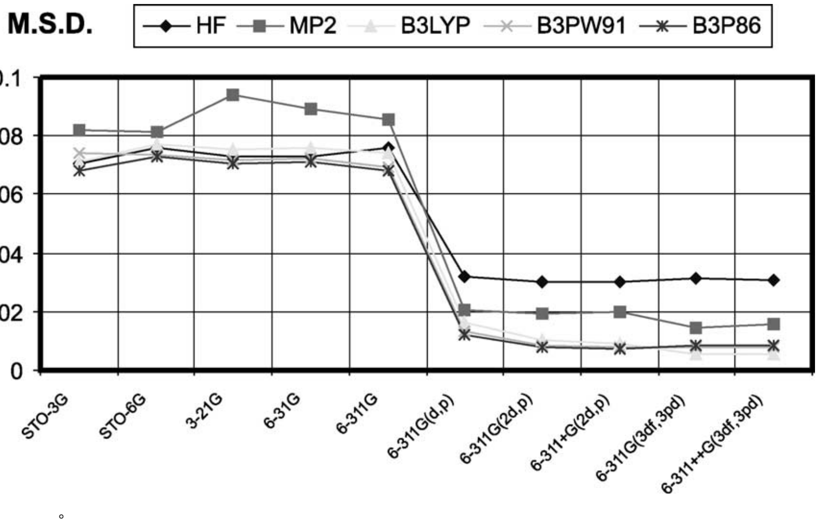

Fig. 4. Mean square deviation (Å). For the molecules: H_2, C_2, N_2, O_2, LiH, BeH, BH, LiF, CF_2, CN_2, CO_2, C_3, H_2O, OF_2, Cl_2, ClO, NS, PN, PO, P_2, SO.

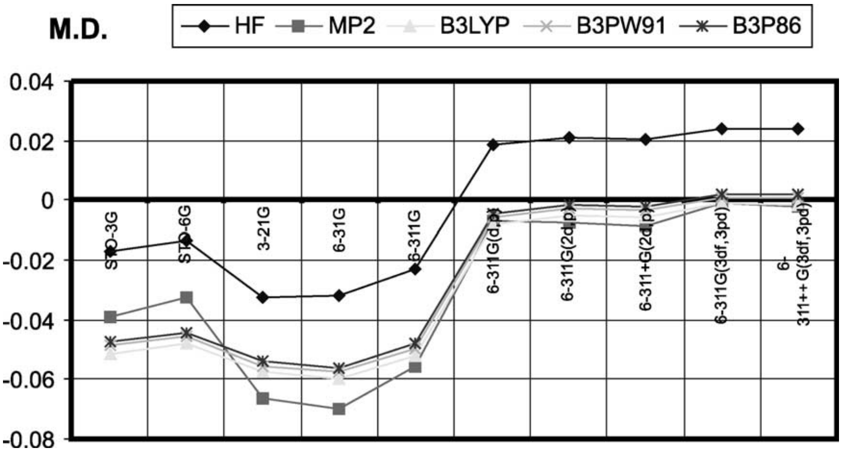

Fig. 5. Mean deviation (Å). For the molecules: CH, NH, OH, FH, BeO, BeF, BO, BF, CO, CF, NO, NF, SO_2, SeO_2, S_2, H_2S, HSi, SiO, CS, HCl.

and:

$$R_e = \frac{\tilde{a}_1 - a_1}{2\cdot a_2} + \frac{\tilde{a}_2}{a_2}\tilde{R}_e \qquad (8)$$

The parameters $(\tilde{a}_1 - a_1)/2\cdot a_2$ and $(\tilde{a}_2/a_2)$ *are dependent on the molecule that is considered*, but the calculated coefficients $\tilde{a}_1$ and $\tilde{a}_2$ can be expressed in a way:

$$\tilde{a}^i(\lambda) = a^i + \tilde{b}(\lambda) + \tilde{c}^i(\lambda)\begin{cases} i = 1, 2\ldots(\text{molecule}) \\ \lambda = 1, 2\ldots(\text{method}) \end{cases}$$

where $\tilde{b}(\lambda)$ represents the *systematic correction* (that depends, exclusively, of the method employed to calculate the energy), and $\tilde{c}^i(\lambda)$ represents the part of the error depending of the molecule under consideration. This allow us to express Eq. (8) as:

$$R_e^i = \left[\frac{\tilde{b}_1 + \tilde{c}_1^i}{2\cdot a_2}\right] + \left[1 + \frac{\tilde{b}_2 + \tilde{c}_2^i}{a_2}\right]\tilde{R}_e^i \qquad (9)$$

If we multiply to both sides by a_2^i we have:

$$a_2^i\cdot R_e^i = \left[\frac{\tilde{b}_1 + \tilde{c}_1^i}{2}\right] + [a_2^i + \tilde{b}_2 + \tilde{c}_2^i]\tilde{R}_e^i \qquad (10)$$

the coefficient a_2^i represent the experimental force constant (k^i) of the ith molecule. Centering our point of

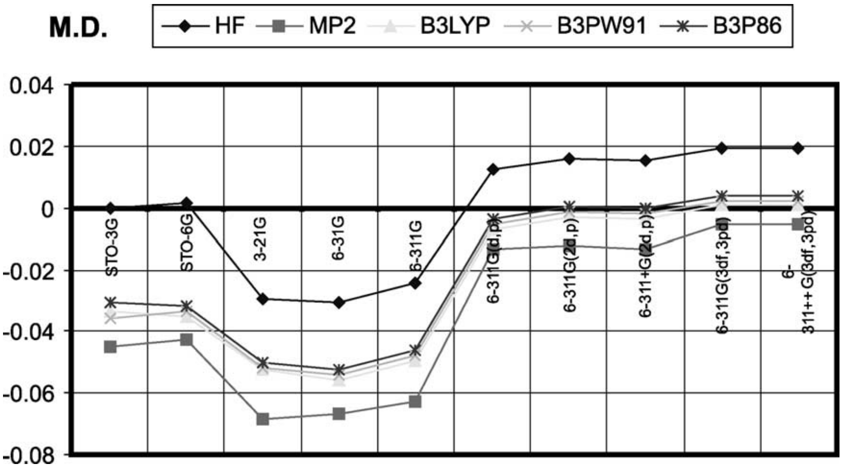

Fig. 6. Mean deviation (Å). For the molecules: H_2, C_2, N_2, O_2, LiH, BeH, BH, LiF, CF_2, CN_2, CO_2, C_3, H_2O, OF_2, Cl_2, ClO, NS, PN, PO, P_2, SO.

view in a concrete set of molecules, we can set:

$$k^i = \bar{k} + \varepsilon^i \tag{11}$$

where $\bar{k}$ represents the mean force constant for the set of molecules and ε_i is the variation of each constant with respect to the average. Using Eq. (11) we can rewrite Eq. (10) as:

$$(\bar{k}+\varepsilon^i)\cdot R_e^i = \left[\frac{\tilde{b}_1+\tilde{c}_1^i}{2}\right] + [(\bar{k}+\varepsilon^i)+\tilde{b}_2+\tilde{c}_2^i]\cdot\tilde{R}_e^i \tag{12}$$

or:

$$R_e^i = \left[\frac{\tilde{b}_1}{2\cdot\bar{k}} + \left(1+\frac{\tilde{b}_2}{\bar{k}}\right)\cdot\tilde{R}_e^i\right] + \left[\frac{\tilde{c}_1^i}{2\cdot\bar{k}} + \frac{\tilde{c}_2^i}{\bar{k}}\cdot\tilde{R}_e^i + \frac{\varepsilon^i}{\bar{k}}(\tilde{R}_e^i - R_e^i)\right] \tag{13}$$

The parameters appearing in the first bracket of Eq. (13) are independent of the considered molecule. As far as the second bracket was little enough we can approximately represent the bond lengths by a linear relation:

$$\tilde{R}_e^i \approx a + b\cdot\tilde{R}_e^i \tag{14}$$

Of course, the second bracket—that resume the influence of random errors—could not be little. But this assumption can be tested by fitting calculated data to the experimental ones, for an adequate sample of molecules. Therefore, we define an *improved calculated bond length*:

$$\tilde{R}_e^{\text{improved}} = \frac{\tilde{b}_1}{2\cdot\bar{k}} + \left[1+\frac{\tilde{b}_2}{\bar{k}}\right]\cdot\tilde{R}_e^i = a + b\cdot\tilde{R}_e^i \tag{15}$$

We have obtained the parameters *a* and *b* for the correlation function (15), using the same sample of 45 molecules employed in the first part of the work. The results are quoted in Table 1. As we can see in Table 2 the use of Eq. (15) improves the calculated bond lengths for all procedures except for Hartree–Fock with STO-*n*G basis sets. Nevertheless the improvements are important only for the small basis sets, and become negligible for the most complete basis sets. The interpretation of these results is easy. As can be seen in Fig. 2. The small basis sets STO-3G, STO-6G, 3-21G, 6-31G, 6-311G, correspond to the methods with the greater systematic error. For the 6-311G(d,p), 6-311G(2d,p), 6-311 + G(2d,p), 6-311G(3df,3pd), 6-311++G(3df,3pd) basis set, the main part of the error is random.

5. Improvement of ab initio bond lengths calculation by reduction of its random error

The improvement discussed in Section 4, decreases the systematic error for the bond lengths calculations, but it does not shorten the random error and, as we have already seen, when good basis sets are used the main part of error is random. Here we develop a method, which decreases this component of the error

Table 1
Coefficient *a* and *b* of the relation: $\tilde{R}_e^{\text{improved}} = a + b\cdot\tilde{R}_e^{\mathrm{i}}$ for 45 molecules

	Hartree–Fock		Moeller–Plesset		B3LYP		B3PW91		B3P86	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
STO-3G	0.0346	0.9662	0.0611	0.9239	0.0453	0.9338	0.0203	0.9521	0.0380	0.9417
STO-6G	0.0403	0.9638	0.0492	0.9350	0.0244	0.9493	0.0178	0.9558	0.0175	0.9571
3-21G	0.1336	0.8775	0.1701	0.8284	0.1271	0.8661	0.1174	0.8743	0.1163	0.8760
6-31G	0.1514	0.8636	0.1690	0.8278	0.1407	0.8545	0.1318	0.8626	0.1305	0.8645
6-311G	0.1607	0.8617	0.1750	0.8279	0.1569	0.8465	0.1432	0.8581	0.1423	0.8598
6-311G(d,p)	0.0482	0.9755	0.0421	0.9601	0.0452	0.9590	0.0323	0.9707	0.0326	0.9716
6-311G(2d,p)	0.0408	0.9835	0.0419	0.9608	0.0308	0.9729	0.0145	0.9872	0.0173	0.9860
···+G(2d,p)	0.0406	0.9834	0.0417	0.9600	0.0314	0.9719	0.0199	0.9825	0.0183	0.9848
···G(3df,3pd)	0.0264	0.9976	0.0185	0.9839	0.0155	0.9879	0.0042	0.9981	0.0028	1.0002
···++G(3df,3pd)	0.0261	0.9977	0.0177	0.9840	0.0152	0.9876	0.0045	0.9976	0.0031	0.9997

Table 2
Mean square deviation in Å for the calculation of improved bond lengths

	Hartree–Fock	Moeller–Plesset	B3LYP	B3PW91	B3P86
STO-3G	0.061(63)	**0.057(76)**	**0.054(74)**	0.055(73)	**0.052(70)**
STO-6G	0.063(65)	**0.057(73)**	0.058(76)	0.056(72)	0.056(71)
3-21G	**0.044(67)**	**0.043(101)**	**0.038(83)**	**0.035(79)**	**0.035(78)**
6-31G	**0.040(68)**	**0.036(100)**	**0.033(85)**	**0.030(81)**	**0.030(79)**
6-311G	**0.045(68)**	**0.033(93)**	**0.037(84)**	**0.034(79)**	**0.034(77)**
6-311G(d,p)	0.023(28)	**0.012(20)**	**0.014(21)**	**0.012(16)**	0.012(15)
6-311G(2d,p)	**0.021(28)**	**0.011(19)**	0.011(14)	0.009(09)	0.009(10)
···+G(2d,p)	**0.021(28)**	**0.011(20)**	**0.010(14)**	0.009(11)	0.009(10)
···G(3df,3pd)	**0.020(31)**	0.010(12)	0.007(08)	0.008(08)	0.008(08)
···++G(3df,3pd)	**0.020(31)**	0.011(12)	0.008(08)	0.008(08)	0.008(09)

Significative figures of errors before improvement are quoted between parenthesis. All procedures with improvements higher than 25% have their figures in bold.

by mixing the results of different procedures of calculation (method + basis set).

As is well known [18] for a linear combination of two independent estimations M_1 and M_2 of a magnitude:

$$M = \alpha_1 M_1 + \alpha_2 M_2 \tag{16}$$

the quadratic error is:

$$\sigma^2 = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 \tag{17}$$

For example, when M is a simple mean of two bond length with the same quadratic error σ_0, the quadratic error of the mean $M = (M_1 + M_2)/2$ will be:

$$\sigma = \frac{\sigma_0}{\sqrt{2}} \tag{18}$$

In our case, we introduce an adequate weight factor (α_i) for each component M_i of M, related to the confidence of each value M_i. We have tested a relationship between α_i and σ_i of potential form:

$$\alpha_i = \frac{k(n)}{\sigma_i^n} \tag{19}$$

where n is a parameter and $k(n)$ is a normalization constant. From Eqs. (17) and (19) we have:

$$\sigma^2 = k^2\left(\frac{\sigma_1^2}{\sigma_1^{2n}} + \frac{\sigma_2^2}{\sigma_2^{2n}}\right) = \frac{\sigma_1^2\sigma_2^{2n} + \sigma_1^{2n}\sigma_2^2}{(\sigma_1^n + \sigma_2^n)^2} \tag{20}$$

The values of σ_i for all the examined methods of theoretical calculation of bond lengths are known (Table 1), and we can choose the parameter n by minimization of Eq. (20). We find that the optimum value of n is 2, independently of the values of σ_1 and σ_2.

Another important subject is that there is not adequate to combine any couple of calculation methods. They must be chosen carefully, because the accuracy of Eq. (17) depends on two restrictive conditions: the errors of M_1 and M_2 must be *random*, and M_1 and M_2 must be *independent*.

In order to accomplish the first condition is necessary, at least, that the systematic errors of M_1 and M_2 are negligible. In Fig. 2 we can see that the combination of Hartree–Fock method with STO-*n*G basis set, or Moeller–Plesset method with 6-311G(d,p) or superior basis sets fulfill this requirement but, for example, the combination of MP2 with 6-31G is clearly inadequate.

In order to test the fulfillment of the second condition We can use the Pearson's correlation coefficient. We have calculated this parameter for the distributions of error of each couple of methods (Table 3). As is well known, the correlation coefficient goes to unity as $d_i(\lambda)$ goes to $d_j(\lambda)$. In other words, this coefficient goes to 1 when the results of methods i and j are *not* independent (see Fig. 7). Consequently, we should choose methods *with a low correlation coefficient* in Table 3.

We have found that, in order to fulfilling conditions of *random error* and independence, a good choice is the B3LYP/6-311G(3df,3pd) and MP2/6-311G(3df,3pd) methods ($\sigma_1 = 0.0080$

Table 3

Correlation coefficients between deviations ($R_i^{\mathrm{exp}} - R_i^{\mathrm{calc}}$) for representative couples of methods

		HF				MP2				B3LYP		
		STO-6G	6-311G	···+G(2d,p)	···++G(3dp,3pd)	STO-6G	6-311G	···+G(2d,p)	···++G(3df,3pd)	STO-6G	6-311G	···G(3df,3pd)
HF	STO-3G	0.999	0.354	0.094	0.181	0.632	0.062	0.145	0.427	0.917	0.379	0.239
	6-31G		0.966	0.269	0.026	0.238	0.737	0.551	0.098	0.327	0.902	0.644
	6-311G(d,p)			0.975	0.913	0.316	0.118	0.089	0.039	0.178	0.062	0.674
	···G(2d,p)			0.993	0.943	0.345	0.179	0.171	0.128	0.215	0.030	0.602
	···+G(2d,p)				0.951	0.378	0.183	0.133	0.068	0.251	0.035	0.602
	···G(3df,3pd)				0.999	0.462	0.426	0.335	0.083	0.364	0.299	0.426
MP2	STO-3G					0.994	0.483	0.399	0.141	0.846	0.506	0.140
	3-21G						0.914	0.769	0.292	0.395	0.784	0.242
	6-31G						0.962	0.832	0.379	0.329	0.868	0.335
	6-311G							0.817	0.316	0.310	0.875	0.390
	6-311G(d,p)							0.920	0.712	0.247	0.645	0.379
	6-311G(2d,p)							0.966	0.678	0.212	0.653	0.283
	···G(3df,3pd)								0.987	0.199	0.082	0.033
B3LYP	STO-3G									0.902	0.573	0.337
	STO-6G										0.538	0.277
	3-21G										0.921	0.452
	6-31G										0.987	0.571
	6-311G(d,p)											0.877
	···+G(2d,p)											0.907

In Figs. 7a and b we can see the distributions corresponding to the maximum and minimum values of this coefficient.

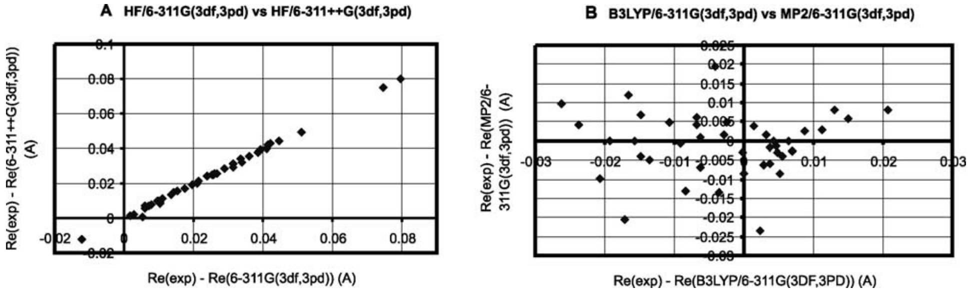

Fig. 7. Distributions of deviations ($R_i^{\text{exp}} - R_i^{\text{calc}}$) for HF/6-311G(3df,3pd)–HF/6-311++G(3df,3pd) in (A), and B3LYP/6-311G(3df,3pd)–MP2/6-311G(3df,3pd) in (B). These are the corresponding to the maximum (0.999) and minimum (0.033) values of the correlation coefficients quoted in Table 3.

and $\sigma_2 = 0.0116$). In this case we should use:

$$M = 0.676286 \cdot M_{\text{B3LYP/6-311G(3df,3pd)}} + 0.323714 \cdot M_{\text{MP2/6-311G(3df,3pd)}} \qquad (21)$$

whose MSD turns to be $\sigma = 0.0066$. This value is not much better than $\sigma_1 = 0.0080$ but later we will see that, in practice, the results of applying relation (21) seems to be considerably better than this foresight. Moreover, the calculation of bond length by two theoretical procedures so independent as B3LYP/6-311G(3df,3pd) and MP2/6-311G(3df,3pd) (Pearson's coefficient of 0.0033) is advisable because it gives a high level of confidence in case of agreement between both results.

6. Some applications

Formula (21) has been got by using the data corresponding to 45 diatomic and triatomic molecules. Of course it should be useful for any other

Table 4
Experimental, calculated and improved bond lengths of the 1,2,3-thiadiazole, 1,2,4-thiadiazole, 1,2,5-thiadiazole, 1,3,4-thiadiazole molecules [17]

1,2,3-Thiadiazole	$R_e(N_1-N_2)$	$R_e(N_1-S_5)$	$R_e(N_2-C_3)$	$R_e(C_3-C_4)$	$R_e(C_3-H_6)$	$R_e(C_4-S_5)$	$R_e(C_4-H_7)$	MSD
MP2/6-311G(3DF,3PD)	1.3171	1.6690	1.3475	1.3842	1.0785	1.6742	1.0768	0.0171
B3LYP/6-311G(3DF,3PD)	1.2755	1.7094	1.3632	1.3667	1.0780	1.6905	1.0763	0.0087
Improved	1.2890	1.6963	1.3581	1.3724	1.0782	1.6852	1.0765	0.0040
Experimental	1.2900	1.6920	1.3660	1.3690	1.0780	1.6890	1.0780	–
1,2,4-Thiadiazole	$R_e(C_1-N_2)$	$R_e(C_1-S_5)$	$R_e(C_1-H_6)$	$R_e(N_2-C_3)$	$R_e(C_3-N_4)$	$R_e(C_3-H_7)$	$R_e(N_4-S_5)$	MSD
MP2/6-311G(3DF,3PD)	1.3203	1.6963	1.0789	1.3542	1.3298	1.0794	1.6331	0.0102
B3LYP/6-311G(3DF,3PD)	1.3048	1.7140	1.0795	1.3646	1.3105	1.0805	1.6532	0.0051
Improved	1.3098	1.7083	1.0793	1.3612	1.3167	1.0801	1.6467	0.0025
Experimental	1.3130	1.7070	1.0790	1.3660	1.3170	1.0780	1.6490	–
1,2,5-Thiadiazole	$R_e(C_1-N_2)$	$R_e(C_1-S_5)$	$R_e(C_1-H_6)$	$R_e(N_2-C_3)$				MSD
MP2/6-311G(3DF,3PD)	1.3450	1.6170	1.3999	1.0794				0.0140
B3LYP/6-311G(3DF,3PD)	1.3189	1.6333	1.4194	1.0805				0.0045
Improved	1.3273	1.6280	1.4131	1.0801				0.0022
Experimental	1.327	1.63	1.417	1.081				–
1,3,4-Thiadiazole	$R_e(C_1-N_2)$	$R_e(C_1-S_5)$	$R_e(C_1-H_6)$	$R_e(N_2-C_3)$				MSD
MP2/6-311G(3DF,3PD)	1.3153	1.705	1.0779	1.3555				0.0114
B3LYP/6-311G(3DF,3PD)	1.2948	1.7274	1.0785	1.3652				0.0133
Improved	1.3014	1.7201	1.0783	1.3621				0.0103
Experimental	1.3200	1.721	1.079	1.371				–

The last column includes the mean square deviation for all bonds of the molecule.

Table 5
Experimental, calculated and improved bond lengths of the methylthiocianate molecule

Methylthiocianate	$R_e(S_1-C_2)$	$R_e(S_1-C_6)$	$R_e(C_2-H_3)$	$R_e(C_2-H_4)$	$R_e(C_6-N_7)$	MSD
MP2/6-311G(3DF,3PD)	1.8089	1.6846	1.0870	1.0849	1.1752	0.0088
B3LYP/6-311G(3DF,3PD)	1.8285	1.6888	1.0875	1.0852	1.1554	0.0089
Improved	1.8222	1.6874	**1.0873**	1.0851	1.1618	0.0070
Experimental	1.8237	1.6842	**1.0725**	1.0854	1.1697	–

The last column includes the mean square deviation for all bonds of the molecule.

molecule, not necessarily diatomic or triatomic. In order to verify this usefulness we have applied Eq. (21) to some representative polayatomic molecules. First, we have recalculated the bond lengths of the molecules recently studied by Glossman-Mitnik [17]. In Table 4 we can see that the attained improvement is fairly good.

Finally, by applying formula (21) to calculate the bond lengths in the methylthiocianate molecule we have been able to detect some problems with its standard experimental geometry. In Table 5 we compare the results of calculated bond lengths by B3LYP/6-311G(3df,3pd), MP2/6-311G(3df,3pd) and by using Eq. (21), with the experimental data recommended in Kuchitsu's book [6]. It can be seen that S_1-C_2, S_1-C_6 and $C_2-H_4(=C_2-H_5)$ experimental and calculated bond lengths are fairly coincident. On the contrary, the experimental and calculated values of the C_2-H_3 and C_6-N_7 bond lengths are too different. In the C_2-H_3 case the coincidence between the results of B3LYP and MP2 calculations (that, as we have proved in Section 6 are *independent*) points clearly to a problem with the experimental data. In the case of C_6-N_7 bond, the situation is not so clear as in the C_2-H_3 case, because experimental data lies between the B3LYP and the MP2 calculations, and this fact makes the possibility of a experimental mistake less evident.

7. Conclusions

1. The improvement of calculating bond lengths by Hartree–Fock, Moeller–Plesset, or DFT methods with basis sets beyond 6-311G(d,p), results unexpectedly poor (Figs. 1 and 2). The best choice to calculate bond lengths could be 6-311G(d,p) or 6-311G(2d,p), specially if we use B3P86 or B3PW91 procedures.
2. The error in calculated bond lengths can be divided in two independent components that we have named 'systematic error' and 'random error'. Further analysis has enabled us to reduce both kinds of errors.
3. It is possible to reduce systematic errors, through a least square fitting between calculated an experimental results for a sample of molecules. The most significant improvements correspond to the following cases:

Method	MSD	MSD (after improvement)	Differences
HF/6-31G	0.0677	0.0396	0.0281
MP2/6-31G	0.1001	0.0356	0.0645
B3LYP/6-31G	0.0853	0.0327	0.0526
B3PW91/6-31G	0.0807	0.0302	0.0505
B3P86/6-31G	0.0792	0.0304	0.0488
HF/6-311G(2d,p)	0.0284	0.0209	0.0075
MP2/6-311G(d,p)	0.0202	0.0119	0.0083
B3LYP/ 6-311G(d,p)	0.0205	0.0137	0.0068
B3PW91/6-311G(d,p)	0.0160	0.0117	0.0043
B3P86/6-311G(d,p)	0.0153	0.0118	0.0035

4. The random error of the calculated bond lengths can be reduced by mixing the results of certain pairs of procedures of calculation. A good choice is to combine the results of B3LYP/6-311G(3df,3pd) and MP2/6-311G(3df,3pd) by using the formula $M = 0.676286 \cdot M_{\mathrm{B3LYP/\dots}} + 0.323714 \cdot M_{\mathrm{MP2/\dots}}$
5. Our work strongly suggest that one of the C–H bond lengths of the standard geometry for methylthiocianate molecule should be considerably shorter than the value quoted in Kuchitsu's (Landolt-Börstein) book.

References

[1] W.G. Richards, T.E.H. Walker, R.K. y Hinkley, A Bibliography of Ab-initio Molecular Wavefunctions, Oxford University Press, Oxford, 1971, and supplements for 1970–73, 1974–77, 1978–80.

[2] R.A. Whiteside, M.J. Frisch, J.A. Pople, The Carnegie-Mellon Quantum Chemistry Archive, Third ed., Carnegie-Mellon University, Pittsburgh, 1983.

[3] M. Toyama, T. Oka, Y. y Morino, J. Mol. Spectrosc. 13 (1964) 193–213.

[4] M.Y. Hargittai, I. Hargittai, Int. J. Quantum Chem. 44 (1992) 1057–1067.

[5] J. Sánchez Márquez, M. Fernández Núñez. FCTL-Vol. XXVII, 1999, pp. 71–82.

[6] Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, New Series, group II, vol. 7, Springer, Berlin, 1976.

[7] CRC Handbook of Chemistry and Physics, 76th ed., CRC Press, 1997.

[8] M. Hargittai, Molecular structure of metal halides, Chem. Rev. 100 (2000) 2233–2301.

[9] K.P. Huber, G. Herzberg, Molecular Spectra and Molecular Structure, IV Constants of Diatomic Molecules, Van Nostrand, New York, 1979.

[10] S.L. Dixon, P.C. Jurs, J. Comput. Chem. 15 (1994) 733–746.

[11] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, GAUSSIAN98, Revision A.9, Gaussian Inc., Pittsburgh, PA, 1998.

[12] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[13] J.P. Perdew, Phys. Rev. B 33 (1986) 8822–8824.

[14] J.P. Perdew, Y. Wang, Phys. Rev. B 45 (1992) 13244.

[15] J.B. Collins, P.v.R. Schleyer, J.S. Binkley, J.A. Pople, J. Chem. Phys. 64 (1976) 5142.

[16] R. Krishnan, J.S. Binkley, R. Seeger, J.A. Pople, J. Chem. Phys. 72 (1980) 650.

[17] D. Glossman-Mitnik, A. Márquez-Lucero, J. Mol. Struct. (Theochem) 548 (2001) 153–163.

[18] R.L. Scheaffer, J.T. McClave, Probability and statistics for engineers, Ed. PWS-KENT publishing co., 1990.