

Population Analysis of Pair Densities – A Study of Basis Set Dependence

R. Ponec^a, F. Uhlik^a, and D. L. Cooper^b

^a*Institute of Chemical Process Fundamentals, Academy of Sciences of Czech Republic, Prague 6, Suchbát 2, 165 02 Czech Republic*

^b*Department of Chemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, United Kingdom*

Received June 1, 1994; revised September 19, 1994; accepted September 26, 1994

Based on work reported by several authors, one of us (RP) has introduced a Mulliken-like population analysis of pair densities. This scheme, introduced first at the semi-empirical level, was subsequently generalised both for SCF and for post-SCF *ab initio* methods. A potential problem is the basis set dependence that can be expected for all kinds of Mulliken-like approaches. The main purpose of the present study is to quantify this basis set dependence for a range of simple molecules, using basis sets ranging from STO-3G to TZVP quality. It is shown, except for low quality basis sets, that the »effective« pair populations derived from *ab initio* SCF calculations are relatively insensitive to the choice of basis set.

INTRODUCTION

There continues to be significant interest in schemes which allow the visualisation of quantum chemical calculations in terms of quantities closely related to more classical chemical concepts of bond order, charge density, valence, and the like.^{1–7} Another broad class of such techniques (amongst many others) is represented by various localisation or population analysis schemes.^{8–11} Into this last category, we may include studies dealing with the population analysis of the pair density.^{12,13}

The pair population analysis, originally derived using a geminal expansion of the pair density, was initially formulated only for semi-empirical methods based on orthonormal basis functions.¹³ However, even in this simple form, the scheme provided remarkably clear and transparent pictures of the bonding in molecules with well localised two-centre two-electron (2c-2e) bonds. The methodology was subsequently generalised to *ab initio* SCF and post-SCF methods (with $S_{pq} \neq \delta_{pq}$), and results were presented for a number of systems.¹⁴ The aim of the present study is to investigate the sensitivity of the calculated populations to the quality of the basis set.

THEORETICAL

The pair population formalism is described in detail in the previous studies^{13,14} and so we present here only a very brief account. As it is well known, the normalisation condition for the two-particle (spinless) density matrix $\mathbf{D}(pq|rs)$ for an N -electron system takes the form

$$\binom{N}{2} = \sum_{pqrs} \mathbf{D}(pq|rs) S_{pr} S_{qs} \quad (1)$$

in which S_{pr} and S_{qs} are overlap integrals in the a.o. basis. It proves useful to rewrite this expression using quantities which are symmetric or antisymmetric with respect to interchange of electron labels:*

$$\binom{N}{2} = \frac{1}{2} \sum_{pqrs} \mathbf{D}(pq|rs) (S_{pr} S_{qs} + S_{ps} S_{qr}) + \frac{1}{2} \sum_{pqrs} \mathbf{D}(pq|rs) (S_{pr} S_{qs} - S_{ps} S_{qr}) \quad (2)$$

In the special case of a closed-shell SCF wavefunction, we may use the equality

$$\mathbf{D}(pq|rs) = \frac{1}{2} \mathbf{D}(p|r) \mathbf{D}(q|s) - \frac{1}{4} \mathbf{D}(p|s) \mathbf{D}(q|r) \quad (3)$$

to rewrite equation (2) in the form

$$\binom{N}{2} = \frac{1}{8} \sum_{pq} \left[(\mathbf{DS})_{pp} (\mathbf{DS})_{qq} + (\mathbf{DS})_{pq} (\mathbf{DS})_{qp} \right] + \frac{3}{8} \sum_{pq} \left[(\mathbf{DS})_{pp} (\mathbf{DS})_{qq} - (\mathbf{DS})_{pq} (\mathbf{DS})_{qp} \right] \quad (4)$$

By direct analogy with a Mulliken analysis, we may define populations by restricting the double summations which appear in this last expression.¹⁴ Specifically, we define 'singlet' and 'triplet' pair populations, Π_{AB}^s and Π_{AB}^t , according to

$$\begin{aligned} \Pi_{AB}^s &= \frac{1}{8} \sum_{p \in A} \sum_{p \in B} \left[(\mathbf{DS})_{pp} (\mathbf{DS})_{qq} + (\mathbf{DS})_{pq} (\mathbf{DS})_{qp} \right] \\ &\quad + \frac{1}{8} (1 - \delta_{AB}) \sum_{p \in B} \sum_{p \in A} \left[(\mathbf{DS})_{pp} (\mathbf{DS})_{qq} + (\mathbf{DS})_{pq} (\mathbf{DS})_{qp} \right] \\ \Pi_{AB}^t &= \frac{3}{8} \sum_{p \in A} \sum_{p \in B} \left[(\mathbf{DS})_{pp} (\mathbf{DS})_{qq} - (\mathbf{DS})_{pq} (\mathbf{DS})_{qp} \right] \\ &\quad + \frac{3}{8} (1 - \delta_{AB}) \sum_{p \in B} \sum_{p \in A} \left[(\mathbf{DS})_{pp} (\mathbf{DS})_{qq} - (\mathbf{DS})_{pq} (\mathbf{DS})_{qp} \right] \end{aligned} \quad (5)$$

* In essence, we insert the identity $I = 1/2(I + P_{12}) + 1/2(I - P_{12})$, in which P_{12} transposes spatial coordinates.

in which A,B denote atomic centres. With these definitions, the normalisation condition now takes the form

$$\sum_{A \leq B} (\Pi_{AB}^s + \Pi_{AB}^t) = \left(\frac{N}{2} \right) \quad (6)$$

The construction of $(DS)_{pq}$ for a closed-shell SCF wavefunction involves a summation over all the doubly-occupied MOs, so that the Π_{AB}^s and Π_{AB}^t represent »total« or »all-electron« quantities. We can, of course, choose to restrict this summation to particular subsets of the MOs so as to define, for example, valence-electron pair populations.

The interpretation of the pair populations is based on the so-called »effective« pair populations¹³⁻¹⁶

$$\Pi_{AB}^{\text{eff}} = \Pi_{AB}^s - \frac{1}{3}\Pi_{AB}^t \quad (7)$$

which satisfy the additional condition

$$\sum_{A \leq B} \Pi_{AB}^{\text{eff}} = \frac{1}{2}N \quad (8)$$

We have found that the two-centre effective pair populations provide information about the connectivity of the atoms and that they are equivalent to the »bond orders« introduced some time ago by Mayer,¹⁷ as a generalization of the so-called Wiberg indices.⁷ The values are negligible except for centres linked by a formal bond in the structural formula. Moreover, the populations are more sensitive to the bond multiplicity than to the nature of the bonded atoms: values for double and triple bonds are roughly twice and three times larger than those for single ones.^{13,15} The one-centre quantities, on the other hand, carry information about core and non-bonded electron pairs.^{14,15}

In order to examine the basis set dependence of the populations, we have carried out *ab initio* SCF calculations for five small molecules which feature 2c-2e bonds. The molecular geometry was optimised with each basis set, and all-electron and valence-electron effective pair populations were then calculated. The different basis sets range in quality from minimal (STO-3G) to more extended TZVP sets. All of the basis sets – namely STO-3G, MIDI4, 4-31G, 6-311G*, 6-31G**, DZP and TZVP – were taken directly from the GAMESS-UK program.¹⁸ Results were also obtained with the semi-empirical MNDO method, as implemented in the MOPAC program.¹⁹

RESULTS AND DISCUSSION

As we have indicated, it is the effective pair populations that are central to the proposed population analysis. In addition to studying the basis set dependence of these quantities, it is useful to compare »all-electron« and »valence-electron« populations. We also assess the description of the bonding implied by the effective pair populations derived from the semi-empirical calculations.

The simplest situation occurs for the H_2 molecule, in that the SCF description of the bonding in this molecule corresponds to 50% covalent character and 50% ionic character for all bond lengths. As a consequence, all of the resulting populations are completely insensitive to the quality of the basis set. Each one-centre (HH) effective population is equal to 1/4 and the corresponding two-centre (H-H) value is exactly 1/2 for all basis sets.

Results for the other molecules considered in the present work – CH_4 , NH_3 , H_2O and HF – are collected in Table I(a)-(d). We distinguish between bonded and non-bonded two-centre values by means of the designation A—B and A...B, respectively, and we denote the one-centre terms AA. The basis sets are listed in the order of total energy. Net Mulliken populations on the hydrogen atoms are recorded in Table II for comparison.

TABLE I

Basic set dependence of all-electron and valence-electron effective pair populations

basic set	(a) CH_4			(b) NH_3		
	population	total	valence	population	total	valence
MNDO	HH	—	0.26	HH	—	0.21
	CC	—	0.98	NN	—	1.87
	C—H	—	0.49	N—H	—	0.49
	H...H	—	0.00	H...H	—	0.00
STO-3G	HH	0.22	0.22	HH	0.18	0.18
	CC	2.14	1.14	NN	3.00	2.00
	C—H	0.50	0.50	N—N	0.48	0.48
	H...H	0.00	0.00	H...H	0.00	0.00
MIDI4	HH	0.20	0.20	HH	0.16	0.16
	CC	2.27	1.27	NN	3.16	2.16
	C—H	0.49	0.49	N—H	0.46	0.46
	H...H	0.00	0.00	H...H	0.00	0.00
4-31G	HH	0.19	0.19	HH	0.12	0.12
	CC	2.34	1.35	NN	3.33	2.34
	C—H	0.48	0.48	N—N	0.43	0.43
	H...H	0.00	0.00	H...H	0.00	0.00
6-311G*	HH	0.22	0.22	HH	0.15	0.15
	CC	2.20	1.20	NN	3.15	2.15
	C—H	0.49	0.49	N—N	0.47	0.47
	H...H	0.00	0.00	H...H	0.00	0.00
6-31G**	HH	0.20	0.20	HH	0.14	0.14
	CC	2.26	1.26	NN	3.21	2.21
	C—H	0.49	0.49	N—H	0.46	0.46
	H...H	0.00	0.00	H...H	0.00	0.00
DZP	HH	0.20	0.20	HH	0.14	0.14
	CC	2.29	1.29	NN	3.18	2.18
	C—H	0.49	0.49	N—N	0.47	0.47
	H...H	0.00	0.00	H...H	0.00	0.00
TZVP	HH	0.21	0.21	HH	0.15	0.15
	CC	2.25	1.25	NN	3.14	2.14
	C—H	0.49	0.49	N—N	0.47	0.47
	H...H	0.00	0.00	H...H	0.00	0.00

TABLE I (continuation)

Basic set dependence of all-electron and valence-electron effective pair populations

basic set	(c) H ₂ O			(d) HF		
	population	total	valence	population	total	valence
MNDO	HH	—	0.17	HH	—	0.13
	OO	—	2.68	FF	—	3.41
	O-H	—	0.48	H-F	—	0.46
	H...H	—	0.00			
STO-3G	HH	0.17	0.17	HH	0.16	0.16
	OO	3.68	2.68	FF	4.36	3.36
	O-H	0.48	0.48	H-F	0.48	0.48
	H...H	0.00	0.00			
MIDI4	HH	0.12	0.12	HH	0.09	0.09
	OO	3.89	2.89	FF	4.51	3.51
	O-H	0.43	0.43	H-F	0.40	0.40
	H...H	0.00	0.00			
4-31G	HH	0.10	0.10	HH	0.07	0.07
	OO	4.00	3.00	FF	4.55	3.55
	O-H	0.40	0.40	H-F	0.37	0.37
	H...H	0.00	0.00			
6-311G*	HH	0.13	0.13	HH	0.10	0.10
	OO	3.83	2.83	FF	4.45	3.45
	O-H	0.46	0.46	H-F	0.45	0.45
	H...H	0.00	0.00			
6-31G**	HH	0.11	0.11	HH	0.09	0.09
	OO	3.89	2.89	FF	4.48	3.48
	O-H	0.44	0.44	H-F	0.43	0.43
	H...H	0.00	0.00			
DZP	HH	0.11	0.11	HH	0.08	0.08
	OO	3.89	2.89	FF	4.49	3.49
	O-H	0.45	0.45	H-F	0.43	0.43
	H...H	0.00	0.00			
TZVP	HH	0.12	0.12	HH	0.09	0.09
	OO	3.85	2.85	FF	4.47	3.47
	O-H	0.46	0.46	H-F	0.45	0.45
	H...H	0.00	0.00			

First of all, we observe for the most flexible basis sets that the pair populations calculated for a given molecule appear to be less sensitive to the choice of basis set than are the Mulliken populations on the hydrogen atoms. In addition, the one-centre »total« and »valence-only« populations calculated with a given basis set differ, to an excellent approximation, only by the numbers of pairs of core electrons. The two-centre populations differ hardly at all, as we might have expected. After taking account of the numbers of core electrons, the semi-empirical (MNDO) values resemble most closely the *ab initio* one calculated with the minimal (STO-3G) set.

Having shown that we can ignore the core electrons, we now describe briefly the structural interpretation of the various valence-only effective pair populations. We concentrate here on the results obtained with the most flexible of the basis sets (TZVP). Before comparing the values with those described earlier for H₂, it seems

TABLE II

Basis set dependence of the net Mulliken populations on the hydrogen atoms in various molecules.

Molecule	MIDI4	4-31G	6-311G*	6-31G**	DZP	TZVP
CH ₄	0.12	0.15	0.09	0.12	0.13	0.11
NH ₃	0.23	0.32	0.23	0.26	0.25	0.23
H ₂ O	0.32	0.40	0.29	0.34	0.34	0.31
HF	0.42	0.48	0.35	0.39	0.41	0.38

entirely reasonable to reduce the XX one-centre populations by the expected number of non-bonded pairs (*i.e.*, 'lone' pairs). In addition, we must be careful not to overcount the XX contribution, by including it in *each* of the H-X bonds. With this in mind, the valence-electron value of $I_{NN}^{\text{eff}} = 2.14$ for NH₃, for example, is 'corrected' to $(2.14 - 1)/3 = 0.38$. This strategy leads to the values listed in Table III. By analogy with the results described previously for H₂, we may interpret these 'corrected' effective pair populations in terms of contributions from covalent and ionic structures, as it is also shown in Table III.

TABLE III

Corrected populations (TZVP basis) appropriate to an individual 2c-2e bond, and their structural interpretation, by means of the analogy with H₂. The values do not always add to 100% because of rounding.

CH ₄	HH	21%	H ⁻ + CH ₃
	CC	31%	H ⁺ - CH ₃
	C-H	49%	H - CH ₃
NH ₃	HH	15%	H ⁻ + NH ₂
	NN	38%	H ⁺ - NH ₂
	N-H	47%	H - NH ₂
H ₂ O	HH	12%	H ⁻ + OH
	OO	42%	H ⁺ - OH
	O-H	46%	H - OH
HF	HH	9%	H ⁻ + F
	FF	47%	H ⁺ - F
	H-F	45%	H - F

We find very little change in the weight of the H-XH_n covalent structure, although the various basis sets do indicate a slight decrease from CH₄ to HF. Much more significant changes occur in the relative weights of the two ionic structures we have considered. We observe that the one-centre HH values decrease from CH₄ to HF. We may interpret this as a decrease in the weight of the H-XH_n⁺ ionic structure. At the same time, there is a corresponding increase in the weight of the H-XH_n⁻ ionic structure, which affects the one-centre XX values. There are analogous changes in the Mulliken populations (see Table II). All of these observations are, of course, entirely in keeping with our expectation from the increased electronegativity of X from CH₄ to HF.

CONCLUSIONS

Provided that sufficiently flexible basis sets are employed, the effective pair populations derived from the two-particle density matrix are somewhat less sensitive to the quality of the basis set than we might have anticipated. In addition, the differences between the »total« and »valence-only« populations coincide with the numbers of pairs of core electrons, as required.

These findings are very useful because they put the pair populations scheme on a much safer footing. In view of its simplicity and (relative) insensitivity to the choice of basis set, we believe that the method could usefully be incorporated into the existing quantum chemical programs as an additional means of interpreting and visualising the molecular electronic structure.

Acknowledgement. – This collaboration was made possible by a three-month Fellowship to RP within the framework of the European Community's »Action for Cooperation in Science and Technology« with Central and Eastern European countries.

REFERENCES

1. J. Cioslowski and S. T. Mixon, *J. Amer. Chem. Soc.* **113** (1991) 4142.
2. J. Ciołowski and P. Surjan, *J. Mol. Struct. (Theochem)* **255** (1992) 9.
3. P. Karafiloglou and J. P. Malrieu, *Chem. Phys.* **104** (1986) 383.
4. P. Karafiloglou, *ibid.* **140** (1990) 373.
5. M. S. Gopinathan and K. Jug, *Theor. Chem. Acta* **63** (1984) 497.
6. K. Jug, *Croat. Chem. Acta* **57** (1984) 941.
7. K. B. Wiberg, *Tetrahedron* **24** (1968) 1083.
8. R. Mulliken, *J. Chem. Phys.* **23** (1955) 1833, 2343.
9. R. McWeeny, *Proc. Roy. Soc. (London)* **A223** (1954) 63.
10. R. Ahlrichs and C. Ehrhardt, *Theor. Chim. Acta* **68** (1985) 231.
11. K. Roby, *Molec. Phys.* **27** (1974) 81.
12. R. Bochicchio, *J. Mol. Struct. (Theochem.)* **228** (1991) 209.
13. R. Ponec and M. Strnad, *Int. J. Quant. Chem.* **50** (1994) 43.
14. D. L. Cooper, R. Ponec, T. Thorsteinnsson, and G. Raos, *Int. J. Quant. Chem.* in press.
15. R. Ponec, *Coll. Czech. Chem. Commun.* **59** (1994) 505.
16. R. Ponec, *Croat. Chem. Acta* **67** (1994) 55.
17. I. Mayer, *Int. J. Quant. Chem.* **29** (1986) 73.
18. M. F. Guest and P. Sherwood, *GAMESS-UK User's Guide and Reference Manual Revision B. O. Ser. C*, Daresbury Laboratory, UK (1992).
19. J. J. P. Stewart, *J. Comput.-Aided Mol. Des.* **4** (1990) 1.

SAŽETAK

Populacijska analiza gustoće parova – ovisnost o raznom skupu

R. Ponec, F. Uhlík i D. L. Cooper

Polazeći od rezultata već danih u literaturi, jedan od autora (RP) predložio je Mullikenovsku populacijsku analizu gustoće parova, i to prvo na semi-empirijskoj razini. Kasnije je ona bila poopćena, kako za samoskladne (SCF) tako i za postupke *ab initio*. Moguće poteškoće pri analizi mogu, kao i u svima Mullikenovskim pristupima, proizići iz ovisnosti o baznom skupu. Glavna je nakana ovoga rada kvantitativno proučavanje ovisnosti o odabranom skupu za niz jednostavnih molekula, s baznim skupovima različite kvalitete od STO-3G do TZVP. S izuzetkom baznih skupova niske kvalitete, pokazano je da se »efektivne« populacije parova prilično neosjetljive na izbor baznog skupa.