J Math Chem (2015) 53:1651–1656
DOI 10.1007/s10910-015-0521-7

BRIEF COMMUNICATION

Least squares estimation of unknown molecular properties and quantum QSPR fundamental equation

Ramon Carbó-Dorca[1,2]

Received: 22 April 2015 / Accepted: 25 May 2015 / Published online: 4 June 2015

Abstract Several characteristic features of the least squares method when solving the quantum QSPR fundamental equation are discussed.

Keywords Quantum multimolecular polyhedra (QMP) · Quantum QSPR fundamental equation · Collective QMP condensed distances and indices · Least squares

1 The QQSPR fundamental equation involving a QMP and a collection of different properties

The quantum QSPR (QQSPR) fundamental equation can be deduced from the collective distances [1–3], which can be defined by taking into account the geometrical structure of any QMP [4], which is nothing else that a geometrical picture associated to the DF tag set, forming a quantum object set [5].

The general QQSPR form of the fundamental equation can be written as [4]:

$$\forall I = 1, M_K : \pi_I \approx \sum_P \omega_P A_{PI}, \tag{1}$$

where M_K is the number of molecules involved belonging to a QMP and possessing known values of some property π, which can be supposed that are collected in the

---

✉ Ramon Carbó-Dorca
ramoncarbodorca@gmail.com

[1] Institut de Química Computacional i Catàlisi, Universitat de Girona, 17071 Girona, Catalonia, Spain

[2] Departamento de Química, Universidad Técnica Particular de Loja, 1101608 Loja, Ecuador

set $\Pi = \{\pi_I \,|I = 1, M_K\}$. On the QQSPR equation right side the set of coefficients $\mathrm{O} = \{\omega_P \,|P = 0, 2, 3, \ldots\}$ might be determined to an arbitrary order. Finally, the set of matrix elements $\mathbf{A} = \{A_{PI} \,|P = 0, 2, 3, \ldots \wedge I = 1, M_K\}$ correspond to the Pth order element of the collective QMP condensed index associated to the Ith molecule [3].

The collective QMP condensed indices can be obtained from the expression of the Pth order QMP collective moment, related to QMP collective distances [6,7], which can be basically constructed with the origin shifted DF set, see for example reference [8], expression:

$$\forall P\colon \mu_P = M^{-1} \sum_L (\rho_L - \rho_C)^P \tag{2}$$

where ρ_C, the QMP centroid, is simply defined as the arithmetic mean [8]:

$$\rho_C = M^{-1} \sum_L \rho_L.$$

Note that the total number of QMP DF vertices: M, might be different from those which possess a known property attached, M_K. In fact, the total number of QMP vertices $M = M_K + M_U$, can be split into those M_K with known property attached and those M_U with unknown property values.

Condensing expression (2) corresponds to obtain the integral of the DF moment [2,3]:

$$\forall P\colon \langle \mu_P \rangle = M^{-1} \sum_L \left\langle (\rho_L - \rho_C)^P \right\rangle$$

and then it can be written:

$$\forall P, L = 1, M : A_{PL} = \left\langle (\rho_L - \rho_C)^P \right\rangle.$$

It can be also easily shown [4] that solving Eq. (1) to obtain the coefficients in the set: O, becomes equivalent to describe some QQSPR operator Ω, which generally speaking can fulfill:

$$\forall L = 1, M : \langle \Omega \rho_L \rangle = \pi_L,$$

supposedly known the DF set $\mathrm{P} = \{\rho_L \,|L = 1, M\}$, associated to the QMP DF vertices. This quantum mechanical expectation value expression above will also hold, whenever the QQSPR operator can be written as a linear combination of a set of Hermitian operators:

$$\Omega \approx \sum_P \omega_P \Omega_P;$$

where, besides the elements of the coefficient set O, it must be taken into account that:

$$\Omega_0 = I \wedge \Omega_1 = 0$$

and

$$\forall P = 0, 2, 3... \wedge L = 1, M\colon \langle \Omega_P \rho_L \rangle = A_{PL}.$$

In order to solve Eq. (1) one can set up an equivalent matrix equation, which can be written with the same form, but using only the M_K elements of the condensed QMP collective indices, associated to known properties:

$$\mathbf{A}\,|\mathbf{w}\rangle = |\boldsymbol{\pi}\rangle \tag{3}$$

where it has been used the formerly defined matrix $\mathbf{A}$, the column vectors collecting the QQSPR operator coefficients: $|\mathbf{w}\rangle = \{\omega_P\}$ and the set known property values: $|\boldsymbol{\pi}\rangle = \{\pi_I\,|I = 1, M_K\}$.

Equation (3) can be solved via a least squares procedure, which provides the intermediate equation:

$$\mathbf{A}^T\mathbf{A}\,|\mathbf{w}\rangle = \mathbf{A}^T\,|\boldsymbol{\pi}\rangle\,, \tag{4}$$

obtained by multiplying Eq. (1) on the left by the transpose matrix: $\mathbf{A}^T$. The well-known optimal result for the coefficients of the QQSPR operator might be easily obtained via:

$$|\mathbf{w}\rangle = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\,|\boldsymbol{\pi}\rangle\,. \tag{5}$$

2 General QQSPR equation

Equation (3) and the least squares optimal solution (5) can be easily generalized when, instead of the property column vector $|\boldsymbol{\pi}\rangle$, a set of properties is known in form of a matrix, whose columns are vectors corresponding to the set of Q different properties, that is:

$$\boldsymbol{\Pi} = \{|\boldsymbol{\pi}_R\rangle\,|R = 1, Q\}$$

Therefore, the coefficient vector $|\mathbf{w}\rangle$ associated to the QQSPR operator might be transformed into a matrix, whose columns correspond to a QQSPR operator adequate to every property considered:

$$\mathbf{W} = \{|\mathbf{w}_R\rangle\,|R = 1, Q\}\,.$$

Within this general formalism, Eq. (3) becomes:

$$\mathbf{AW} = \boldsymbol{\Pi} \tag{6}$$

and the corresponding optimal solution in the least squares sense will be converted into:

$$\mathbf{W} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\boldsymbol{\Pi}. \tag{7}$$

3 Estimating unknown property values and the nature of the optimal least squares solution of the QQSPR equation

Equation (7) can be used in Eq. (6) in order to obtain unknown properties for some of QMP DF vertices, which have not been attached to any known property, and then it can be written:

$$\mathbf{A}_U\mathbf{W} \approx \mathbf{\Pi}_U \rightarrow \mathbf{A}_U\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T\mathbf{\Pi}_K \approx \mathbf{\Pi}_U. \tag{8}$$

This result corresponds to an expression, which indicates that the unknown estimated property values just become the result of transforming, via a matrix operator, the known ones. To see how this statement can hold, it is sufficient to construct the matrix:

$$\mathbf{T}_{UK} = \mathbf{A}_U\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T; \tag{9}$$

then, it is straightforward to write:

$$\mathbf{T}_{UK}\mathbf{\Pi}_K = \mathbf{\Pi}_U;$$

therefore, the matrix (9) corresponds to the matrix expression of some operator, which is able to transform vectors belonging to the known property values set into vectors which belong to the unknown property values set.

4 Accuracy of the least squares solution of the QQSPR fundamental equation

When it is used the same matrix $\mathbf{A}_K$ into Eq. (6) instead of $\mathbf{A}_U$, then expression (8) can be rewritten as:

$$\mathbf{A}_K\mathbf{W} \approx \mathbf{\Pi}_K \rightarrow \mathbf{A}_K\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T\mathbf{\Pi}_K \approx \mathbf{\Pi}_K^{calc}$$

and thus, whenever it can be written:

$$\mathbf{\Pi}_K \approx \mathbf{\Pi}_K^{calc}, \tag{10}$$

also it can be obtained:

$$\mathbf{A}_K\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T\mathbf{\Pi}_K - \mathbf{\Pi}_K^{calc} \approx \mathbf{0} \rightarrow \left(\mathbf{A}_K\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T - \mathbf{I}\right)\mathbf{\Pi}_K \approx \mathbf{0}.$$

Taking into account that the original property matrix is not null by construction, so: $\mathbf{\Pi}_K \neq \mathbf{0}$, then it can be also considered that necessarily:

$$\mathbf{A}_K\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T - \mathbf{I} \approx \mathbf{0} \rightarrow \mathbf{A}_K\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1}\mathbf{A}_K^T \approx \mathbf{I} \rightarrow \mathbf{T}_{KK} \approx \mathbf{I}.$$

Then it can be easily deduced that the accuracy of the calculation, represented by the Eq. (10), and which has to be associated to the endomorphism:

$$\mathbf{T}_{KK}\mathbf{\Pi}_K = \mathbf{\Pi}_K^{calc},$$

has to be obviously associated to the nearness of the endomorphism $\mathbf{T}_{KK}$ to the unit matrix. This will be exactly fulfilled when the matrix $\mathbf{A}_K$ becomes non-singular, and thus it might be written:

$$\left(\mathbf{A}_K^T\mathbf{A}_K\right)^{-1} = \mathbf{A}_K^{-1}\left(\mathbf{A}_K^T\right)^{-1} \to \mathbf{T}_{KK} = \mathbf{I}. \tag{11}$$

In this case, because Eq. (11) holds, then Eq. (7) can be written in a simplified form as:

$$\mathbf{W} = \mathbf{A}_K^{-1}\mathbf{\Pi}_K,$$

therefore Eq. (8) can be also simplified in this circumstance to:

$$\mathbf{A}_U\mathbf{W} \approx \mathbf{\Pi}_U \to \mathbf{A}_U\mathbf{A}_K^{-1}\mathbf{\Pi}_K \approx \mathbf{\Pi}_U$$

and the transformation operator (9) now could be written simply as:

$$\mathbf{T}_{UK} = \mathbf{A}_U\mathbf{A}_K^{-1}.$$

5 Conclusions

Constructing a QMP, where some DF vertices might be also associated to a set of property values, an approximate set of operators can be obtained. This can be done just constructing the QQSPR fundamental equation and solving it via a general least squares technique. The QQSPR operators can be employed afterwards to evaluate the properties of the remnant QMP DF vertices possessing unknown property values.

In setting up and solving in this way the problem, one has several points to note, which can be found among the characteristics of QQSPR fundamental equation results [4]:

1. Nothing opposes to the fact that a general QQSPR equation might be built using known molecular property values attached to some QMP DF vertices. Its solution permitting to obtain a set of Hermitian operators, which can be further used to estimate values of molecular properties, unknown in other molecular structures forming the QMP.
2. This previous fact corresponds to establishing a causal relationship between QMP structures and aromaticity descriptors [9].
3. The basis of this relationship is based into quantum mechanical theoretical grounds.

Whenever Eq. (6) can be solved yielding results of the recomputed properties within reasonable error limits, and the expression (7) defining the QQSPR operators is reachable, then one can be confident that there has been obtained a global theoretical expression, which permits to estimate unknown values of molecular properties, constructed within the quantum mechanical limits of the theory.

Acknowledgments The author wishes to specify that this work has been developed under a Prometeo fellowship, and acknowledges the SENESCYT, Government of Ecuador, for granting it.

Conflict of interest None.

References

1. R. Carbó-Dorca, Multimolecular polyhedra and QSPR. J. Math. Chem. **52**, 1848–1856 (2014)
2. R. Carbó-Dorca, Quantum polyhedra, definitions, statistics and the construction of a collective quantum similarity index. J. Math. Chem. **53**, 171–182 (2015)
3. R. Carbó-Dorca, Collective vectors, condensed indices and quantum similarity, IQCC Technical Report TC-2015-2. Frontiers in computational chemistry (in preparation)
4. R. Carbó-Dorca, Quantum multimolecular polyhedra, collective vectors, quantum similarity and quantum QSPR fundamental equation, IQCC Technical Report TC-2015-4
5. R. Carbó-Dorca, E. Besalú, Shells, point cloud huts, generalized scalar products, cosines and similarity tensor representations in vector semispaces. J. Math. Chem. **50**, 210–219 (2012)
6. R. Carbó-Dorca, Collective euclidian distances and quantum similarity. J. Math. Chem. **51**, 338–353 (2013)
7. R. Carbó-Dorca, Communications on quantum similarity (4): collective distances computed by means of similarity matrices, as generators of intrinsic ordering among quantum multimolecular polyhedra, IQCC Technical Report TC-2015-1. WIREs (submitted for publication)
8. R. Carbó-Dorca, Mathematical aspects of the LCAO MO first order density function (5): centroid shifting of MO ShF basis set, properties and applications. J. Math. Chem. **51**, 289–296 (2013)
9. M. Bunge, *Causality and Modern Science: Third Revised Edition* (Dover, New York, 2011)