

# Quantum QSAR and the eigensystems of stochastic quantum similarity matrices

Ramon Carbó-Dorca

*Institut de Química Computacional, Universitat de Girona, Campus de Montilvi, 17071 Girona, Catalonia, Spain*

Received 19 June 2000

Row or column stochastic quantum similarity matrices (SM) are by construction non-symmetric square arrays. Thus, although representing a source of partial order within quantum object (QO) sets, they are not so easily connected to QO descriptors as the parent symmetric quantum SM do. In this paper, the eigensystems of such stochastic SM are analyzed and connected to the original quantum SM eigenvalues and eigenvectors. Simple relationships are found, providing row or column stochastic quantum SM with the power to be used as QO descriptors in both classical or quantum QSAR frameworks.

**KEY WORDS:** quantum similarity measures, quantum objects, quantum similarity matrices, stochastic quantum similarity matrices, generalized matrix eigensystems, quantum QSAR

## 1. Introduction

Recently, Klein described in detail the potential uses of *partially ordered sets* (posets) in the field of theoretical chemistry [1]. Partial order within a set arises when some object descriptor, associated to the elements of a given set, is not symmetric with respect to the involved object order. An appropriate example of such an occurrence can be linked to the matrices arising from a non-symmetric topological similarity measure, proposed by Mezey [2]. In this particular case the topological measure  $T(A, B)$ , involving two objects  $A$  and  $B$ , was defined in such a way as to yield a different value when computed in reverse order, that is,  $T(A, B) \neq T(B, A)$ .

At the same time, in the last years, it has been demonstrated in our laboratory that *quantum similarity measures* (QSM) constitute a universal unbiased source of *quantum object* (QO) descriptors (see for more details [3–8]). However, the QSM  $Z(A, B)$ , involving two QO  $A$  and  $B$ , are symmetric measures, that is,  $Z(A, B) = Z(B, A)$ , and consequently, the associated quantum *similarity matrices* (SM) generated in such manner become uninteresting for poset construction over *QO sets* (QOS). See appendix A, [9–13] and recent reviews [14–19] for more details on the employed definitions in this paragraph.

On the other hand, the structure and construction algorithms leading to the precise definition of quantum *stochastic SM* (SSM) [20] have been recently discussed. Due

to the fact that such SSM are non-symmetric, they seem to constitute a potential missing link between QSM and poset definition over QOS. Alternatively, it appears at first sight difficult to promote such non-symmetric SSM as candidates to a QO descriptor source, within a *quantitative structure–activity relationships* (QSAR) framework. This is so, because the QSAR dimension reduction problem is based on the descriptor matrix eigenvectors [21,22], and usually non-symmetric matrix eigensystems are not so easy to handle as the symmetric counterparts: see, for example, [23,24].

Taking into account all these considerations, the present paper will show in a first place, after properly and briefly defining QSM and SM, how quantum SM can be trivially transformed into row or column SSM. Then, once the quantum SSM structure is set, the eigensystems of such matrices will be studied and compared. Next, it will be discussed how a simple procedure interrelate the eigenvalues and eigenvectors of the involved stochastic matrix structures with a new geometrical point of view, in order to connect QO structure and properties. Such a procedure has been recently named *quantum QSAR* (QQSAR) [20]. Due to the present study, a novel set of QO descriptors, based on quantum similarity concepts, could be taken into account and used in QSAR modeling.

## 2. Quantum similarity measures

The first precise definition of QSM has been proposed several years ago (see, e.g., [25,26]). Actually, any kind of QSM can be considered made by an application of the direct product of several QO tags, belonging to some QOS, into the set of the positive definite real numbers  $\mathbb{R}^+$ ; see [14–17] and also appendix A.

In the simplest way, a QSM may be defined by means of two QO  $\{\omega_A; \omega_B\}$ , using their tags, constructed in terms of quantum-mechanical *density functions* (DF)  $\{\rho_A; \rho_B\}$ . Both DF tags can be also connected by means of a chosen *positive definite* (PD) operator  $\Omega$ , through the volume integral

$$z_{AB}(\Omega) = \iint \rho_A(\mathbf{r}_1)\Omega(\mathbf{r}_1; \mathbf{r}_2)\rho_B(\mathbf{r}_2) \, d\mathbf{r}_1 \, d\mathbf{r}_2. \quad (1)$$

Then, for instance, substituting the positive definite operator  $\Omega$  in equation (1), by the Dirac's delta function  $\delta(\mathbf{r}_1 - \mathbf{r}_2)$ , then an *overlap-like* QSM is obtained:

$$z_{AB} = \int \rho_A(\mathbf{r})\rho_B(\mathbf{r}) \, d\mathbf{r}. \quad (2)$$

Similarly, when a Coulomb operator  $\Omega(\mathbf{r}_1; \mathbf{r}_2) = |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$  is employed in equation (1) a *Coulomb-like* QSM appears connected with the integral form:

$$z_{AB}(\mathbf{r}^{-1}) = \iint \rho_A(\mathbf{r}_1)|\mathbf{r}_1 - \mathbf{r}_2|^{-1}\rho_B(\mathbf{r}_2) \, d\mathbf{r}_1 \, d\mathbf{r}_2. \quad (3)$$

A QSM  $z_{AA}$ , involving only a unique QO tag, is usually called a *Quantum Self-Similarity Measure* (QSSM):

$$z_{AA}(\Omega) = \iint \rho_A(\mathbf{r}_1)\Omega(\mathbf{r}_1; \mathbf{r}_2)\rho_A(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \quad (4)$$

QSM, defined in this way, can be interpreted as generalized volumes and they can be easily linked with quantum-mechanical expectation values [12], see also appendix C for a discussion on this crucial point. Many alternative QSM definitions are possible, and throughout its historical development, quantum similarity has become an extremely flexible and general theoretical tool [18].

### 3. Quantum similarity matrices

Given a well-designed QOS, the collection of QSM involving all the QO pairs can be ordered in the form of a symmetric matrix:  $\mathbf{Z} = \{z_{IJ}\}$  and  $\mathbf{Z} = \mathbf{Z}^T$ . Such an ordered QSM array is called a *Similarity Matrix* (SM). Any SM can be partitioned in terms of its columns (or rows)<sup>1</sup>, that is:

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N) = \{\mathbf{z}_I\}, \quad (5)$$

where every column corresponds to a precise QO, belonging to the studied QOS [11,14–18,25,26].

Thus, if  $\omega_A$  is a given QO and  $\rho_A$  is associated to its DF tag, then the corresponding column of the SM  $\mathbf{z}_A$ , is in connection to the  $\omega_A$  QO DF tag, and can be consequently considered as a *discrete* QO representation. This situation can be symbolically written as

$$\omega_A \leftrightarrow \rho_A \leftrightarrow \mathbf{z}_A;$$

or, in a better way, employing the tagged set notation, as

$$(\omega_A; \rho_A) \leftrightarrow (\omega_A; \mathbf{z}_A),$$

taking into account the QO tagged set structure, customarily made by the ordered pairs (quantum system; DF).

In this manner, the QSM collection constructed by means of the QSM integrals, involving any chosen  $I$ th QO with respect to *all* the elements of the QOS, defines an  $N$ -dimensional discrete *tag set*  $\{\mathbf{z}_I\}$ ,  $N$  being the cardinality of the QOS, which can substitute the former infinite-dimensional DF tags.

A new *discrete QOS* (DQOS), of the same cardinality as the original QOS, can be constructed in this way: the *tagged set* objects are the same quantum systems of the QOS

<sup>1</sup> The most common matrix-to-hypermatrix transformation procedure is performed by means of a column matrix partition. However, here, later on, a row matrix transformation shall be used, demanding a row partition. At the moment to distinguish which one of the two possible vector partition forms is adopted, then bra-ket formalism will be used to note the difference between row and column partition vector forms.

as before, but the columns or rows of the SM substitute the initial DF tag set elements by appropriate discrete vectors. The DQOS discrete tags set naturally defines a polyhedron in  $N$ -dimensional space. The DQOS tag set, for obvious reasons, can be called a *QO point cloud* [19].

#### 4. Stochastic quantum similarity matrices

The computation of SM over a QOS providing a new DQOS structure, as discussed above, produces a set of  $N$ -dimensional tags, which can be associated to the original infinite-dimensional DF tags. Despite the strict positive definiteness of the SM column set elements,  $\{\mathbf{z}_I\}$ , which appear as a consequence of the QSM definition presented in equation (1), the connection between the  $N$ -dimensional tags and the DF is not immediately made evident. However, it can be easily deduced after taking into account the nature of the involved DF tags [14–19,25,26], which can be considered in turn either as positive definite functions or projection operators.

The present section will describe a SM transformation, producing a new collection of  $N$ -dimensional tags, which are characterized with such a structure that they can be looked at as forming a discrete probability distribution. This last possibility shall be considered as an expected plausible outcome of QSM theory, due to the quantum-mechanical origin of all the QO tags employed so far.

##### 4.1. Stochastic transformation of quantum similarity matrices

$N$ -dimensional quantum SM columns or rows,  $\{\mathbf{z}_I\}$ , or, simply, their elements, are made, by construction and in computational practice, of positive definite rational numbers, although being theoretically real. This characteristic property can be resumed by saying that the set of the SM columns or rows, describing the QO point cloud, is a vector subset belonging to a *Vector Semispace* (VSS):  $\{\mathbf{z}_I\} \subset \mathcal{V}_N(\mathbb{R}^+)$ . VSS are vector spaces, which lack of reciprocal elements or negative scalars. See [10,11] and appendix A for more details. It must be noticed, for example, that the QO point cloud, defined by the DQOS tag set elements and commented a few lines ago in the previous section, defines a characteristic set of points in some  $N$ -dimensional VSS, because due to the nature of the QSM definition all their components are made by strictly positive numbers.

The VSS structure in general and, in particular, the construction of SM precludes that, in any case, the sum of every SM row (or column) elements is a positive real number, that is:

$$\forall I \quad \langle \mathbf{z}_I \rangle = \sum_J z_{IJ} \in \mathbb{R}^+.$$

This positive definiteness of the elements of any SM can be only assured by the appropriate definition and construction of the QSM. It must be stressed, as it has been done in all previous work on the subject, that in the QSM definition (1) and the following particular QSM construction, the active involved DF must be convex (see definition A.4 in

appendix A); that is, everywhere positive definite and integrable, and the present weight operators should be positive definite as well.

These above-commented row (or column) SM sums can be used as row (or column) scale factors in order to trivially obtain a new row (or column) set belonging to the same VSS, that is,  $\mathbf{s}_I = \langle \mathbf{z}_I \rangle^{-1} \mathbf{z}_I$ , but possessing afterwards the imposed form of a discrete probability distribution. That is, the following equalities can be easily written:

$$\langle \mathbf{s}_I \rangle = \langle \langle \mathbf{z}_I \rangle^{-1} \mathbf{z}_I \rangle = \langle \mathbf{z}_I \rangle^{-1} \langle \mathbf{z}_I \rangle = 1.$$

The set of  $N$  rows  $\mathbf{S} = \{\mathbf{s}_I\}$ , ordered forming a square ( $N \times N$ ) matrix  $\mathbf{S}$ , produce a non-symmetric *stochastic SM* [27] as a result.

A trivial compact way to create a *row stochastic SM* from any kind of quantum SM may be readily described by first constructing the diagonal matrix  $\mathbf{D}$ , whose elements are made by the sums of row (or column) SM elements:

$$\mathbf{D} = \text{Diag}(\langle \mathbf{z}_1 \rangle, \langle \mathbf{z}_2 \rangle, \dots, \langle \mathbf{z}_N \rangle), \quad (6)$$

and then producing the matrix product

$$\mathbf{S} = \mathbf{D}^{-1} \mathbf{Z}. \quad (7)$$

In the same manner, a *column stochastic SM* will be defined straightforwardly as the transpose of the previous definition:

$$\mathbf{S}^T = \mathbf{Z} \mathbf{D}^{-1},$$

where one must take into account the symmetrical structure of the original SM. The term *stochastic SM* (SSM) can be used as a generic reference to both SSM kinds, defined so far.

The row  $\{\langle \mathbf{s}_I | \}$  or column  $\{ | \mathbf{s}_I \rangle \}$  SSM partition vector sets, being associable to a collection of discrete probability distributions, may be even better connected to the DF tag set  $\{\rho_I\}$  of the original QOS, than the equivalent rows or columns of the attached SM  $\mathbf{Z}$ . In fact, they can be used as another alternative tag set part, which, when combined with the microscopic quantum systems belonging to the object set  $\{\omega_I\}$  of the original QOS, finally produce a new attached tagged set made of discrete QO, which can be called a *Discrete Stochastic QOS* (DSQOS).

For instance, taking into account the same considerations as those used before when previously discussing the nature of the SM rows, the connection between the original QOS elements with the SSM rows

$$(\omega_I; \rho_I) \leftrightarrow (\omega_I; \langle \mathbf{s}_I |) \quad \forall I$$

defines the elements of a DSQOS.

#### 4.2. Inward symmetrisation of stochastic quantum similarity matrices and stochastic quantum similarity indices

The elements of the DSQOS tag set can be directly used as QO discrete descriptors, admitting the actual implications

$$\forall I \quad \omega_I \leftrightarrow \rho_1 \leftrightarrow \langle \mathbf{s}_I |,$$

in a similar manner as it has been previously discussed, when dealing with the significance and the interpretation of the nature of the SM elements. The unique problem lies in the fact that, on the contrary to the SM  $\mathbf{Z}$ , the row (or column) SSM  $\mathbf{S}$  is not symmetric. However, this does not constitute a restrictive problem when alternative manipulations of the DSQOS are envisaged, and even has other possible uses, not contained in the symmetric QO descriptor structures.

##### 4.2.1. Partial order and SQSM

A *partial order* [1] over this kind of DSQOS can be easily associated to the characteristic mathematical structure of SSM. In this manner, the possibility to construct the non-symmetric SSM  $\mathbf{S}$  from a simple manipulation of the symmetric SM  $\mathbf{Z}$  completes QSM theory in a general and elegant way, providing QOS with a partial ordering formalism.

Because of the usual symmetric structure of the associated DQOS SM, the possibility of transforming DQOS into posets was still lacking in the quantum similarity theoretical structure. Thus, as a consequence of the discussion performed up to now, QOS can be transformed into DQOS, and even more easily into DSQOS: there a poset structure can be studied and employed, due to the non-symmetrical structure of the tag set SSM. The broad characteristic features of posets defined through QSM SSM will certainly be studied elsewhere. Here, some particular linear algebra aspects of the poset structure will be only considered.

##### 4.2.2. Inward matrix product symmetrisation

In addition to the classical symmetrisation techniques, which customarily use the sum or the classical product of the studied matrix and its transpose [28], there for the same purpose can be also described a simpler algorithm, involving an *inward matrix product* (IMP) [20,29,30].

An IMP between two known matrices,  $\mathbf{A}$  and  $\mathbf{B}$ , bearing the same arbitrary dimension, can be defined without problems as another matrix  $\mathbf{C}$ , with the same dimension form; see appendix B for more details. Using  $(n \times m)$  matrices as a typical but quite general example, the following straightforward algorithm can be designed for the IMP definition:

$$\mathbf{C} = \mathbf{A} * \mathbf{B} \quad \implies \quad \forall i, j \quad c_{ij} = a_{ij}b_{ij}. \quad (8)$$

IMP possesses a standard set of properties, which have already been described in detail and for this reason will not be repeated here; see [12,29,30] and also appendix B. Thus, it is trivial to realise that the commutative IMP

$$\mathbf{R} = \mathbf{S}^T * \mathbf{S} = \mathbf{S} * \mathbf{S}^T \quad (9)$$

of the stochastic matrix  $\mathbf{S}$  with its transpose  $\mathbf{S}^T$  produces a symmetric matrix as a result, that is,  $\mathbf{R} = \mathbf{R}^T$ .

#### 4.2.3. Stochastic similarity indices

The symmetric matrix  $\mathbf{R}$ , as defined above, could be used in the same way as the original SM,  $\mathbf{Z}$ . However, unit homogenisation with respect to the initial SM  $\mathbf{Z}$  makes preferable to employ the IMP square root (see appendix B.3 for more details), computed over the symmetric IMP as defined in equation (9), which is in turn readily constructed by the algorithm:

$$\mathbf{Q} = \mathbf{R}^{[1/2]} \implies \forall i, j \quad q_{ij} = \sqrt{r_{ij}}. \quad (10)$$

In Fortran 95 language, all the operations starting with the known SM  $\mathbf{Z}$  and leading to the computation of the matrix  $\mathbf{Q}$  can be immediately written in an extremely simple and short code. See [31] for instance and [20] for some pieces of code as illustrative examples. See also appendix B for more details on IMP powers and functions.

#### 4.2.4. Relationship between stochastic quantum similarity indices and Carbó index

One can also consider the new symmetric matrix  $\mathbf{Q}$  as holding *Stochastic Quantum Similarity Indices* (SQSI).

It is worthwhile to analyse, at this point, how the new kind of SQSI may be related to previous well-defined *quantum similarity indices* (QSI), which, as the so-called Carbó index [25,26], are present in the literature since a long time ago.

An expression of the elements of matrix  $\mathbf{Q}$ , as defined in equation (10), in terms of the original symmetric SM, leads to the equality sequence:

$$q_{ij} = \sqrt{r_{ij}} = \sqrt{s_{ij}s_{ji}} = \sqrt{\frac{z_{ij}z_{ji}}{\langle \mathbf{z}_i | \langle \mathbf{z}_j \rangle}} = \frac{z_{ij}}{\sqrt{\langle \mathbf{z}_i | \langle \mathbf{z}_j \rangle}}, \quad (11)$$

which resembles the Carbó similarity index [25,26], when observing its expression in continuous form, as can be studied below. The so-called Carbó similarity index is defined over the involved QO DF tag couples, using the associated QSM SM elements, and is written as

$$\kappa_{ij} = \frac{z_{ij}}{\sqrt{z_{ii}z_{jj}}} \quad (12)$$

This discussion does not complete at all the possible ways any QSM SM can be transformed into a QSI matrix. When discussing below in section 6 the possible transformations of the QQSAR fundamental equation, a related index to the SQSI one described in equation (11) will appear as a natural consequence of the stochastic scaling of the SM.

## 5. Stochastic quantum similarity matrices eigensystems

After defining the row and column SSM, as in the preceding paragraphs, for the purpose to study approximate solutions of the QQSAR fundamental equation (see, for example, [20,30] and appendix C), it appears to be very interesting to analyse the eigensystems of such matrices and their possible relationships. In fact, this subject constitutes the core of the present paper.

Suppose known the eigensystem equation of the row SSM as defined in equation (7), which can be written in the following way:

$$\mathbf{S}\mathbf{C} = \mathbf{C}\mathbf{\Sigma}, \quad (13)$$

where the diagonal matrix  $\mathbf{\Sigma}$  contains the eigenvalues of the row SSM  $\mathbf{S}$ . Because of the fact that the SSM is non-symmetric, then the eigenvector matrix  $\mathbf{C}$  is no longer orthogonal. However, equation (13) can be easily related to a symmetric eigensystem equation. Remembering the structure of the row SSM, the secular equation (13) can be also explicitly written as

$$\mathbf{D}^{-1}\mathbf{Z}\mathbf{C} = \mathbf{C}\mathbf{\Sigma}. \quad (14)$$

Now, employing the definite positive nature of the elements of the diagonal matrix

$$\mathbf{D} = \text{Diag}(\langle \mathbf{z}_I \rangle),$$

then it is easy to build up the square root real diagonal matrix

$$\mathbf{D}^{1/2} = \text{Diag}(\langle \mathbf{z}_I \rangle^{1/2}),$$

as well as its inverse

$$\mathbf{D}^{-1/2} = \text{Diag}(\langle \mathbf{z}_I \rangle^{-1/2}).$$

Both diagonal square roots can be used to transform equation (14) into the new secular equation

$$\mathbf{D}^{-1/2}\mathbf{Z}\mathbf{C} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{\Sigma},$$

which in turn can be written in a symmetric form as

$$\mathbf{D}^{-1/2}\mathbf{Z}\mathbf{D}^{-1/2}\mathbf{D}^{1/2}\mathbf{C} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{\Sigma}.$$

Thus, using the associative property of the matrix product, one can rewrite the previous equation as

$$\mathbf{Q}\mathbf{U} = \mathbf{U}\mathbf{\Sigma}, \quad (15)$$

where the following definitions have been used:

$$\mathbf{Q} = \mathbf{D}^{-1/2}\mathbf{Z}\mathbf{D}^{-1/2} \quad \text{and} \quad \mathbf{U} = \mathbf{D}^{1/2}\mathbf{C}.$$

However, by construction, the matrix  $\mathbf{Q}$  is symmetric, and consequently, the eigenvector matrix  $\mathbf{U}$  is always orthogonal. The matrix  $\mathbf{Q}$ , as defined for eigensystem equation



transformation purposes, becomes the same as the already defined symmetric matrix of equations (10) or (11). The eigensystem (15) is, thus, easily solvable by means of standard diagonalization methods [23,24]. Moreover, its eigenvalues are the same as those initially sought for in equations (13) or (14), while the initial row SSM eigenvectors are readily obtained from those of the symmetric matrix  $\mathbf{Q}$  forming the unitary matrix  $\mathbf{U}$ :

$$\mathbf{C} = \mathbf{D}^{-1/2}\mathbf{U}.$$

With respect to the column SSM  $\mathbf{S}^T$ , a similar eigensystem can be considered known:

$$\mathbf{S}^T\mathbf{X} = \mathbf{X}\mathbf{\Delta}, \quad (16)$$

where the diagonal matrix  $\mathbf{\Delta}$  has the ordered set of eigenvalues corresponding to the non-orthogonal eigenvector matrix  $\mathbf{X}$ . The equation (16) eigensystem can be now written, owing to the structure of the column SSM, as

$$\mathbf{ZD}^{-1}\mathbf{X} = \mathbf{X}\mathbf{\Delta} \quad (17)$$

and subsequently transformed into the equivalent symmetric structure as the one associated to equation (15). This can simply be done taking into account the existence of the diagonal square root matrices and writing first:

$$\mathbf{ZD}^{-1/2}\mathbf{D}^{-1/2}\mathbf{X} = \mathbf{X}\mathbf{\Delta},$$

followed by multiplying both sides on the left by the inverse square root diagonal matrix. Then, in this way, equation (15) is obtained again

$$\mathbf{D}^{-1/2}\mathbf{ZD}^{-1/2}\mathbf{D}^{-1/2}\mathbf{X} = \mathbf{D}^{-1/2}\mathbf{X}\mathbf{\Delta}.$$

However, while it is obviously evident that the eigenvalue matrix of the eigensystem (16) coincides with these of equations (14) and (15), that is,  $\mathbf{\Delta} = \mathbf{\Sigma}$ , the actual eigenvector matrix is related with the orthogonal matrix  $\mathbf{U}$  by means of the relationship

$$\mathbf{X} = \mathbf{D}^{1/2}\mathbf{U}.$$

So the row and column SSM possess the same eigenvalues, coinciding with those of the symmetric matrix  $\mathbf{Q}$ , and their eigenvectors are related through the square root of the diagonal stochastic transformation matrix  $\mathbf{D}$  to the orthogonal matrix  $\mathbf{U}$  made of the  $\mathbf{Q}$  matrix eigenvectors. Both non-orthogonal eigenvector matrices are related by a simple equation, owing to their construction from the orthogonal matrix  $\mathbf{U}$ :

$$\mathbf{X} = \mathbf{D}\mathbf{C}.$$

All the reported properties for the secular equations of SSM present very similar characteristics to the relationships found between the components of the secular equations of a matrix and its transpose, see, for instance, [23].

The properties of the secular equations attached to the two kinds of SSM can be employed to further study the structure of both matrices. Therefore, the spectral decomposition of the row SSM can be easily written in the following terms:

$$\mathbf{S} = \mathbf{C}\mathbf{\Sigma}\mathbf{C}^{-1} = \mathbf{D}^{-1/2}\mathbf{U}\mathbf{\Sigma}\mathbf{U}^T\mathbf{D}^{1/2} = \mathbf{D}^{-1/2}\mathbf{Q}\mathbf{D}^{1/2},$$

as well as the following decomposition is straightforwardly associated to the column SSM:

$$\mathbf{S}^T = \mathbf{X}\mathbf{\Sigma}\mathbf{X}^{-1} = \mathbf{D}^{1/2}\mathbf{U}\mathbf{\Sigma}\mathbf{U}^T\mathbf{D}^{-1/2} = \mathbf{D}^{1/2}\mathbf{Q}\mathbf{D}^{-1/2}.$$

It is shown in such a manner how the non-symmetric nature of the SSM does not interfere with the possible use of symmetric matrix diagonalization procedures.

## 6. QQSAR fundamental equation and stochastic quantum similarity matrices

It is easy to construct from the QSM framework the QQSAR fundamental equation, connecting any QOS SM with any QO property vector. Details can be found in [3–9,20,30] and appendix C. The QQSAR fundamental equation is conveniently written as a linear system:

$$\mathbf{Z}\mathbf{w} = \boldsymbol{\pi}, \quad (18)$$

where the SM  $\mathbf{Z}$  is connected with a QO, *experimental* property vector  $\boldsymbol{\pi}$ , by means of a vector  $\mathbf{w}$ , representing a discrete form of an *unknown* operator, see also appendix C for more details.

Equation (18) can be easily related with the corresponding SSM. Indeed, multiplying on the left by the stochastic diagonal matrix inverse transformation both sides of equation (18), it is obtained:

$$\mathbf{D}^{-1}\mathbf{Z}\mathbf{w} = \mathbf{D}^{-1}\boldsymbol{\pi} \implies \mathbf{S}\mathbf{w} = \mathbf{p} \quad \text{and} \quad \mathbf{p} = \mathbf{D}^{-1}\boldsymbol{\pi}.$$

In the same manner, the column stochastic matrix can intervene in equation (18), just using the simple matrix algebra:

$$\mathbf{Z}\mathbf{D}^{-1}\mathbf{D}\mathbf{w} = \boldsymbol{\pi} \implies \mathbf{S}^T\mathbf{v} = \boldsymbol{\pi} \quad \text{and} \quad \mathbf{v} = \mathbf{D}\mathbf{w}.$$

These two previous transformations can be employed at the same time, producing a new QQSAR fundamental equation, where a related matrix to the symmetric matrix  $\mathbf{Q}$  is the leading term:

$$\mathbf{D}^{-1}\mathbf{Z}\mathbf{D}^{-1}\mathbf{D}\mathbf{w} = \mathbf{D}^{-1}\boldsymbol{\pi} \implies \mathbf{A}\mathbf{v} = \mathbf{p} \quad \text{and} \quad \mathbf{A} = \mathbf{D}^{-1}\mathbf{Z}\mathbf{D}^{-1}.$$

While the matrix  $\mathbf{A}$ , described above, being symmetric can be processed in the usual manner to solve the QQSAR fundamental equation (see appendix D for additional details), the equations associated to the SSM  $\mathbf{S}$  and  $\mathbf{S}^T$  must be analyzed using spectral decompositions of the appropriate type, as previously discussed in section 5. For example, if the QQSAR fundamental equation

$$\mathbf{S}\mathbf{w} = \mathbf{p}$$

has to be solved, while preserving the intrinsic equation constraints inherent to the positive definite nature of their elements, as has been discussed in [20,30], the detailed procedure described in appendix D has to be followed. In this particular case, the adequate IMP decomposition of the QQSAR fundamental equation can be used:

$$\begin{aligned} \mathbf{S} &= \mathbf{R} * \mathbf{R} & \text{and} & & \mathbf{R} &= \mathbf{S}^{[1/2]}, \\ \mathbf{w} &= \mathbf{x} * \mathbf{x} & \text{and} & & \mathbf{x} &= \mathbf{w}^{[1/2]}, \\ \mathbf{p} &= \mathbf{q} * \mathbf{q} & \text{and} & & \mathbf{p} &= \mathbf{q}^{[1/2]}. \end{aligned}$$

So, an approximate solution of the fundamental QQSAR equation can be found solving the unrestricted equation, which can be now written as

$$\mathbf{R} * \mathbf{x} = \mathbf{q}.$$

The resulting approximate unrestricted equation can be solved using the procedure outlined in appendix D, just taking into account that the IMP square root of the SSM can be also written in terms of the SM, as it can be easily deduced that

$$\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{Z}^{[1/2]},$$

owing to the fact that

$$\mathbf{S} = \mathbf{D}^{-1} \mathbf{Z} \implies \forall i, j \quad s_{ij} = d_{ii}^{-1} z_{ij} \implies \forall i, j \quad \sqrt{s_{ij}} = d_{ii}^{-1/2} \sqrt{z_{ij}}.$$

In this manner, the role of the matrices  $\mathbf{D}^{-1}$  and  $\mathbf{Z}$  in the SSM  $\mathbf{S}$  eigensystem, as previously discussed, is substituted by the IMP square roots  $\mathbf{D}^{-1/2}$  and  $\mathbf{Z}^{[1/2]}$ , respectively. That is, defining the auxiliary matrix  $\mathbf{T}$  as

$$\mathbf{T} = \mathbf{D}^{-1/4} \mathbf{Z}^{[1/2]} \mathbf{D}^{1/4} = \mathbf{Q}^{[1/4]},$$

if the  $\mathbf{T}$  eigensystem is built up, using the secular equation

$$\mathbf{T} \mathbf{V} = \mathbf{V} \boldsymbol{\tau},$$

where  $\boldsymbol{\tau}$  is the diagonal eigenvalue matrix; then, the spectral decomposition of the matrix  $\mathbf{R}$  can be constructed by the non-symmetric matrix product

$$\mathbf{R} = \mathbf{D}^{-1/4} \mathbf{V} \boldsymbol{\tau} \mathbf{V}^T \mathbf{D}^{1/4}.$$

So, in this case, the procedure described in appendix D can be applied with minimal differences. The same can be said for the column SSM, with the logical changes in the procedure.

## 7. Conclusions

There has been described how from SM made from QSM, non-symmetric SSM can be constructed. The procedure for solving the QQSAR fundamental equation when the descriptor matrix is symmetric can be easily extended into the SSM non-symmetric case. This property appears from the fact that SSM eigensystems can be easily related to

symmetric secular equations, just by performing simple matrix algebra manipulations. In this manner row or column SSM appear as quite interesting candidates to be used in QQSAR problems in order to find causal relationships between the unbiased and universal QO descriptor structure contents and QO properties. Alternatively, the SSM columns or rows can be also employed as QO descriptors in classical QSAR procedures under principal component analysis statistical protocol.

### Acknowledgements

The author wants to acknowledge the Foundation M.F. de Roviralta as well as the European Commission contract #ENV4-CT97-0508, which have supported this work.

### Appendix A. Definitions related to quantum similarity measures

For more details on the following definitions and concepts, see [9,10].

**Definition A.1** (Tagged sets). Let us suppose known a given set, the object set  $\mathcal{S}$ , and another set, made of some chosen mathematical elements, which will be hereafter called tags, forming a tag set,  $\mathcal{T}$ . A *tagged set*  $\mathcal{Z}$  can be constructed by means of the ordered product  $\mathcal{Z} = \mathcal{S} \times \mathcal{T}$ :

$$\mathcal{Z} = \{\forall \theta \in \mathcal{Z} \mid \exists s \in \mathcal{S} \wedge \exists t \in \mathcal{T}: \theta = (s, t)\}.$$

**Definition A.2** (Quantum object, QO). A QO can be defined as an element of a tagged set: quantum systems in well-defined states are taken as the object set part and the corresponding DF constitute the tag set part.

**Definition A.3** (Vector semispace, VSS). A VSS over the positive definite real field  $\mathbb{R}^+$  is a vector space with the vector sum part provided by a structure of Abelian *semigroup*.

By an additive semigroup, an additive group without the presence of reciprocal elements is understood here. All VSS elements can be seen as directed towards the region of the positive axis hyperquadrant. It can be accepted if necessary that *null elements* are both included in the scalar field as well as in the VSS structure.

**Definition A.4** (First-order electronic DF, eDF). The first-order eDF form, as expressed within MO theory, can be defined by means of the linear combination:

$$\rho(\mathbf{r}) = \sum_i w_i |\varphi_i(\mathbf{r})|^2. \quad (\text{A.1})$$

This MO eDF can be written in a general way, as a double sum of products of function pairs, coupled with a set of matrix coefficients. See also appendix B.4 for more details.

However, a simple matrix diagonalization, followed by a unitary MO basis set transformation, can revert DF to the formal expression in equation (A.1). The coefficient set  $\mathbf{w} = \{w_i\} \subset \mathbb{R}^+$ , interpreted as MO occupation indices, corresponds to a collection of positive real numbers. A unit norm convention can be adopted:

$$\int |\varphi_i|^2 \mathbf{d}\mathbf{r} = 1 \quad \forall i \quad \implies \quad \int \rho(\mathbf{r}) \mathbf{d}\mathbf{r} = \sum_i w_i \int |\varphi_i|^2 \mathbf{d}\mathbf{r} = \sum_i w_i = 1,$$

and this results in considering the coefficient set  $\mathbf{w} = \{w_i\}$  as a discrete probability distribution.

**Definition A.5** (Convex conditions). By the term “convex conditions” applied to an  $n$ -dimensional vector it is understood:

$$\mathcal{K}_n(\mathbf{w}) \equiv \left\{ \mathbf{w} \in \mathcal{V}_n(\mathbb{R}^+) \wedge \sum_i w_i = 1 \right\}.$$

The set of the vector elements,  $\mathbf{w} = \{w_i\}$ , can be used instead in the convex conditions symbol, that is:

$$\mathcal{K}_n(\{w_i\}) \equiv \left\{ \forall i: w_i \in \mathbb{R}^+ \wedge \sum_i w_i = 1 \right\}.$$

The coefficients of a first-order eDF fulfill a set of convex conditions.

**Definition A.6** (General quantum similarity measures). A general QSM  $G(\Omega)$  can be considered a positive definite multiple scalar product defined by a contracted  $\nu$ -direct product of a QOS,  $\mathcal{T}$ :

$$G(\Omega) : \bigotimes_{K=1}^{\nu} \mathcal{T} \rightarrow \mathbb{R}^+.$$

This allows us to mix  $\nu$  DF  $\{\rho_I(\mathbf{r}), I = 1, \dots, \nu\}$  of the QOS with  $\omega$  PD operators, collected into a set  $\Omega = \{\Omega_K(\mathbf{r}), K = 1, \dots, \omega\}$ , belonging to the same VSS, for example:

$$G(\Omega) = \int \left[ \prod_{K=1}^{\omega} \Omega_K(\mathbf{r}) \right] \left[ \prod_{I=1}^{\nu} \rho_I(\mathbf{r}) \right] \mathbf{d}\mathbf{r},$$

where the coordinate vector  $\mathbf{r}$  shall be taken here as a general position vector.

## Appendix B. Inward matrix product: Definitions, properties and examples

Fortran 95 contains as a built-in feature [31] the Hadamard, Schur or, perhaps better to call it: *inward matrix product* (IMP). Such a matrix product is defined between two matrix structures, belonging to the same matrix space. As IMP has been previously used in several papers, related with quantum-chemical applications [20,29,30], it will be

presented here as a first step to describe its connection with DF structure and generalization. IMP is related to the DVS product structure, and as such can be seen as some sort of generalization of diagonal matrices product to general matrix VS. Curiously enough, IMP is scarcely referenced in the current literature, except on shy looking footnotes in a few books (see, e.g., [32,33] for the definitions and [34] for the origins).

IMP can be widely used in quantum mechanics for various purposes. In this section, some devoted pages are included, just to remark the simplicity of the concepts which can be built up around it and the interesting problems where IMP can be employed.<sup>2</sup>

IMP can be defined in a simpler way than the classical matrix product, as the associated structure mimics the matrix addition and the product of scalars. The next definition tries to provide a general form of this matrix operation.

**Definition B.1** (Inward matrix product, IMP). Consider any arbitrary hypermatrix space over a field  $M_{(\times \mathbf{n})}(\mathbf{R})$ . Let  $\mathbf{A}, \mathbf{B} \in M_{(\times \mathbf{n})}$ . An IMP involving the hypermatrix pair is a closed operation, resulting in a new hypermatrix  $\mathbf{P} \in M_{(\times \mathbf{n})}$ , and symbolized by  $\mathbf{P} = \mathbf{A} * \mathbf{B}$ , whose elements are defined by the algorithm:

$$\forall(\mathbf{i}) \quad p(\mathbf{i}) = a(\mathbf{i})b(\mathbf{i}).$$

Above, the elements of the involved hypermatrices are identified by means of an index vector  $(\mathbf{i}) \equiv (i_1; i_2; \dots; i_p)$ . Thus, one can consider that the hypermatrix space dimension is given by  $(\times \mathbf{n}) \equiv (n_1 \times n_2 \times \dots \times n_p)$ . The notation follows a previous one, employed when dealing with NSS structures (see, e.g., [35–38]).

Having presented the simple definition of IMP, then some of the most interesting properties and applications will be provided. IMP acts over hypermatrix spaces almost as if hypermatrices were treated as a product of scalars. This attractive feature can be employed in quantum chemistry computational problems, see, for example, [20,29,30], as well as in the development of new theoretical structures, which can present the peculiarity of being easily transferable to a high level programming language like Fortran 95.

### B.1. IMP properties

The following properties can be attached to the IMP defined over the elements of an arbitrary hypermatrix space  $M_{(\times \mathbf{n})}$ : IMP is distributive with respect to the matrix sum, as well as associative, and commutative [29].

The interest in defining such a matrix product appears from the possibility to attach to it the most usual features of a multiplicative composition rule. The following properties can be attached to the IMP.

<sup>2</sup> The term matrix or hypermatrix can be used indistinctly in reference to IMP manipulations.

Let be:  $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots \in \mathbf{M}_{(\times n)}$ . IMP defined over them are:

- (1) distributive with respect to matrix sum:  $\mathbf{A} * (\mathbf{B} + \mathbf{C}) = \mathbf{A} * \mathbf{B} + \mathbf{A} * \mathbf{C}$ ,
- (2) associative:  $\mathbf{A} * \mathbf{B} * \mathbf{C} = \mathbf{A} * (\mathbf{B} * \mathbf{C}) = (\mathbf{A} * \mathbf{B}) * \mathbf{C}$ ,
- (3) commutative:  $\mathbf{A} * \mathbf{B} = \mathbf{B} * \mathbf{A}$ .

From the inspection of these properties, it is easy to see that IMP can operate over matrix spaces in the same way as in *diagonal matrix vector spaces* (DVS) the classical matrix product acts.

### B.2. IMP unit element and inverse

Also, an inward unit element exists, which can be called the *unity* matrix,  $\mathbf{1} \in \mathbf{M}_{(\times n)}$ , such that  $\mathbf{1} * \mathbf{A} = \mathbf{A} * \mathbf{1} = \mathbf{A}$ . Using the real multiplication unit it can be defined as  $\mathbf{1} = \{1(\mathbf{i}) = 1 \forall(\mathbf{i})\}$ . The existence of an IMP inverse is subject to the following important limitations imposed by the following definition:

**Definition B.2** (IMP invertible matrices). Iff  $\mathbf{A} = \{a(\mathbf{i})\} \wedge \forall(\mathbf{i}): a(\mathbf{i}) \neq 0$  then  $\mathbf{A}$  can be called *inwardly invertible* or *regular*. A new matrix defines the IMP inverse of a matrix  $\mathbf{A}$ :  $\mathbf{A}^{[-1]} = \{a^{[-1]}(\mathbf{i})\}$ , with elements, which are computed as follows:  $\forall(\mathbf{i}) a^{[-1]}(\mathbf{i}) = (a(\mathbf{i}))^{-1}$ . This definition produces the sequence of equalities  $\mathbf{A} * \mathbf{A}^{[-1]} = \mathbf{A}^{[-1]} * \mathbf{A} = \mathbf{1}$ .

These IMP properties are sufficient to define a commutative algebra over any matrix vector space. One can refer to this kind of algebra as *Hadamard* or *Schur algebra*.

### B.3. IMP powers and functions

IMP powers of a given hypermatrix  $\mathbf{A}$  are readily defined as  $\mathbf{A}^{[p]} = \{a(\mathbf{i}, \mathbf{j})^p\}$ . The square bracket enveloping the exponent is used here to distinguish an IMP power from the one defined involving classical products, as it has been previously used in the IMP inverse definition. For example, whenever  $\mathbf{Z} = \mathbf{A} * \mathbf{A}$ , then the matrix  $\mathbf{A}$  can be also considered as the IMP square root of  $\mathbf{Z}$ :

$$\mathbf{A} = \mathbf{Z}^{[1/2]} \implies \forall \mathbf{i}, \mathbf{j} \quad a(\mathbf{i}, \mathbf{j}) = \sqrt{z(\mathbf{i}, \mathbf{j})}.$$

IMP functions of a given hypermatrix are also easy to define:  $\phi[\mathbf{Z}] = \{\phi(z(\mathbf{i}, \mathbf{j}))\}$ . IMP algebra is tightly related to diagonal matrix computational algorithms and the above definitions are the consequence of another shared isomorphic characteristic between DVS and VS associated to an IMP.

#### B.4. A quantum-chemical example

Under LCAO MO approach, the first-order density function (A.1) can be expressed as a double sum over the AO basis set labels:

$$\rho(\mathbf{r}) = \sum_{\mu} \sum_{\nu} P_{\mu\nu} |\mu\rangle \langle \nu|,$$

where  $\{P_{\mu\nu}\}$  are the elements of the charges–bond orders matrix and  $\{|\mu\rangle = \chi_{\mu}(\mathbf{r})\}$  represents the AO basis set. Both matrix elements can be cast into a pair of arrays as

$$\mathbf{P} = \{P_{\mu\nu}\} \quad \text{and} \quad \mathbf{X}(\mathbf{r}) = \{X_{\mu\nu}(\mathbf{r}) = |\mu\rangle \langle \nu|\}.$$

Then, a new hybrid matrix can be expressed as an IMP:

$$\mathbf{M}(\mathbf{r}) = \mathbf{P} * \mathbf{X}(\mathbf{r}).$$

This allows obtaining the first-order DF simply as a sum involving all matrix elements [35–38]:

$$\rho(\mathbf{r}) = \langle \mathbf{M}(\mathbf{r}) \rangle = \sum_{\mu} \sum_{\nu} M_{\mu\nu}(\mathbf{r}).$$

### Appendix C. Expectation values and fundamental QQSAR equation

When studying a quantum-mechanical system, the expectation value of some observable  $\omega$ , in a well-defined QO system state, can be written by means of the integral

$$\langle \omega \rangle = \int W(\mathbf{r}) \rho(\mathbf{r}) \, d\mathbf{r}, \quad (\text{C.1})$$

where  $W(\mathbf{r})$  is an associated Hermitian operator to be determined for a given QOS, and  $\rho(\mathbf{r})$  the system state density function.

On the other hand, the expression (C.1), from a QQSAR point of view, can be interpreted as a scalar product [3–9,20,30,39], that is,

$$\langle \omega \rangle = \langle W | \rho \rangle. \quad (\text{C.2})$$

Taking into account the unknown nature of the QQSAR operator  $W(\mathbf{r})$ , one can consider that it can be decomposed as a product of the operator leading to the expectation value  $W_{\omega}(\mathbf{r})$ , still unknown and to be determined, by a known PD weight operator,  $\Omega(\mathbf{r}, \mathbf{r}_0)$ . Thus, the expectation value as presented in expression (C.1) can be written now as

$$\langle \omega \rangle = \langle W_{\omega} | \Omega | \rho \rangle. \quad (\text{C.3})$$

This is the same as to transform equation (C.1) into the equivalent more general integral

$$\langle \omega \rangle = \iint W_{\omega}(\mathbf{r}) \Omega(\mathbf{r}, \mathbf{r}_0) \rho(\mathbf{r}_0) \, d\mathbf{r} \, d\mathbf{r}_0. \quad (\text{C.4})$$



It is easy to see that the PD weight operator  $\Omega(\mathbf{r}, \mathbf{r}_0)$  can particularly be chosen as the Dirac's delta function  $\delta(\mathbf{r} - \mathbf{r}_0)$ , and, thus, in doing so, equation (C.1) is recovered. In order to distinguish the expectation value general definition, as presented in equation (C.4), from the usual choice in equation (C.1), where the weight can be considered a unit operator, the general integral (C.4) can be named as a *weighted expectation value* integral expression. It is interesting to note that expression (C.4) has a formal structure highly resembling the two-system QSM integral as given in equation (1).

It is only necessary to express the unknown operator of equation (C.4) in terms of a DF basis set to arrive to the fundamental QQSAR equation. Now, whenever both operator and DF in the integral (C.1) can be considered to belong to the same VSS, and expressing approximately the unknown operator,  $W_\omega(\mathbf{r})$ , as a linear combination of the QOS tag set, acting as a basis,

$$W_\omega(\mathbf{r}) \approx \sum_I w_I \rho_I(\mathbf{r}), \quad (\text{C.5})$$

a new relationship is obtained after substituting expression (C.5) in equation (C.4), while taking into account the operator decomposition and the explicit expression of a QO  $A$ :

$$\langle \omega_A \rangle = \sum_I w_I \iint \rho_I(\mathbf{r}) \Omega(\mathbf{r}, \mathbf{r}_0) \rho_A(\mathbf{r}_0) \, d\mathbf{r} \, d\mathbf{r}_0 = \sum_I w_I z_{IA}(\Omega). \quad (\text{C.6})$$

Admitting equation (C.6) holds for every QO in a given QOS, and that the SM  $\mathbf{Z}$ , obtained employing the weight operator  $\Omega(\mathbf{r}, \mathbf{r}_0)$ , is symmetric, then an equivalent equation can be written in matrix form as

$$\mathbf{Z}\mathbf{w} = \boldsymbol{\pi}, \quad (\text{C.7})$$

where  $\boldsymbol{\pi}$  is a column vector containing the expectation values or the QO property of interest and  $\mathbf{w}$  is another column vector, containing the unknown coefficient set  $\{w_I\}$  of the operator expression (C.5). At the light of all the discussed aspects of the problem, it seems that equation (C.7) can be named from now on the *fundamental QQSAR equation*.

#### Appendix D. Quantum QSAR modeling

The procedure used here for dealing with the QQSAR model is based on IMP and the PD nature of the quantum SM or SI, which can be employed to construct it. The possibility to transform the experimental activity data into a PD vector is also taken into account<sup>3</sup>. The whole problem has been studied in detail from various points of view [3–9,14–17,40–45], so it will be just outlined here.

As discussed in the previous section, the QQSAR models can be written as a matrix equation like expression (C.7), involving the chosen SM and the QOS experimental data

<sup>3</sup> Many QSAR properties possess intrinsically PD values or can be scaled as well as origin shifted to fulfill this property.

to be described. All known matrices belong to some VSS of the appropriate dimension. In order to keep the model describing PD property values, the solution could be forced to be an element of some VSS too. So, using the symbol  $\mathbf{A} >^* 0$  to indicate that a chosen matrix  $\mathbf{A}$  has all its values defined in  $\mathbb{R}^+$ , that is, to represent in a shorter form the property  $\mathbf{A} = \{a_{ij}\} \Rightarrow \forall i, j a_{ij} \in \mathbb{R}^+$ ; then the constrained QQSAR model can be written

$$\mathbf{Z} >^* 0 \text{ and } \boldsymbol{\pi} >^* 0 : \quad \mathbf{Z}\mathbf{w} = \boldsymbol{\pi} \quad \Longrightarrow \quad \mathbf{w} >^* 0. \quad (\text{D.1})$$

#### D.1. Fundamental QQSAR equation approximate solution

In order to take into account the implication, which can be considered as a constraint, imposed to the solutions to be found when solving the equation (D.1) linear system, an algorithm can be easily designed as follows.

It is obvious that if the PD restriction on the vector  $\mathbf{w}$  holds, there shall exist three real matrices,  $\mathbf{T}$ ,  $\mathbf{x}$  and  $\mathbf{p}$ , which can be computed as the IMP square root of the system matrices, that is,  $\mathbf{T} = \mathbf{Z}^{[1/2]}$ ,  $\mathbf{x} = \mathbf{w}^{[1/2]}$ ,  $\mathbf{p} = \boldsymbol{\pi}^{[1/2]}$ , in the same fashion as this IMP matrix power was defined in appendix B.3. So, equation (D.1), bearing the corresponding constraint on the unknowns' vector, can be written as

$$(\mathbf{T} * \mathbf{T})(\mathbf{x} * \mathbf{x}) = (\mathbf{p} * \mathbf{p}). \quad (\text{D.2})$$

This suggests the possibility to construct an alternative approximate system. Indeed, just exchanging the sites of classical and inward matrix products on the left of equation (D.2) it is obtained:

$$(\mathbf{T}\mathbf{x}) * (\mathbf{T}\mathbf{x}) = (\mathbf{p} * \mathbf{p}), \quad (\text{D.3})$$

leading to the approximate reduced system

$$\mathbf{T}\mathbf{x} = \mathbf{p}, \quad (\text{D.4})$$

which does not need any restriction on the unknown vector elements and, thus, can be finally solved, considering the approximate nature of the solution yielding the QQSAR model.

This way, chosen here in order to achieve an approximate solution of the fundamental QQSAR equation, takes the following path. First, the eigensystem of the symmetric matrix  $\mathbf{T}$  is obtained, and an approximate spectral decomposition of this matrix is used employing a cutoff value  $\varepsilon$  on the eigenvalues, in order to get rid of numerical noise. This can be written explicitly using a logical Kronecker's delta [35–38] as

$$\mathbf{T} \approx \sum_i \delta(\tau_i > \varepsilon) \tau_i |i\rangle \langle i|, \quad (\text{D.5})$$

where  $\{\tau_i\}$  is the spectrum and  $\{|i\rangle\}$  the eigenvector system of the matrix  $\mathbf{T}$ . Then, the same can be done for the evaluation of an approximate inverse:

$$\mathbf{T}^{-1} \approx \sum_i \delta(\tau_i > \varepsilon) \tau_i^{-1} |i\rangle \langle i|. \quad (\text{D.6})$$

Finally, in this manner the approximate solution can be written using

$$\mathbf{x} \approx \mathbf{T}^{-1} \mathbf{p} = \sum_i \delta(\tau_i > \varepsilon) \tau_i^{-1} |i\rangle \langle i | \mathbf{p}\rangle, \quad (\text{D.7})$$

so, the approximate solution of the original system (D.1),  ${}_a\mathbf{w}$ , with the adequate constraint, is simply computed using the inward product  ${}_a\mathbf{w} = \mathbf{x} * \mathbf{x}$ . This is sufficient to obtain a set of estimated property values:

$${}_a\boldsymbol{\pi} = \mathbf{Z}({}_a\mathbf{w}). \quad (\text{D.8})$$

As discussed before, this algorithm can be employed over any constrained linear system problem, so the matrix  $\mathbf{Z}$  may be used directly, as well as any symmetric manipulation of the stochastic transformations, like the one present in equation (9).

## References

- [1] J. Klein and D. Babić, *J. Chem. Inf. Comput. Sci.* 37 (1997) 656.
- [2] P.G. Mezey, *Shape in Chemistry* (VCH, New York, 1993).
- [3] R. Carbó, E. Besalú, L. Amat and X. Fradera, *J. Math. Chem.* 18 (1995) 237.
- [4] D. Robert and R. Carbó-Dorca, *J. Chem. Inf. Comput. Sci.* 38 (1998) 469.
- [5] L. Amat, D. Robert, E. Besalú and R. Carbó-Dorca, *J. Chem. Inf. Comput. Sci.* 38 (1998) 624.
- [6] D. Robert and R. Carbó-Dorca, *J. Chem. Inf. Comput. Sci.* 38 (1998) 620.
- [7] D. Robert, L. Amat and R. Carbó-Dorca, *J. Chem. Inf. Comput. Sci.* 39 (1999) 333.
- [8] X. Gironés, L. Amat and R. Carbó-Dorca, *SAR and QSAR in Environmental Research* 10 (1999) 545.
- [9] R. Carbó-Dorca and E. Besalú, *J. Mol. Struct. (Theochem)* 451 (1998) 11.
- [10] R. Carbó-Dorca, Fuzzy sets and Boolean tagged sets, vector semispaces and convex sets, quantum similarity measures and ASA density functions, diagonal vector spaces and quantum chemistry, in: *Advances in Molecular Similarity*, Vol. 2, eds. R. Carbó-Dorca and P.G. Mezey (JAI Press, London, 1988) p. 43.
- [11] R. Carbó-Dorca, *J. Math. Chem.* 23 (1998) 353.
- [12] R. Carbó-Dorca, L. Amat, E. Besalú, X. Gironés and D. Robert, Quantum molecular similarity: Theory and applications to the evaluation of molecular properties, biological activities and toxicity, Technical report IT-IQC-99-16, Institute of Computational Chemistry, in: *Proceedings of the IV Girona Seminar on Molecular Similarity* (Kluwer Academic, Dordrecht) (to be published).
- [13] R. Carbó-Dorca, *J. Math. Chem.* 22 (1997) 143.
- [14] R. Carbó-Dorca, E. Besalú, L. Amat and X. Fradera, Quantum molecular similarity measures: Concepts, definitions and applications to QSAR, in: *Advances in Molecular Similarity*, Vol. 1, eds. R. Carbó-Dorca and P.G. Mezey (JAI Press, London, 1996) p. 1.
- [15] R. Carbó-Dorca, L. Amat, E. Besalú and M. Lobato, Quantum molecular similarity, in: *Advances in Molecular Similarity*, Vol. 2, eds. R. Carbó-Dorca and P.G. Mezey (JAI Press, London, 1998) p. 1.
- [16] R. Carbó and E. Besalú, Theoretical foundations of quantum similarity, in: *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, ed. R. Carbó (Kluwer Academic, Amsterdam, 1995) p. 3.
- [17] R. Carbó-Dorca, L. Amat, E. Besalú, X. Gironés, D. Robert, Quantum mechanical origin of QSAR: Theory and applications, Technical report IT-IQC-99-1, Institute of Computational Chemistry, Computational Medicinal Chemistry, special issue, *J. Mol. Struct. (Theochem)* (to appear).
- [18] R. Carbó, B. Calabuig, L. Vera and E. Besalú, *Adv. Quantum Chem.* 25 (1994) 253.
- [19] R. Carbó and B. Calabuig, *J. Chem. Inf. Comp. Sci.* 32 (1992) 600.

- [20] R. Carbó-Dorca, Stochastic transformation of quantum similarity matrices and their use in quantum QSAR (QQSAR) models, Technical report IT-IQC-99-18, Institute of Computational Chemistry, Int. J. Quant. Chem. (1999) (in press).
- [21] P.H.A. Sneath and R.R. Sokal, *Numerical Taxonomy* (W.H. Freeman, San Francisco, 1973).
- [22] W.J. Krzanowski (ed.), *Recent Advances in Descriptive Multivariate Analysis*, Oxford Science Publications (Clarendon Press, Oxford, 1995).
- [23] E. Durand, *Solutions numériques des équations algébriques*, Vol. II (Masson, Paris, 1961).
- [24] J.H. Wilkinson and C. Reinsch, *Linear Algebra* (Springer, Berlin, 1971).
- [25] R. Carbó, M. Arnau and L. Leyda, Int. J. Quant. Chem. 17 (1980) 1185.
- [26] R. Carbó and B. Calabuig, Int. J. Quant. Chem. 42 (1992) 1681.
- [27] I.M. Vinogradov (ed.), *Encyclopaedia of Mathematics*, Vol. 9 (Kluwer Academic, Dordrecht, 1993) p. 9.
- [28] R. Carbó and L. Domingo, *Álgebra Matricial y Lineal*, Serie Schaum (McGraw-Hill, Madrid, 1987).
- [29] K.D. Sen and R. Carbó-Dorca, J. Mol. Struct. (Theochem) 501 (2000) 173.
- [30] R. Carbó-Dorca, Inward matrix products: Extensions and applications to quantum mechanical foundations of QSAR, Technical report IT-IQC-99-15, Institute of Computational Chemistry, J. Mol. Struct. (Theochem) (submitted).
- [31] Lahey/Fujitsu Fortran 95, Version 5.5, Lahey Computer Systems, Incline Village (Nevada) (1999).
- [32] C. Roos, T. Terlaky and J.-P. Vial, *Theory and Algorithms for Linear Optimization* (Wiley, New York, 1997).
- [33] R.A. Horn and Ch.A. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).
- [34] I.M. Vinogradov (ed.), *Encyclopaedia of Mathematics*, Vol. 4 (Kluwer Academic, Dordrecht, 1989).
- [35] R. Carbó and E. Besalú, J. Math. Chem. 13 (1993) 331–342.
- [36] R. Carbó and E. Besalú, Computers Chem. 18 (1994) 117–126.
- [37] R. Carbó and E. Besalú, J. Math. Chem. 18 (1995) 37–72.
- [38] R. Carbó and E. Besalú, in: *Strategies and Applications in Quantum Mechanics*, eds. Y. Ellinger and M. Defranceschi (Kluwer Academic, Dordrecht, 1996) pp. 229–248.
- [39] R. Carbó-Dorca, E. Besalú and X. Gironés, Extended density functions, Technical report IT-IQC-99-2, Institute of Computational Chemistry Adv. Quantum Chem. (in press).
- [40] R. Carbó, E. Besalú, L. Amat and X. Fradera, J. Math. Chem. 19 (1996) 47.
- [41] L. Amat, R. Carbó-Dorca and R. Ponec, J. Comp. Chem. 19 (1998) 1575.
- [42] R. Ponec, L. Amat and R. Carbó-Dorca, J. Comput.-Aided Mol. Design 13 (1999) 259.
- [43] R. Ponec, L. Amat and R. Carbó-Dorca, J. Phys. Org. Chem. 12 (1999) 447.
- [44] A. Gallegos, D. Robert and R. Carbó-Dorca, Structure–toxicity relationships of polycyclic aromatic hydrocarbons using molecular quantum similarity, Technical report IT-IQC-00-03, Institute of Computational Chemistry.
- [45] D. Robert, L. Amat and R. Carbó-Dorca, Quantum similarity QSAR: Study of inhibitors binding to thrombin, trypsin and factor Xa, including a comparison with CoMFA and CoMSIA methods, in: *Proceedings of the Third Congress of the Int. Soc. for Theoret. Chem. Phys.*, Mexico (1999), Int. J. Quant. Chem. (submitted).
- [46] J.H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon Press, Oxford, 1965).