

The Significance of Parameters in Charge Equilibration Models

T. Verstraelen,^{*†} P. Bultinck,[‡] V. Van Speybroeck,[†] P. W. Ayers,[§] D. Van Neck,[†] and M. Waroquier[†]

[†]Center for Molecular Modeling, QCMM Alliance Ghent-Brussels, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium

[‡]Department of Inorganic and Physical Chemistry, QCMM Alliance Ghent-Brussels, Ghent University, Krijgslaan 281 (S-3), B-9000 Gent, Belgium

[§]Department of Chemistry, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada

 Supporting Information

ABSTRACT: Charge equilibration models such as the electronegativity equalization method (EEM) and the split charge equilibration (SQE) are extensively used in the literature for the efficient computation of accurate atomic charges in molecules. However, there is no consensus on a generic set of optimal parameters, even when one only considers parameters calibrated against atomic charges in organic molecules. In this work, the origin of the disagreement in the parameters is investigated by comparing and analyzing six sets of parameters based on two sets of molecules and three calibration procedures. The resulting statistical analysis clearly indicates that the conventional least-squares cost function based solely on atomic charges is in general ill-conditioned and not capable of fixing all parameters in a charge-equilibration model. Methodological guidelines are formulated to improve the stability of the parameters. Although in this case a simple interpretation of individual parameters is not possible, charge equilibration models remain of great practical use for the computation of atomic charges.

I. INTRODUCTION

A thorough understanding and efficient computation of molecular interactions is of fundamental importance in computational chemistry.¹ Electrostatic interactions are typically the strongest intermolecular interactions and therefore play a dominant role in many important processes in extended molecular systems such as drug-receptor interactions in proteins,² host–guest interactions in zeolites,^{3–5} reactions in solvents,^{6–8} and so on.

Molecular electrostatics can to a large extent be rationalized in terms of the molecular charge distribution condensed into the atoms that compose the molecule, i.e., through the assignment of effective atomic charges. Although the concept of charges on the atoms in the molecule (AIM) is evident for most chemists, it is far less obvious to define them in a quantum-mechanical context.^{9–11} AIM charges can be computed in a multitude of ways. One group of methods relies on the introduction of AIM subspaces within the molecule. Many such density or wave function partitioning schemes were proposed over the past 60 years to divide molecules into atomic parts, e.g., Mulliken population analysis (MPA),¹² Lowdin population analysis,¹³ Hirshfeld partitioning,¹⁴ natural population analysis (NPA),¹⁵ Bader's AIM scheme,¹⁶ Hirshfeld-I partitioning,¹⁷ and so on. Besides this class of methods, atomic charges can also be fitted to reproduce the electrostatic potential around a molecule,¹⁸ with for example the Merz–Kollman¹⁹ or RESP scheme.²⁰ Nevertheless, a quantum mechanically unique definition of the atom in the molecule, and thus the atomic charge, remains problematic.⁹ At best, one can test and assess a population scheme using as criteria its basis-set and geometry independence, and its accuracy with respect to the molecular electrostatic potential. Benchmark studies showed that the Hirshfeld-I¹⁷ scheme performs well in this regard.^{21–23} Despite its attractive features, both from the application and information

theoretical viewpoints, it is not feasible to compute Hirshfeld-I charges on extended systems with many thousands of atoms. The prerequisite for any type of charge population scheme is the computation of the total electron density of the molecular system with a reliable electronic structure method, which becomes prohibitive for large systems, or even for smaller systems when the molecular environment is included explicitly.

An alternative approach for the computation of effective atomic charges is through charge equilibration models.²⁴ These models are derived from rigorous density functional theory but are computationally much less demanding than full electronic structure computations. As a consequence, they are sometimes considered a kind of semiempirical density functional theory.²⁵ Charge equilibration models allow an efficient computation of the charge distribution in large molecular systems, while the corresponding electronic structure computation requires an enormous computational effort. This advantage of charge equilibration models enables many attractive applications, including high-throughput screening of drug candidates due to the possibility of computing diverse molecular properties very quickly.^{26,27} Moreover, charge equilibration models are a cornerstone in advanced molecular mechanics developments such as polarizable force fields^{28–30} and reactive force fields.³¹

Charge equilibration models have an intrinsic weakness: one needs to determine many model parameters with ad hoc calibration methods before one can actually apply charge equilibration models. Optimally, one could hope that the calibrated parameters have a clear physical interpretation, i.e., that one can find an ultimate set of parameters that both satisfy intuitive

Received: January 4, 2011

Published: May 02, 2011

Table 1. Non-Exhaustive Overview of EEM Parameters (in eV) Reported in the Literature^a

reference		H	C	N	O	F	S	Cl	Br
Mortier and Baekelandt (a): HF/STO-3G MPA	χ	4.4	5.7	10.6	8.5				
	η	27.6	18.2	26.4	22.2				
Rappe (b)	χ	4.5280	5.343	6.899	8.741	10.874	6.928	8.564	7.790
	η	13.8904	10.126	11.760	13.364	14.948	8.972	9.892	8.850
Menegon (c): PM3 CM1	χ	3.885	4.656	7.175	10.075				
	η	18.404	14.264	13.452	20.690				
Bultinck (d): B3LYP/6-31G* MPA	χ	1.0	5.25	8.80	14.75	15.0			
	η	35.9	18.0	18.78	28.68	39.54			
Bultinck (d): B3LYP/6-31G* NPA	χ	1.0	8.49	13.45	27.06	39.18			
	η	38.88	18.3	20.92	39.26	88.2			
Verstraelen (e): METS, MP2/Aug-cc-pVDZ MPA	χ	9.73	6.94	10.42	25.59	29.26	5.68	5.99	-3.84
	η	19.50	12.65	19.85	40.05	39.91	9.87	32.08	33.64
Verstraelen (e): NETS, MP2/Aug-cc-pVDZ NPA	χ	1.81	4.65	6.26	8.65	17.79	3.54	3.45	1.15
	η	19.36	11.47	12.18	14.92	39.32	8.24	33.09	32.77
Verstraelen (e): HETS, MP2/Aug-cc-pVDZ Hirshfeld-I	χ	6.37	7.37	8.03	9.23	16.31	6.98	10.19	8.11
	η	15.54	10.74	10.93	12.85	38.37	7.86	31.89	31.94

^a (a) Refs 32, 39. (b) Ref 35. (c) Ref 40. (d) Refs 26, 37, 38. (e) Ref 36. The acronyms METS, NETS, and HETS refer to calibrations of the EEM model as discussed in this work, using Mulliken, natural, and Hirshfeld-I charges, respectively.

expectations and reproduce a set of reference charges used for the calibration.

Two important charge equilibration models have been proposed in the literature: the electronegativity equalization method³² (EEM) and the split charge equilibration (SQE)³³ model. All of the model parameters in these models can be given a physical meaning. Since charge equilibration is fundamentally rooted in density functional theory, and the model parameters are closely related to properties of atoms in a vacuum³² and electrostatic properties of extended systems,³⁴ one would expect that the corresponding calibrated parameters can be given a physical interpretation and are unique. Unfortunately, the EEM parameter sets reported in the literature show large fluctuations. A nonexhaustive overview of EEM parameters^{32,35–40} calibrated for organic systems is given in Table 1. Even the electronegativity and hardness parameters of carbon are scattered over ranges of 4.65–8.49 eV and 10.13–18.30 eV, respectively. As shown below, part of the differences may be due to the use of different electrostatic interaction models, e.g., point charge versus Gaussian charge. This tabulation raises the concern that EEM parameters may not have any physical interpretation, simply because no consensus can be extracted from the sets of EEM parameters in the literature. This is also problematic from the practical perspective: which set of parameters is trustworthy for computational applications? A thorough analysis of the spread on the parameters and its origin is required to re-establish the credibility of charge equilibration models.

The primary objective of this work is therefore an in-depth investigation of the reasons underlying the spread in the reported EEM parameters, and a similar analysis for the SQE model. As a side effect, many practical guidelines will be formulated to improve the quality of calibrated charge equilibration parameters.

The paper is structured as follows: The EEM and SQE models are introduced in section II. The next section, section III, first reviews the conventional parameter calibration methods, which are also used in this work to investigate the spread in the calibrated parameters. Then, the various analysis techniques that

are introduced to understand the origin of the fluctuations in the parameters are highlighted. In section IV, we discuss the results of the parameter calibrations and analysis techniques. Finally, section V summarizes the findings of this work and formulates some guidelines for future EEM and SQE calibrations.

II. THEORETICAL CHARGE EQUILIBRATION MODELS

EEM. The EEM model describes charge equilibration in molecular systems based on the effective atomic charges, their electrostatic interaction, and additional local energy terms. It is basically a second order expansion of the molecular energy in terms of atomic partial charges and has the following general mathematical form:

$$E_{\text{EEM}} = v^T q + \frac{1}{2} q^T J q \quad (1)$$

In this expression, q is a column vector of length N containing the effective atomic charges, v is a column vector of length N containing the first-order coefficients, and J is an $N \times N$ a matrix containing the second-order coefficients.

The electronegativity equalization method³² (EEM) postulates specific mathematical expressions for the elements J and v based on density functional theory. The diagonal elements of matrix J are the atomic hardness parameters, η_i . The off-diagonal elements model the electrostatic interactions between atom pairs. In the original work of Mortier et al.,³² a simple point charge model is used, leading to $J_{ij} = 1/r_{ij}$ for the off-diagonal elements, where r_{ij} is the distance between atoms i and j . (Atomic units are used throughout.) Smeared-charge electrostatic interactions were introduced in subsequent papers by various authors to account for the finite size of the atomic charge distribution.^{28,35,36,41,42} A particular advantage of the smeared-charge potentials is that one can introduce lower bounds for the hardness parameters to guarantee a positive definite matrix J .^{36,43} The elements in the vector v are the atomic electronegativity parameters, χ_i . Although the electronegativity and hardness parameters should correspond approximately to their gas-phase

values,^{32,44} they are in practice calibrated to quantum mechanical reference data (e.g., atomic charges from some AIM method or the molecular electrostatic potential from a quantum chemical calculation) for a large training set of molecules,^{26,36–38,45–48} as further discussed in section III.

The equilibrium charge distribution is the vector of atomic charges that minimizes the energy in eq 1 under a total charge constraint. The conventional approach is to introduce a Lagrange multiplier, λ , as follows:

$$\tilde{E}_{\text{EEM}} = v^T q + \frac{1}{2} q^T J q - \lambda(d^T q - Q_{\text{tot}}) \quad (2)$$

where d is a column vector with N elements, all equal to one. The equations for the equilibrium charges require that the derivatives of \tilde{E}_{EEM} toward the partial charges are zero and that the total charge constraint is satisfied, which leads to a set of linear equations. These equations are conveniently written in block matrix notation:

$$\begin{bmatrix} J & -d \\ -d^T & 0 \end{bmatrix} \begin{bmatrix} q \\ \lambda \end{bmatrix} = -\begin{bmatrix} v \\ Q_{\text{tot}} \end{bmatrix} \quad (3)$$

where dotted lines do not only separate block matrices but also vectors, as in $[q : \lambda]^T$. These equations reveal that the electronegativity of all atoms, $\partial E_{\text{EEM}} / \partial q_p$, must be equal to the Lagrange multiplier. Hence, one often uses the term *equalized electronegativity* or *molecular electronegativity* for the Lagrange multiplier, and one refers to this method as the *electronegativity equalization method*. In the remainder of the text, we will use the symbol χ_{mol} instead of λ .

It is clear that the charges minimizing the energy \tilde{E}_{EEM} in eq 2 are invariant when changing all electronegativity parameters by the same amount: $\chi_i \rightarrow \chi_i + \alpha$. Therefore, the individual χ_i 's are only determined up to an additive constant when only reference charges are used for the calibration. The following extension of the EEM model involving a charge bath is used in the calibration process to avoid serious deviations from the physical atomic electronegativities. A molecule is allowed to exchange charge with a charge bath that has an electronegativity, χ_{bath} , equal to the Mulliken electronegativity⁴⁹ of the molecule as derived from the quantum mechanical (QM) computation,

$$E_b = v^T q + \frac{1}{2} q^T J q - \chi_{\text{bath}}(d^T q - Q_0) + \frac{1}{2} \eta_{\text{bath}}(d^T q - Q_0)^2 \quad (4)$$

where Q_0 is the target value for the total charge of the molecule and η_{bath} determines the strength of the restraint penalty. The condition for a minimum is

$$v^T q + J q = \chi_{\text{bath}} d - \eta_{\text{bath}}(d^T q - Q_0)d \quad (5)$$

As one can see, the total charge constraint of the molecule, $d^T q - Q_0 = 0$, can only be realized when the atomic equalized electronegativities (left side of eq 5) are all equal to the bath electronegativity. Otherwise, the molecule would exchange charge with the bath. Conventional calibration methods try to reproduce QM atomic charges and are therefore sensitive to errors in the total charge as well. As a result, the optimized parameters will not show deviations between the QM Mulliken electronegativity⁴⁹ and the EEM equalized electronegativity. The hardness of the charge bath is used to control the relative importance of a correct

equalized electronegativity and the correct total charge. In the limit of a very high bath hardness, the restraint behaves like a conventional constraint, and the cost function becomes again insensitive to errors in the equalized electronegativity. In this work, a bath hardness of 5 eV is used, which results in a weak influence of the QM Mulliken electronegativities in the calibration procedure. In previous works, where a strict charge constraint is used during the calibration, the reference value for the electronegativity parameters is not determined by the training data. Instead, an arbitrary constraint on one of the atomic electronegativities is imposed.³⁷ We will demonstrate in section IV that the charge-bath approach indeed fixes the reference value for the atomic electronegativity parameters based on QM data.

One drawback of the EEM is that the model systematically overestimates the molecular polarizability of insulators in the limit of large molecules. In the macroscopic limit, the polarizability per unit volume of an insulator becomes constant, while the EEM predicts a quadratic increase with system size.^{29,33,43,50} One pernicious consequence is that EEM predicts vanishing multipole moments for large systems due to internal charge redistribution.⁴³ A solution is provided by the SQE model as introduced in the next section.

SQE. The second model relies on the introduction of a new set of variables, the so-called split-charges,³³ p_{ij} , which are linearly related to the effective charges as follows:

$$q_i = \frac{Q_{\text{tot}}}{N} + \sum_{j \sim i} p_{ji} \quad (6)$$

where the summation runs over all atoms j that are covalently bonded to atom i . The symbol p_{ji} stands for the charge transferred from atom j to atom i ; therefore, $p_{ji} = -p_{ij}$. This notation implicitly assumes that charge transfer is only possible between fragments that are connected through covalent bonds, which does not pose any difficulties for the molecules studied in this paper. The total charge constraint is realized by treating Q_{tot} as a constant after the transformation, or by using Q_{tot} as a variable in the charge bath approach. In the case of molecular ions, this approach assumes that the excess charge is distributed equally over the entire molecule. Although one may obtain a more accurate model by placing the excess charge in specific atoms, this choice is the simplest possible and does not introduce additional parameters.

For the remainder of the paper, it is convenient to introduce a matrix notation for the split charges, which allows for a straightforward substitution of eq 6 into the energy expression in eq 1. For each bond, we introduce an index k going from 1 to N_{cov} , where N_{cov} is the total number of covalent bonds. For each bond k that connects atom i_k and j_k , let $t_k = p_{j_k i_k}$ with $j_k < i_k$. This choice eliminates the need for the constraint $p_{ij} = -p_{ji}$. The matrix notation for eq 6 becomes

$$q = U u \quad \text{with} \quad U = \left[T \mid \frac{d}{N} \right] \quad \text{and} \quad u = \begin{bmatrix} t \\ \vdots \\ Q_{\text{tot}} \end{bmatrix} \quad (7)$$

where T is the transfer matrix with N rows and N_{cov} columns and t is the column vector with N_{cov} split charges. Each column k of matrix T contains only two nonzero elements corresponding to the atoms j_k and i_k that are connected by bond k . By construction, the first nonzero element of each column is +1 and the second is -1. For example, the matrix U of a water molecule, where the oxygen

atom comes first, has the following form:

$$\begin{bmatrix} 1 & 1 & 1/3 \\ -1 & 0 & 1/3 \\ 0 & -1 & 1/3 \end{bmatrix}$$

The split charge equilibration³³ (SQE) is an improvement of the EEM with an additional second-order term borrowed from the atom–atom charge transfer (AACT) model²⁹ to yield the correct trend of the polarizability for large systems, and with an extra first-order term to obtain an improved fitting of the equilibrium charges. When the parameters in the extra terms are set to zero, one retrieves the original EEM form. The additional second order term, $\sum_{k=1}^{N_{\text{cov}}} \kappa_k t_k^2$, is diagonal in the split charges and represents a harmonic energy term for the charge transferred through a chemical bond. The bond hardness parameter, κ_k , is a constant that depends on the type of bond. Bonds in conductor-like systems have a low bond hardness, while bonds in dielectric systems with low polarizability have a high bond hardness.³⁴ The linear term is a bond correction for the atomic electronegativity parameters. For each bond k , the electronegativity parameter of atom j_k is increased by $\Delta\chi_{j_k i_k}$ and the electronegativity parameter of i_k is decreased by the same amount. By construction, the correction parameters obey $\Delta\chi_{j_k i_k} = -\Delta\chi_{i_k j_k}$. The parameter $\Delta\chi_{j_k i_k}$ is a constant associated with the type of bond between atoms i_k and j_k . It mainly influences the charge distribution between bonded atoms and can be interpreted as an ad hoc correction to the atomic electronegativity parameters due to the direct molecular environment.

We must introduce a convention for the electronegativity correction parameters, in analogy with the definition of t_k , to obtain a practical matrix notation for the SQE model. For every bond k , we define $c_k = \Delta\chi_{j_k i_k}$ with $j_k < i_k$, which is now a set of independent parameters. The vector with corrections to the atomic electronegativity parameters is written as

$$v' = Tc \quad (8)$$

where T is the transfer matrix and c is a vector of length N_{cov} with electronegativity correction parameters. In terms of the split charges and the total charge, one can write the EEM energy (eq 1) as

$$E_{\text{EEM}} = v^T U u + \frac{1}{2} u^T U^T J U u \quad (9)$$

The SQE energy,

$$E_{\text{SQE}} = (v + v')^T U u + \frac{1}{2} u^T (U^T J U + J') u \quad (10)$$

has a comparable form but contains an additional linear and quadratic term. The matrix J' is a diagonal matrix with $N_{\text{cov}} + 1$ rows and columns. The first N_{cov} diagonal elements are the bond hardness parameters; the last diagonal element is zero.

III. COMPUTATIONAL METHODS

Parameter Calibration. The parameters in the EEM (η_i and χ_i) and the SQE model (η_i , χ_i , κ_k and c_k) must be calibrated with respect to a set of training data in order to obtain a useful model. In this process, one has to introduce atom and bond types and associate unique parameters with atoms or bonds of the same type. Several papers in the literature have used atomic charges, based on QM computations for a set of organic molecules, as

Table 2. Main Characteristics of the Two Training Sets ($\mathbf{P}^{27,37,38}$ and \mathbf{T}^{36}) Used for the Calibration of Charge Equilibration Parameters

	set P	set T
# molecules	166	500
# H atoms	1080	1899
# C atoms	714	1651
# N atoms	121	657
# O atoms	108	496
# F atoms	65	99
# S atoms		267
# Cl atoms	30	116
# Br atoms		44
# neutral	166	419
# positive		32
# negative		49

reference data.^{26,32,37–39,51} One postulates an objective function to measure the mismatch between the model charges and the QM charges:

$$X = \sum_{m=1}^M \sum_{i=1}^{N_m} w_{m,i} (q_{m,i}^{\text{QM}} - q_{m,i}^{\text{MODEL}})^2 \quad (11)$$

The first sum runs over the molecules in the set of training data for which QM computations have been carried out. The second sum runs over all atoms within molecule m . For each atom, the difference between the QM atomic charge, $q_{m,i}^{\text{QM}}$, and the charge obtained with the equilibration model for a given set of parameters, $q_{m,i}^{\text{MODEL}}$, is squared and multiplied with a weight $w_{m,i}$. In the conventional treatment of the least-squares procedure,⁵² the weights are related to the measurement error on the reference data. In this context, the reference data consists of theoretical numbers without measurement errors, and one must resort to ad hoc schemes to define the weights.^{36,37} Through the model charges, the cost function, X , becomes a function of the parameters. This cost can be minimized to find the parameters that result in an optimal performance for the selected training set.

A local minimization of the cost function (eq 11) may lead to multiple and disparate solutions for two reasons. Both may explain the large variety of EEM parameters reported in the literature. In the first place, due to the nonlinear dependence of the model charges on the second-order parameters, the cost function in eq 11 can in principle have multiple local minima. However, even if one finds the global minimum, the cost function may have (around the global minimum) a few directions of low curvature, which implies that very different sets of parameters are degenerate and thus equally good. The exact position of the minimum of the cost function along those directions is sensitive to small changes in the training data, and is therefore significantly affected by the model-development choices like the selection of molecules in the training set, the definition of atom types, and the population scheme used to define the QM atomic charges. We will investigate the relative importance of both mechanisms and their impact on the interpretation of calibrated parameters.

The calibration procedure used in this work is largely based on a previous EEM/SQE benchmark paper,³⁶ and we refer to that work for a detailed description. Our in-house code QFIT is used for the calibration of the parameters. The implementation of QFIT relies extensively on the MOLMOD library, which is also

Table 3. Summary of the Three Parameter Calibration Procedures

procedure	model	atom types	population scheme
1	EEM	element	Hirshfeld-I
2	EEM	element	Mulliken
3	SQE	force field	Hirshfeld-I

the basis for other software projects such as ZEOBUILDER,⁵³ MD-TRACKS,⁵⁴ and TAMKIN.⁵⁵ Below, we review the most important aspects of the calibration for this paper.

Two different sets of molecules are used for the calibration. The first set (P) consists of 166 organic molecules with elements H, C, N, O, F, and Cl; it is based on the work of Bultinck et al.^{26,37,38} The second set (T) contains 500 organic molecules comprising the elements H, C, N, O, F, S, Cl, and Br and is taken from the benchmark paper of Verstraelen et al.³⁶ A summary with the main characteristics of each set is given in Table 2. Lewis structures of each molecule are given in the first section of the Supporting Information. The geometry of each molecule is optimized at the MP2/CC-pVDZ level, followed by a single point calculation of the wave function at the MP2/Aug-CC-pVDZ level,^{56,57} using the Gaussian 03⁵⁸ program. Mulliken¹² and Hirshfeld-I¹⁷ charges are derived from the MP2/Aug-CC-pVDZ wave functions. For each molecule, the single point computation is repeated on the same geometry with one additional electron and one electron less to compute the molecular electronegativity with the Mulliken definition,⁴⁹ not to rely on Koopmans' theorem.⁵⁹ The parameters are calibrated in six different ways, using three different procedures applied to the sets P and T. The first two procedures are based on the EEM and group parameters per element, i.e., all atoms of a given element share the same parameters. The first procedure employs Hirshfeld-I charges as reference data, while the second uses Mulliken charges. The third procedure is based on the SQE and combines the atomic number and the number of bonds to define force-field atom (and bond) types and further uses Hirshfeld-I charges. Previous work showed that force-field atom types are only of fundamental importance in the context of the SQE model.^{36,37} The three different procedures are summarized in Table 3. The six combinations of training set and calibration procedure are denoted in the remainder of the text as P1, P2, P3, T1, T2, and T3. Note that T1, T2, and T3 correspond to calibrations HETS (Hirshfeld-I + EEM + trivial atom types + static cost function), METS (Mulliken + EEM + trivial atom types + static cost function), and HSFS (Hirshfeld-I + SQE + force field atom types + static cost function), respectively, in the benchmark paper of Verstraelen et al.³⁶ The parameters found in this work are slightly improved because of improvements in the conjugate gradient routine used to minimize the cost function. The prevalent atomic elements and force-field atom (and bond) types in both training sets are tabulated in the second section of the Supporting Information.

During the calibration, a lower bound is imposed on all second-order parameters to guarantee a positive definite matrix J in eq 1 for all possible molecular geometries.⁴³ The atoms in the EEM and SQE model in this paper are treated as Gaussian effective charge distributions. The distribution centered at the atom has the form of an S-type Gaussian:

$$\rho_i(\mathbf{r}) = q_i(R_i^2 \pi)^{-3/2} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_i|^2}{R_i^2}\right) \quad (12)$$

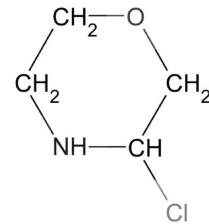


Figure 1. The chemical structure of 3-chloromorpholine.

where R_i is the radius of atom i and \mathbf{r}_i its position. The atomic radius parameters are based on the covalent radius of the atom.⁶⁰ For hydrogen three times the covalent bond radius is used. For all other atoms, 1.5 times the covalent bond radius is used. The minimal value for the atomic hardness parameter is equal to the self-interaction of the charge distribution of the atom divided by q_i^2 :

$$\eta_{i,\min} = \frac{1}{q_i^2} \int \frac{\rho_i(\mathbf{r}) \rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' = \sqrt{\frac{2}{\pi}} \frac{1}{R_i} \quad (13)$$

The radii and minimal values for the atomic hardness parameters are tabulated in section 3 of the Supporting Information. The bond hardness parameters must be larger than zero. With these lower bounds, the hardness matrix is always positive definite.⁴³ This is not only interesting from a physical point of view, as it ensures that the energy function in eq 1 has a minimum. It also renders the calibration of the parameters more tractable, as the cost function in eq 11 diverges for parameters that make the hardness matrix singular for one of the molecules in the training set.^{61,62} The restriction of the parameters to positive definite hardness matrices therefore confines the parameters to the region where the cost function is well-behaved, which is beneficial for the optimization of the parameters.

Parameter Scans. The minimum of the cost function in eq 11 does not necessarily correspond to a sharply determined point in the multidimensional parameter space. The minimum may also lie in an elongated valley where many disparate parameter vectors lead to virtually the same (minimal) value of the cost function. Elongated valleys imply a low local curvature of the cost function at the position of the optimal parameters. Because the cost function is multidimensional, this local curvature is characterized by the Hessian, i.e., the symmetric matrix of second-order derivatives with respect to the parameters. Eigenvectors of the Hessian associated with relatively small eigenvalues correspond to displacements of the optimal parameters that lead to a relatively small increase in the cost function. The spectrum of the Hessian, computed from finite differences of the analytical first-order derivatives, is analyzed for all six calibrations in this paper. In general one can use the condition number of the Hessian, defined as the ratio of the highest over the lowest eigenvalue of the Hessian, to assess the presence of low-curvature eigenvectors at the minimum of the cost function and the resulting lack of robustness of the parameters toward changes in the training data.

In order to assess the impact of parameter displacements along a low eigenvector, we compute for a test molecule (3-chloromorpholine, see Figure 1) the AIM charges resulting from EEM and SQE, equalized electronegativity, and global hardness for different displacements of the optimal parameters along this low eigenvector. Specifically, we consider a series of parameter vectors along a line segment in the parameter space that is centered on the

optimal parameters, is parallel to an eigenvector, and has a length of 20 eV. The points on this line segment are characterized by the mathematical relation:

$$\mathbf{x}_{\text{scan},i}(\gamma) = \mathbf{x}_{\text{opt}} + \gamma \mathbf{a}_i \quad (14)$$

where \mathbf{x}_{opt} is the optimized parameter vector, \mathbf{a}_i is an eigenvector associated with the scan, and γ is the line parameter going from -10 eV to $+10$ eV. This type of analysis is only discussed for the calibration P1, but the same trends are present in the other calibrations.

Influence of the Training Set. There is a long history in computational chemistry of the use of low-frequency vibrational modes in molecular systems to search for or explain conformational changes.⁶³ For example, in biochemistry, one typically expands the relative vector of Cartesian coordinates of two protein conformers into the basis of lowest vibrational eigenmodes of one of the conformers to investigate if extended motion along these modes could lead to a transition of one conformer to the other.^{64–67} We use a similar analysis technique to compare two parameter vectors that are derived with the same procedure but based on different training sets, e.g., to compare P1 with T1. The difference between the two parameter vectors is expanded in the basis of eigenvectors of the Hessian of either P1 or T1. This allows one to compute, for each eigenvector, the percentage contribution to the difference between two parameter vectors. Because calibration T1 contains more atom types and hence more parameters, we restrict the relative parameter vector to the parameters that are present in both P1 and T1. Likewise, the Hessian of T1 is reduced to those parameters present in P1 by omitting the columns and rows that correspond to the parameters that are only present in T1.

Let the basis of eigenvectors be \mathbf{a}_i and let the parameter vectors be \mathbf{x}_{P1} and \mathbf{x}_{T1} ; then, the percentage contribution to the difference between two parameter vectors due to basis vector i , Pct_i , is defined as

$$\text{Pct}_i = \frac{((\mathbf{x}_{\text{P1}} - \mathbf{x}_{\text{T1}}) \cdot \mathbf{a}_i)^2}{\|\mathbf{x}_{\text{P1}} - \mathbf{x}_{\text{T1}}\|^2} \times 100\% \quad (15)$$

The same analysis is used to compare the parameters obtained with P2 versus T2 and P3 versus T3.

Influence of the Charge Population Scheme. Strictly speaking, there are no measurement errors on the QM atomic charges since all data are obtained from theoretical computations. Nevertheless, we do not expect the QM atomic charges to be totally robust with respect to the initial guess of the wave function, limited convergence criteria, numerical errors, and so on. This lack of robustness adds an unpredictable contribution to the atomic charges, which we can treat as measurement errors. Such errors on the training data will result in errors on the calibrated model parameters. This error will also contribute to the deviation between two parameters derived from different training sets. The norm of the difference between the parameters of P1 and T1 can be compared with the norm of the difference between parameters of P2 and T2. Because procedures 1 and 2 only differ in the choice of QM charge population scheme, the relative magnitude of the two norms is an indication for the relative robustness of the underlying population schemes.

Multiple Minima. In the present subsection, we introduce a method to provide an indication for the absence of multiple minima in the cost function of T1 when the second-order parameters are restricted to the region where the hardness matrix is guaranteed to be positive definite.

As argued before, AIM charges are contaminated to some extent by numerical errors. Even in the absence of these numerical errors, there will be a residual error between the charges predicted by the EEM or the SQE and the QM training data, simply because both equilibration models are approximations that cannot capture all features and trends in the QM computations.

Assume that the numerical errors and the elusive trends in the QM atomic charges can be modeled with a simple uncorrelated Gaussian error that is completely responsible for the deviation between the QM data and the model predictions. The variance of this Gaussian noise is not known a priori. This assumption is only an approximate description because elusive trends are not true random noise. However, it provides a useful statistical model for the errors between model and reference charges.

In the vicinity of the optimal parameters, we can approximate our cost function by a generalized linear least-squares cost function, except that our cost function (eq 11) contains weight factors that replace the measurement errors that are typically present in a χ^2 -type cost function.⁵² The propagation of errors from the training data to the estimated parameters is mathematically analogous, and therefore the inverse of the Hessian of our cost function is a model for the covariance matrix of the parameters,⁵² up to a constant factor because we do not know the variance of the error on the training data a priori.

We validate this hypothesis, i.e., the relation between the Hessian of the cost function and the uncertainty in the parameters, by studying the covariance of the parameter vectors obtained from 100 distinct cross-validation calibrations for T1. The difference between the 100 runs is the selection of the molecules in the training set: for each case, a random representative subset of 400 molecules is taken from the full set of 500 molecules. The same subsets are used as for the cross-validation in our EEM/SQE benchmark paper,³⁶ and we refer to that work for further details on the generation of the subsets. The covariance matrix of the 100 parameter vectors and the Hessian of the cost function in T1 can be compared in terms of eigenvalues and corresponding eigenvectors. The same analysis is carried out for T2, but not for T3. The latter calibration fits more than 100 parameters, which makes it impossible to estimate the covariance of the parameters using 100 samples.

One can further test whether the parameter vectors obey the multivariate normal distribution dictated by the Hessian of the cost function. One applies an affine transformation to the 100 parameter vectors that brings the average to the origin and turns the covariance matrix of the transformed parameter vectors into a unit matrix. In this new coordinate system, the squared norm of the parameter vectors should follow the χ^2 distribution with the number of degrees of freedom equal to the number of model parameters.

A strict validation of our statistical model for the training data is not the purpose of this analysis. We are rather interested in a procedure to examine and invalidate the existence of multiple minima for this type of cost function. The initial first-order parameters are sampled between -50 and $+50$ eV. The initial second-order parameters are selected randomly between 5 and 50 eV above their lower bound. This should be sufficient to trigger optimizations into different local minima if they would exist. The tests for the statistical model of the parameter distribution must fail dramatically if multiple minima are present in our cost function. The spread of the 100 parameter vectors over multiple minima cannot be predicted by the Hessian of the cost function

of calibration T1. By consequence, if the tests do not fail, it is very unlikely that multiple minima are present in the cost function of T1.

Parameter Correlations in EEM. In this subsection, we only consider the EEM model for a more detailed analysis of the statistical correlations between the atomic electronegativity and hardness parameters. Consider at first instance a molecule in which each atom has its own (η_i, χ_i) parameter pair. The equations to determine the charges, q_i , in this molecule are given by

$$\chi_i + q_i \eta_i + \sum_{j=1, j \neq i}^N q_j J_{ij} = \chi_{\text{mol}} \text{ and } \sum_{j=1}^N q_i = Q_{\text{tot}} \quad (16)$$

where J_{ij} represents the electrostatic interaction between atoms i and j , χ_{mol} is the equalized electronegativity, and Q_{tot} is the total charge. Once these equations are solved, one can change the parameter χ_i of each atom and update the value of η_i according to

$$\eta_i = \frac{\chi_{\text{mol}} - \chi_i - \sum_{j=1, j \neq i}^N q_j J_{ij}}{q_i} \quad (17)$$

without changing the equilibrium charges.³⁷ If one substitutes the QM atomic charges and the molecular electronegativity into eq 17, one finds a linear relation for all possible optimal (η_i, χ_i) parameter pairs associated with atom i . It was to be expected that the optimal parameters are not unique, since there are twice as many parameters as there are atomic charges to be reproduced.

The statistical correlations so far consider independent parameters for each atom in the entire set of molecules, while in practice the same parameters are used for atoms with the same element (or force-field atom type). Therefore, we introduce a graphical representation for every element in scheme P1 to analyze the statistical correlation data of all of the corresponding independent atomic parameters. Given an element, we construct a plot in the (η_i, χ_i) plane with the lines according to eq 17 for all atoms alike in all of the molecules in the training set. When this bundle of lines features an approximate common intersection point, we expect that the calibrated P1 parameters for that pair to be close to the intersection. In other words, a contour plot of the cost function of P1 in the (η_i, χ_i) plane, keeping all other parameters constant at their optimal value, should reveal a minimum in the vicinity of the approximate intersection point of the bundle of lines. When a bundle of lines is almost parallel for a given element, the corresponding optimal (η_i, χ_i) parameter pair should be ill-defined, and the contour plot should reveal an elongated valley instead of a unique minimum. Note that our cost function is designed to take also into account the electronegativity of each molecule in the training set; otherwise one should not expect any correspondence between the contour plots and the lines defined by eq 17.

IV. RESULTS AND DISCUSSION

For ease of reference, results from the methods described in the various subsections of section III, are discussed in the corresponding subsections of section IV.

Parameter Calibration. In total, six calibrations were carried out with a subsequent detailed analysis of the significance of the parameters. The optimal parameters can be found in the Supporting Information, section 4. Before analyzing the results, we briefly review the main statistical properties of the calibrations.

The R^2 and RMSE values in Table 4 immediately reveal the major differences between the six calibrations. First of all: procedures 1 and 3 are superior in accuracy compared to procedure 2. The latter is based on Mulliken charges, while the former two use Hirshfeld-I charges. It is known that Mulliken charges are ill-behaved when large basis sets with diffuse functions are used,^{12,15,68} while Hirshfeld-I charges are much less basis set dependent.²³ The R^2 and RMSE values confirm that Mulliken charges have a larger degree of arbitrariness, which makes them more difficult to reproduce with charge equilibration models. A second observation is that the SQE (in procedure 3) is more accurate than the EEM (procedure 1). Both observations are in line with earlier work.^{33,36} When comparing results from training set P with those of T, the reference data based on set P are easier to reproduce. This is related to the fact that set P contains fewer molecules and that set T is generated with an algorithm that tries to maximize the diversity of chemical functional groups.

As discussed above, a high condition number is an indicator for less robust parameters, i.e., parameters that are more sensitive to irrelevant details in the training data. The optimal parameters are ill-conditioned in all six cases. All of the condition numbers of the Hessian are at about 5 or more orders of magnitude above the minimal and ideal value, i.e., 1. The condition number is clearly higher for calibration T3, and the Hessian is virtually singular in case P3. This is partly due to the larger number of parameters in the SQE model and to the introduction of multiple force-field atom types for the same element. Also note that set P is designed to be used for EEM calibrations without extensive use of force-field atom types,²⁶ while set T is specifically constructed with force-field atom and bond types in mind.³⁶ It is therefore expected that set P does not contain sufficient information to fix all of the parameters in procedure 3. This is also visible in the actual parameters of P3, e.g., the electronegativity parameter of atom type N1 is 24.53 eV, which is excessively high. The Hessians of T1, T2, and T3 are positive definite in this work, which was not the case in the EEM/SQE benchmark paper.³⁶ The conjugate gradient optimizer used in this work is extended with a diagonal preconditioner, which significantly improves the rate of convergence compared to a conventional conjugate gradient optimizer.

Calibrations P2, T2, and P3 clearly have weaknesses and should not be used for general applications. They are however interesting for this paper, revealing the typical problems inherent in calibrated EEM and SQE parameters.

Correlated Parameters in a Diatomic Molecule. It is instructive to show how the atomic charges are invariant to certain changes in parameters when the SQE model is applied to a neutral heteronuclear diatomic molecule. The charges are given by

$$q_1 = -q_2 = \frac{\chi_1 - \chi_2}{\eta_1 + \eta_2 + \kappa - \frac{2}{r_{ij}}} \quad (18)$$

where the last term in the denominator is twice the electrostatic interaction between the two atoms. As mentioned earlier, one can replace the electrostatic term by other functional forms, which will clearly affect the optimal values of the hardness parameters. There are in total five parameters to fix only one unknown, leaving four degrees of freedom in the parameters that cannot be fixed in a calibration where charges are used as reference data. The following degrees of freedom remain:

- An arbitrary constant can be added to both electronegativity parameters without changing the charge q_1 . This was already

Table 4. Key Statistical Characteristics of Each Calibration (P1, T1, P2, T2, P3, and T3)^a

quantity	(a) R ² [%]		(b) RMSE [e]		(c) condition number		(d) # parameters	
	training set	P	T	P	T	P	T	P
procedure 1	95.47	89.61	0.0646	0.1149	2.74×10^5	4.15×10^5	12	16
procedure 2	67.05	53.74	0.2674	0.3006	9.83×10^5	8.09×10^4	12	16
procedure 3	99.36	98.18	0.0242	0.0481	6.79×10^6	7.0×10^6	70	135

^a (a) The Pearson correlation coefficient between the QM and model charges, (b) the root-mean-square error (RMSE) between QM and model charges, (c) the condition number of the Hessian of the cost function (defined as the ratio of the highest over the lowest eigenvalue of the Hessian), and (d) the number of parameters in the charge equilibration model.

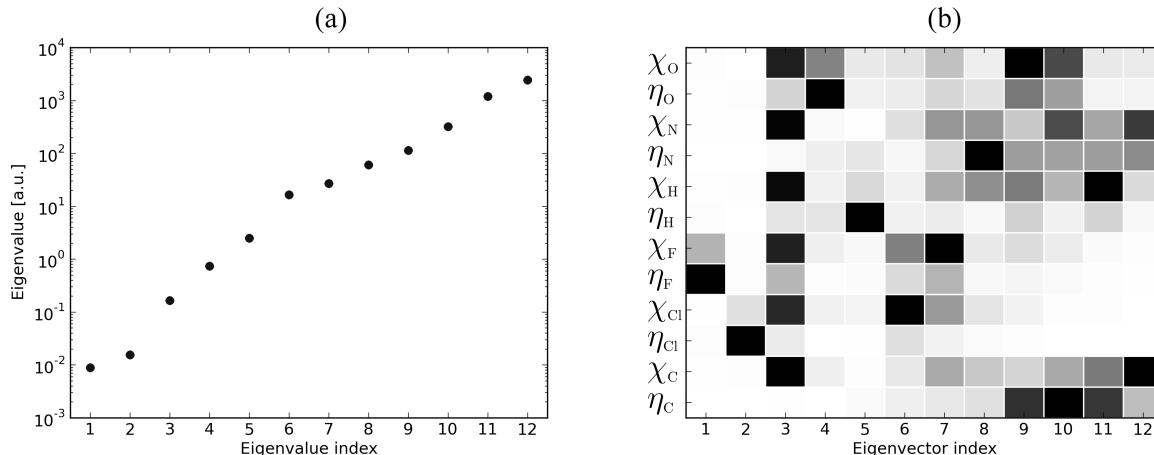


Figure 2. Graphical representation of the eigenvalues (a) and the eigenvectors (b) in calibration P1. Each column in part b corresponds to an eigenvector, and each row corresponds to a parameter. The cells are colored on the basis of the absolute value of the corresponding matrix element: the darker the color, the larger the absolute value.

mentioned in section II, and this invariant is also present in more general molecules.

- 2 One can also increase the electronegativity difference and compensate for this by increasing the parameters in the denominator of eq 18. This type of parameter correlation is discussed at the end of section III and section IV when the EEM model is applied to more general molecules.
- 3 The last two invariants consist of an increase (or decrease) in one hardness parameter that is exactly compensated with opposite changes in the other hardness parameters. In the case of the EEM, such problems were not observed, probably because the training sets contain a large set of molecules to which this invariant does not simply apply.

Parameter Scans. Before we proceed with the eigenvector-following analysis, we first examine the eigenvalue decomposition of the Hessian. The results for calibration P1 are depicted in Figure 2. Similar plots for all other calibrations are included in the Supporting Information, section 5. Note that the eigenvectors could not be properly visualized for P3 and T3 due to the large number of parameters. Figure 2b visualizes the eigenvectors as follows. Each column in the figure corresponds to an eigenvector of the Hessian of the cost function used in calibration P1, sorted from low to high eigenvalue. Each row corresponds to a parameter in the P1 calibration. The cells are colored on the basis of the absolute value of the corresponding coefficient of the eigenvector: the darker the color, the larger the absolute value. Figure 2 reveals the following two trends. First, the lowest two eigenvectors of the Hessian indicate that the P1 cost function is

not sensitive to a linear combination of the ($\eta_i \chi_j$) parameter pair of each halogen. To a lesser extent, the same problem is visible for oxygen. Similar conclusions can be drawn from the Hessian of T1. Second, the influence of the charge bath is visible in eigenvector 3; i.e., the cost function is (weakly) sensitive to a constant shift of all electronegativity parameters. If a strict total charge constraint was used during the parametrization, this eigenvector would have a corresponding eigenvalue of zero because the cost function is then completely insensitive to the reference value for the electronegativity parameters. Both observations also apply to the calibrations based on the Mulliken charges, P2 and T2.

Figure 3 contains the results for two linear scans through the parameter space in the context of calibration P1. The first scan (Figure 3a and b) is taken along the eigenvector with the lowest-but-one eigenvalue of the Hessian. The scan with the lowest eigenvalue is not used because it mainly affects the fluorine parameters, while there is no fluorine present in the test molecule (3-chloromorpholine, see Figure 1). The second scan is similar but based on the highest eigenvalue (Figure 3c and d). All 15 atomic charges (3a and c) and the molecular electronegativity and hardness^{59,70} (3b and d) are computed for the test molecule in both scans. The hatched region is excluded during the calibration due to the lower bounds on the second-order parameters. In the red region, the hardness matrix of the test molecule has at least one negative eigenvalue. The beginning, middle, and end point of each scan and the eigenvectors are given in section 6 of the Supporting Information.

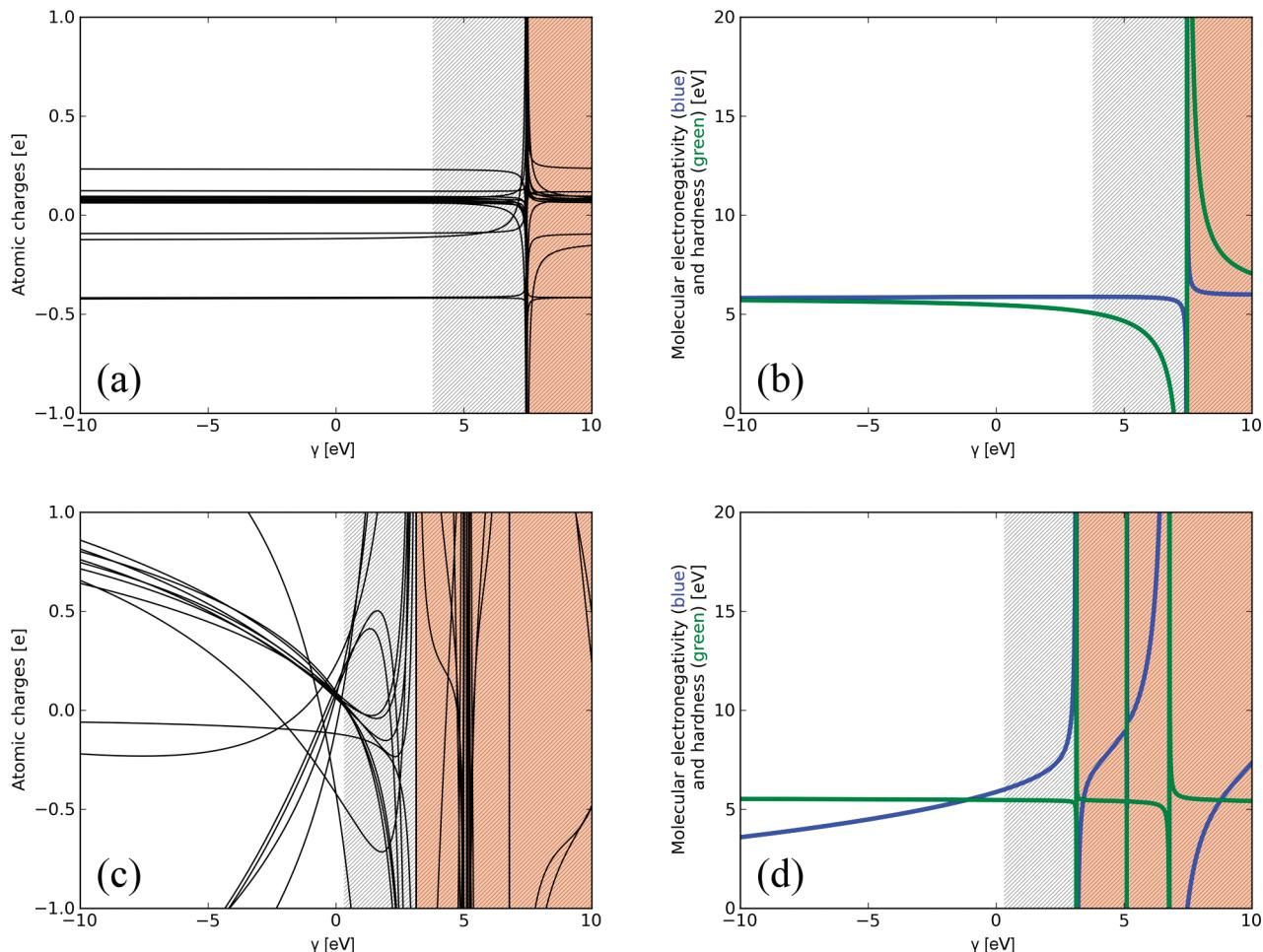


Figure 3. Evolution of the 15 atomic charges (parts a and c) and molecular electronegativity and hardness (blue and green respectively in parts b and d) during two linear scans through the parameter space for a test molecule (3-chloromorpholine, see Figure 1). The optimal P1 parameter vector corresponds to $\gamma = 0$ in each plot. In the top row (a,b) the scans are taken along the eigenvector of the P1 Hessian with the lowest-but-one eigenvalue. In the bottom row (c,d) the highest eigenvalue is used. The hatched region is excluded during the calibration due to the lower bounds on the second-order parameters. In the red region, the hardness matrix of the test molecule is no longer positive definite.

The first scan shows virtually no changes in atomic charges as the parameter vector is displaced in the region that is allowed during the calibration (see Figure 3a). Only when the parameters lead to a hardness matrix that is almost singular do the charges change significantly. At the edge of the red region, the hardness matrix has one zero eigenvalue; the charges (and therefore also the cost function) diverge.^{61,62} The molecular electronegativity shows the same behavior as the charges: it is practically constant except near the singularity in the hardness matrix. The molecular hardness is more sensitive than the electronegativity. The weak sensitivity of the charges along this eigenvector is a general feature for many molecules in the training set, which explains the low curvature of the cost function for this eigenvector.

The scan along the eigenvector with the largest eigenvalue goes through three points where the hardness matrix becomes singular (Figure 3d). At each of these three points, all computed quantities diverge. The charges show stronger fluctuations. The molecular electronegativity is sensitive, and the molecular hardness is not. The fact that the cost function has the highest curvature along this direction is solely due to high sensitivity of the atomic partial charges for changes in parameters along this direction.

The figures discussed above show that the parameter space is full of transitions where one of the eigenvalues of the hardness matrix (of one of the molecules in the training set) changes sign. Each of these transitions takes place at a hypersurface in the parameter space that separates feasible regions in which the atomic charges remain finite. At the boundary of each feasible region, the cost function diverges toward plus infinity, and consequently, each region contains at least one local minimum. However, there is only one feasible region for which all hardness matrices in the training set are positive definite, which will be denoted below as *the positive definite feasible region*. All other feasible regions are called *nonpositive definite*.

The contour plot of the cost function P1 in Figure 4 shows the multitude of feasible regions (for example, when $\eta_H < \eta_{H,\min}$ or $\eta_C < \eta_{C,\min}$). The region excluded during the calibration is darker and separated by a thick dashed line from the allowed region. From the practical point of view, one should avoid allowing the calibration algorithm to converge to parameters that are locally optimal in one of these tiny *nonpositive definite* feasible regions. It is not just that it is difficult to find the feasible region with the lowest minimum. The real problem is that one can always find a molecular structure that has a singular hardness matrix for those

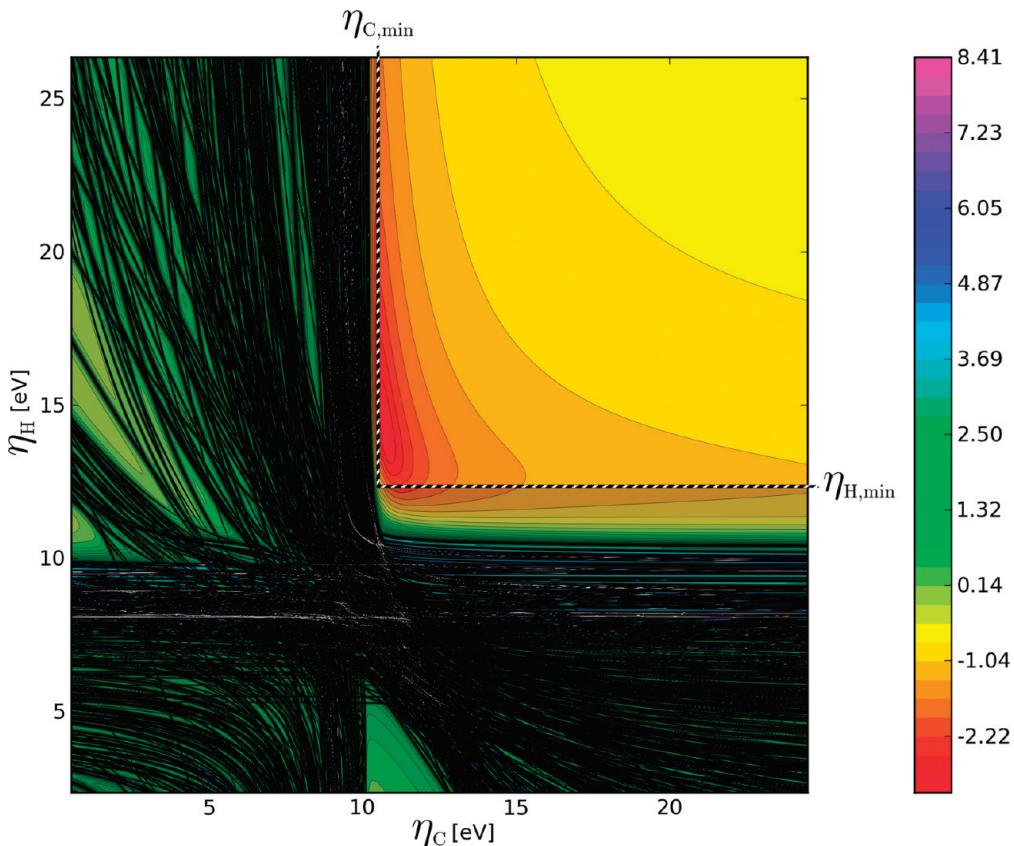


Figure 4. A contour plot of the cost function of P1 with two atomic hardness parameters as variables (η_C and η_H). All other parameters are kept at their optimal values. The allowed region during the calibration comprises the top-right corner of the scan, bordered by the dashed line. The contours represent iso-curves for the logarithm of the cost function.

Table 5. The Values of the Parameters Present in Both P1 and T1 and Their Differences^a

parameter	P1 [eV]	T1 [eV]	difference [eV]
χ_H	5.583	7.019	-1.436
χ_C	6.263	8.002	-1.738
χ_N	6.847	8.610	-1.763
χ_O	8.173	9.699	-1.525
χ_F	8.625	11.150	-2.525
χ_S		7.510	
χ_{Cl}	6.602	8.426	-1.823
χ_{Br}		7.769	
η_H	14.126	15.483	-1.357
η_C	10.743	10.718	0.025
η_N	10.887	10.877	0.010
η_O	13.068	12.724	0.344
η_F	17.235	21.199	-3.963
η_S		7.870	
η_{Cl}	11.266	15.291	-4.025
η_{Br}		18.933	

^aThe norm of the difference is 7.354 eV.

only one minimum in the positive definite feasible region when all degrees of freedom are considered.

Influence of the Training Set. The parameters for calibrations P1 and T1 are given in Table 5. They show that the same calibration procedure applied to two different sets of training data can lead to significant differences in the optimal parameters. Similar differences are present when comparing P3 with T3 and become even more pronounced in sets P2 and T2. This subsection relates such differences to the low eigenmodes of the Hessian belonging to the cost functions used for the calibration. The discussion and data in this subsection are based on calibrations P1 and T1, but the same explanation is valid for the other calibrations. The data for P2 versus T2 and P3 versus T3 are given in the Supporting Information, section 7.

The comparison between P1 and T1 is troubled by the sulfur and bromine parameters χ_S , η_S , χ_{Br} , and η_{Br} that are present in T1, but absent in P1. Set P does not contain molecules with sulfur or bromine. For this reason, the parameter vector of T1 is projected on the parameter space of P1 by simply omitting the additional four parameters. An analogous projection is applied to the Hessian of T1; i.e., all rows and columns corresponding to those four parameters are deleted. After the projection operations, one can compute the difference between the parameter vectors of P1 and T1 and expand the difference on the basis of eigenvectors of the Hessian of P1 or the projected Hessian of T1. The same type of projection must be carried out when comparing P2 with T2 or P3 with T3.

The results for the comparison of P1 and T1 are given in Tables 5, 6, and 7. Table 5 also reports the difference between the

locally optimal parameters. Note that all nonpositive definite feasible regions have at least one hardness parameter that is below its lower bound. In the allowed region, only one local minimum is visible, although this does not prove that there is

Table 6. Decomposition of the Vector Representing the Difference between the P1 and T1 Parameters in the Basis of P1 Hessian Eigenvectors^a

index	relative eigenvalue	coefficient [eV]	contribution [%]
1	3.646e-6	-4.619e+0	39.44
2	6.341e-6	-4.117e+0	31.35
3	6.807e-5	-3.788e+0	26.52
4	3.053e-4	1.656e-1	0.05
5	1.017e-3	1.143e+0	2.41
6	6.794e-3	-5.107e-2	0.00
7	1.098e-2	3.133e-1	0.18
8	2.499e-2	1.175e-1	0.03
9	4.695e-2	-5.788e-2	0.01
10	1.322e-1	3.197e-3	0.00
11	4.907e-1	-3.959e-2	0.00
12	1.000e+0	-2.426e-2	0.00

^a The eigenvalues of the P1 Hessian are given relative to the highest eigenvalue.

Table 7. Decomposition of the Vector Representing the Difference between the P1 and T1 Parameters in the Basis of Projected T1 Hessian Eigenvectors^a

index	relative eigenvalue	coefficient [eV]	contribution [%]
1	4.636e-06	-4.470e+00	36.95
2	1.283e-05	-4.162e+00	32.02
3	1.213e-03	-2.614e-01	0.13
4	1.579e-03	1.401e+00	3.63
5	3.437e-03	-2.227e+00	9.17
6	3.970e-03	1.105e+00	2.26
7	1.920e-02	-2.828e+00	14.79
8	3.264e-02	-6.524e-01	0.79
9	6.963e-02	-3.314e-01	0.20
10	2.208e-01	-3.148e-02	0.00
11	3.544e-01	1.831e-01	0.06
12	1.000e+00	-5.094e-02	0.00

^a The eigenvalues of the projected T1 Hessian are given relative to the highest eigenvalue.

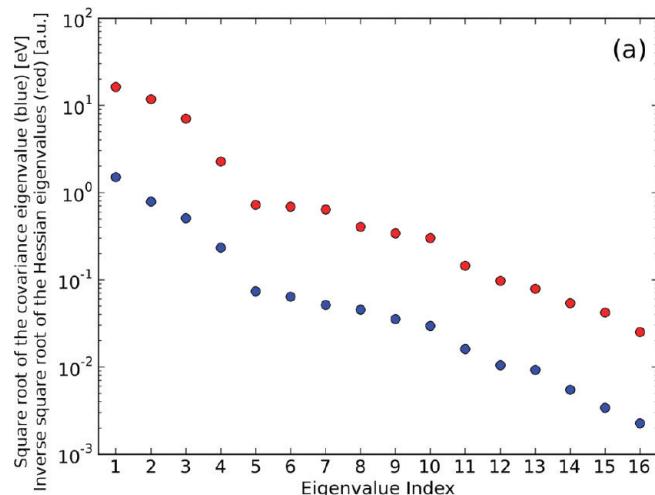
parameters of P1 and T1. Note that the significant differences fall into two categories: (i) all electronegativity parameters in T1 are about 1.8 eV higher than those in P1; (ii) the parameters of the halogens differ more than those of the other atoms. Tables 6 and 7 give the decomposition of the relative parameter vector in the basis of the Hessian of P1 and the projected Hessian of T1, respectively. The percentage contributions of the lowest three eigenmodes in case P1 (Table 6) add up to 97%. In Table 7, the first two eigenmodes only account for 69% of the difference, and some higher modes also have a non-negligible contribution. When comparing P2 with T2 or P3 with T3, the same pattern is found: at least one of the two Hessians has low eigenmodes that form a nearly complete basis for the differences between the parameters. We conclude that cost functions based on training set P or T essentially have the same ill-conditioned minimum, but due to the subtle differences in the reference data, the exact position of the optimal point is shifted along the directions of low curvature on the cost function surface.

Influence of the Charge Population Scheme. The norm of the difference between the parameters of P2 and T2 (83.164 eV) is much larger than the norm of the difference between P1 and T1 (7.354 eV). This means that the parameters obtained with procedure 2 are much more sensitive to the selection of molecules in the training set. The only difference between procedures 1 and 2 is that 1 uses Hirshfeld-I charges while 2 is based on Mulliken charges. The lack of robustness of the Mulliken charges when using large basis sets for the QM computation, i.e., its sensitivity to irrelevant details in the wave function,^{12,15,68} clearly causes two undesirable effects: (i) Mulliken charges are difficult to reproduce with equilibration models (as discussed above), and (ii) the position of optimal parameters along the directions of low curvature of the cost function are heavily scattered. In the context of generalized linear least squares methods, one can show that the noise on the parameters is a linear function of the noise on the training data.⁵² Because our cost function can be locally approximated as a linear least squares cost function, the same relation applies. Similar conclusions can be drawn for any QM population scheme that is not fully robust. For example, ESP fitted charges tend to be sensitive to the choice of grid points, while the choice of grid points is irrelevant from the physical perspective.¹⁸ Therefore, we expect that ESP-fitted charges are, just like Mulliken charges, of little use for the calibration of statistically sound EEM or SQE parameters.³⁸

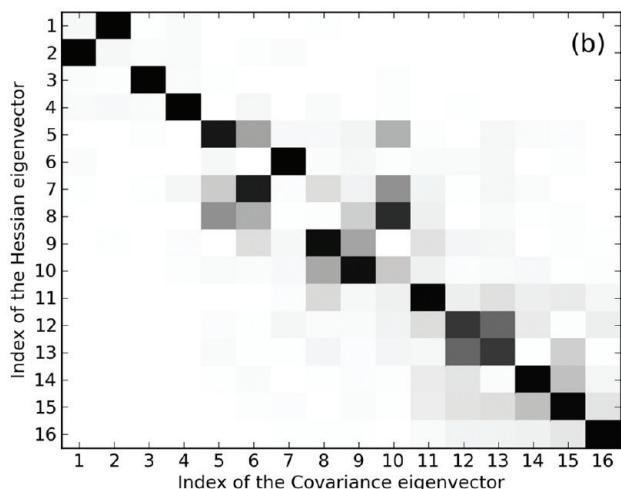
Multiple Minima. The cross-validation runs for T1 and T2 confirm the reliability of the calibration procedure. For each run, RMSE and R^2 values are computed to test how well the parameters based on a selection of 400 (out of 500) molecules can reproduce the data of the 100 remaining molecules. In line with previous work,³⁶ the performance measures obtained in the cross-validation are very comparable to the numbers obtained in Table 4. The remainder of the discussion is limited to the results based on calibration T1. Similar data for calibration T2, leading to the same conclusions, are included in the Supporting Information in section 8.

The sampling covariance matrix of the 100 parameter vectors is closely related to the Hessian of the cost function. In principle, our covariance estimate does not converge to the true covariance of the parameters because each training subset of 400 molecules shares at least 300 molecules with all other subsets. Therefore, our current approach systematically underestimates the true covariance. This is of little concern because the model for the parameter covariance, based on the inverse of the Hessian of T1, is only known up to a constant factor. In Figure 5a, the square root of the sampling covariance eigenvalues are plotted in blue. The first eigenmode corresponds to a spread of about 1 eV on the parameters. The inverse of the square root of the Hessian eigenvalues are plotted as red dots. Both data series follow the same trend up to a constant factor, a first confirmation that the Hessian of the cost function minimum can be used to explain the fluctuations on the EEM parameters. The eigenvectors of the covariance and the Hessian are compared in Figure 5b, which is a visual representation of the overlap matrix of both sets of eigenvectors. Each cell of the overlap matrix is colored according to its absolute value: the darker the color, the larger the absolute value. In the ideal picture, only the diagonal elements would be 100% black, while all off-diagonal elements would remain white. Figure 5b reveals some contamination between eigenvectors with similar eigenvalues, but the overall correspondence between the two sets of eigenvectors is good.

The statistical explanation of the variation on the parameters can be further validated by studying the distribution of the



(a)



(b)

Figure 5. Comparison of the covariance of the 100 parameter vectors from the cross-validation of calibration T1 with the inverse of the Hessian of the cost function of T1. (a) The square roots of both matrices. Note the square root of a covariance eigenvalue can be interpreted as the standard deviation of the parameters along the corresponding eigenvector. (b) Visual representation of the overlap matrix of the eigenvectors of covariance and the Hessian. Each cell of the overlap matrix is colored on the basis of its absolute value: the darker the color, the larger the absolute value.

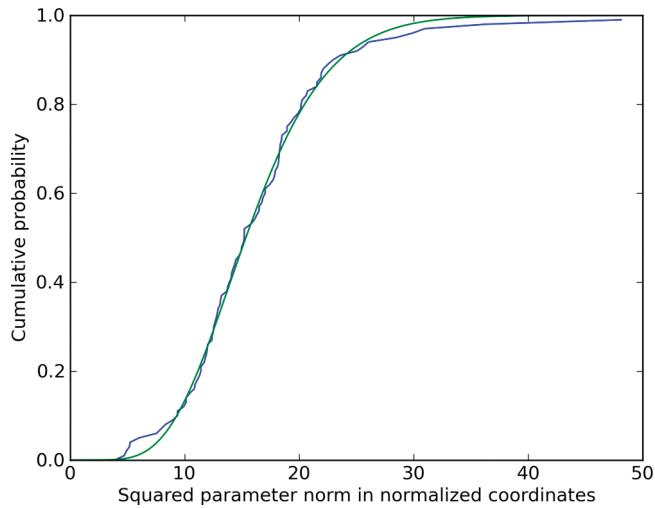


Figure 6. Comparison of the cumulative frequency distribution of the squared norm of the 100 parameter vectors from the cross-validation of T1 in normalized coordinates (blue) and the expected cumulative χ^2 distribution with 16 degrees of freedom (green).

distance of each parameter vector from the average. First, we apply an affine transformation to the parameters such that the average vector is zero and the covariance matrix becomes unity. The squared norm of the parameter vectors in these normalized coordinates should be distributed according to a χ^2 distribution with 16 degrees of freedom, i.e., the number of parameters in T1. A comparison of the cumulative frequency plot with the exact cumulative χ^2 distribution is presented in Figure 6. Small aberrations are found in the tails of the distribution, but there is no indication that firm outliers are present.

Both tests point out that the Hessian of the cost function explains the uncertainty on the parameters of T1. This is only possible if all 100 parameter vectors essentially describe the same minimum but are slightly displaced due to differences in the training data. If the 100 parameter vectors would be due to truly different minima in the cost function, one would not expect such

a good correspondence between the sampling covariance matrix and the inverse of the Hessian of T1. Therefore, we conclude that cost function T1 has only one minimum in the positive definite feasible region. This justifies the use of a local minimization algorithm for the calibration of the parameters instead of a computationally more expensive global minimizer.

Finally, we must stress that this analysis can only work under two important conditions. First, the calibration of the parameters must be restricted to one feasible region, e.g., by imposing lower bounds on the second-order parameters. Otherwise, one can find at least one minimum in the cost function for each feasible region. Second, one must use a training set for the calibration and the subsets for the cross-validation that contain sufficient information to determine all parameters. If not, the covariance of the parameters diverges.

Parameter Correlations in EEM. All results so far point out that parameters in the EEM and SQE models are inherently correlated to a high degree, even when the molecules in the training sets are carefully selected to contain sufficient information to estimate the parameters. In this subsection, we show that correlated parameters are practically unavoidable with cost functions that are based solely on atomic partial charges. The results in this subsection are based on calibration P1.

Figure 7a and b contain lines defined in eq 17, for carbon and fluorine atoms, respectively. The optimal parameters for carbon and fluorine are indicated with a black cross. The corresponding contour plots of the cost function in the same planes are shown in Figure 7c and d, respectively. The same color scale is used as in Figure 4. The plots for other elements (H, N, O, and Cl) are given in section 9 of the Supporting Information. As discussed in the methods, the optimal parameters for a given element should coincide with the approximate crossing point of the bundle. This is always the case, except for a constant shift of all of the electronegativity parameters of about +2 eV. Such a deviation is to be expected since our cost function is only weakly sensitive to the reference value for the electronegativity parameters. The correlation between the line plots and the corresponding contour plots is striking. Both the positions and the shapes of the minima in the contour plots have a surprising direct link with the reference data.

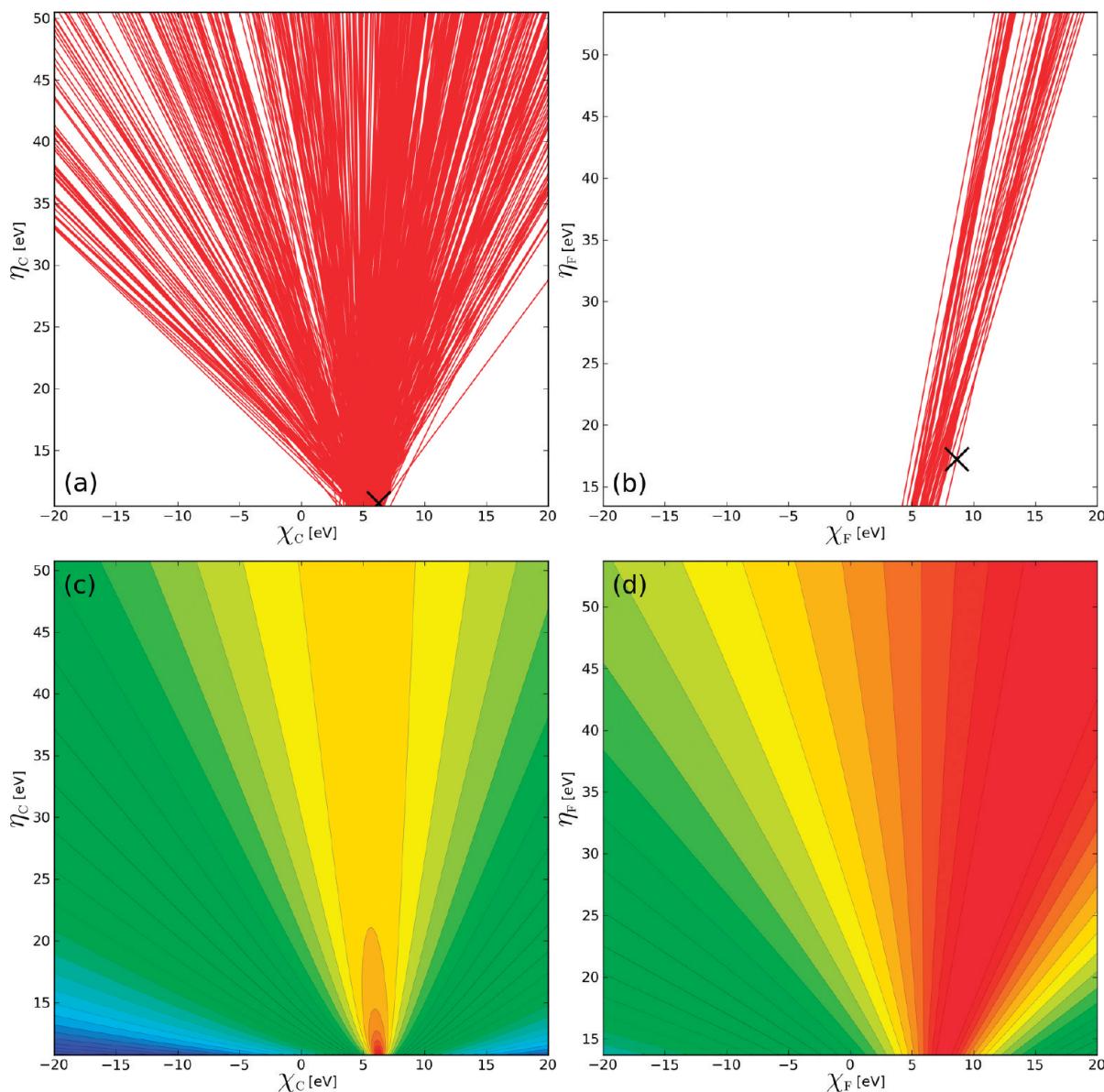


Figure 7. The lines for (a) carbon and (b) fluorine associated with the QM atomic charges in calibration P1. (See text for the definition of the lines.) The position of the optimal parameters for each element is indicated with a black cross. Parts c and d contain the corresponding contour plots of the cost function in the same planes, using the same ranges for χ_i and η_j . The color code for the contour plots is the same as in Figure 4.

This is not only useful to construct an initial guess for the parameters. It also shows why the parameters of the halogens can vary easily: all of the lines associated with fluorine (or chlorine) have approximately the same slope, and hence there is no clear intersection point. The slope of the lines is determined by the charge on the corresponding atoms, see eq 17, which implies that atoms with little variation in the partial charge are inherently difficult to parametrize.^{37,38}

Given these insights, one can think of potential improvements in the calibration procedures. (i) One could define a weight $w_{m,i}$ per element (or force-field atom type in the case of SQE) in the cost function (eq 11) that is inversely proportional to the variance on the atomic charges within the atom type.⁷¹ With this choice, atom types whose charges show little variation in the training set receive a much higher weight. This increases the curvature of the cost function, i.e., its sensitivity, for parameters

that are otherwise underdetermined. In this paper, the weights were still based on the prevalence of atomic elements or force-field atom types.³⁶ (ii) A second improvement is based on the observation that the parameter correlations are always found in pairs, i.e., the hardness and electronegativity parameters of the same element. One could use two separate cost functions: one to determine the second-order parameters (e.g., based on linear response data) and one for the linear parameters (e.g., based on equilibrium charge distributions). This is of course only helpful when both separate cost functions have a low condition number. An alternative remedy is tested in the benchmark paper of Verstraelen et al.:³⁶ an additional term is added to the cost function based on linear response data, which only depends on second-order parameters. The reduced condition number of the Hessian of such an extended cost function shows that this approach leads to less dispersed parameters.

V. CONCLUSION

Three sets of parameters for the EEM and SQE models are calibrated in this work using state-of-the art methods and resulted in three practically useful calibrations (P1, T1, and T3). These three sets of parameters are calibrated with Hirshfeld-I charges for a sufficiently large set of molecules compared to the number of parameters in the corresponding models. Three other calibrations (P2, T2, and P3) are used to illustrate the typical problems that can arise in the calibration of charge equilibration parameters. Despite the careful calibration protocols, our analysis demonstrates that the EEM and SQE parameters are still statistically ill-defined numbers that have no absolute interpretation and can only be considered as regression parameters. Irrelevant details in the training set can lead to large fluctuations on the parameters, but cross-validation tests confirm that these fluctuations do not affect the charge distribution predicted by the EEM or SQE model.

Using a variety of analysis techniques, we investigated the two potential issues that may lead to parameters that are not unique: (i) the presence and nature of multiple minima in the cost function and (ii) directions of low curvature in the minimum of the cost function. Screened electrostatics combined with suitable and reasonable lower limits on the second-order parameters restrict the calibration to a single feasible region in the parameter space. Numerical tests indicate that this region contains only one local minimum, which leaves the lack of curvature of the cost function as the only potential difficulty. The lack of curvature is present in all six calibrations in this work and seems to be a general feature, which is not surprising given the large number of parameters. In practice, this means that there is a large hyper-ellipsoid in the parameter space that contains all of the parameter vectors that are nearly optimal. Every effort to find the true minimum in this ellipsoid is completely in vain, as its exact position is very sensitive to irrelevant details in the training set. It is therefore also impossible to give a physical interpretation to individual parameters.

A certain amount of noise on the calibrated EEM and SQE parameters is unavoidable, but several measures are proposed in the paper to improve their robustness. It is compulsory to design a well-balanced training set in which each atom (and bond) type is sufficiently prevalent. On the basis of the training data, one should construct a cost function that includes a maximum of information from the training data. In addition to the atomic charges, our cost function is also sensitive to the molecular electronegativity, and one can further enrich the cost function with linear response data. One should also try to reduce the noise in the training data; e.g., it is advantageous to use Hirshfeld-I charges instead of Mulliken charges because the former are less basis set dependent. We also expect that the cost function will become better behaved when the weights associated with the charges in the cost function are defined per element (or atom type) such that they are inversely proportional to the variance on the charge of all atoms within the same atom type.

Our investigation raises many issues in the calibration of EEM and SQE parameters, and even more so in the direct interpretation of the calibrated parameters. We hope our analysis of the calibration procedures will be a fruitful source of inspiration for future work on charge equilibration models.

■ ASSOCIATED CONTENT

5 Supporting Information. The following are included:
 (1) Lewis structures of the molecules in set P and T; (2) prevalence of atomic elements and force-field atom types in

training sets P and T; (3) atomic radii used for the screened electrostatic interactions and corresponding minimum values for the atomic hardness parameters; (4) calibrated parameters for P1, T1, P2, T2, P3, and T3; (5) plots of eigenvalues (and eigenvectors) of the Hessians in T1, P2, T2, P3, and T3; (6) numerical description of the scans through the parameter space in case P1; (7) comparison of parameters of P2 with T2 and of P3 with T3 and the decomposition of the relative parameter vectors; (8) results for the covariance analysis based on the cross-validation data for calibration T2; (9) lines and contour plots similar to Figure 7, but for elements H, N, O, and Cl. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Toon.Verstraelen@UGent.be.

■ ACKNOWLEDGMENT

This work is supported by the Fund for Scientific Research—Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/6/27. Funding was also received from the European Research Council under the European Community's Seventh Framework Programme (FP7(2007-2013) ERC grant agreement number 240483). The authors would also like to thank the Ghent University for the computational resources (Stevin Supercomputer Infrastructure). P.W.A. thanks NSERC and the Canada Research Chairs for financial support.

■ REFERENCES

- (1) Politzer, P.; Murray, J. S. *Theor. Chem. Acc.* **2002**, *108*, 134–142.
- (2) Neves-Petersen, M. T.; Petersen, S. B. Protein electrostatics: A review of the equations and methods used to model electrostatic equations in biomolecules - Applications in biotechnology. In *Biotechnology Annual Review*; El-Gewely, M. R., Ed.; Elsevier: Amsterdam, The Netherlands, 2003; Vol. 9, pp 315–395.
- (3) Labat, F.; Fuchs, A. H.; Adamo, C. *J. Phys. Chem. Lett.* **2010**, *1*, 763–768.
- (4) Verstraelen, T.; Szyja, B. M.; Lesthaeghe, D.; Declerck, R.; Van Speybroeck, V.; Waroquier, M.; Jansen, A. P. J.; Aerts, A.; Follens, L. R. A.; Martens, J. A.; Kirschhock, C. E. A.; van Santen, R. A. *Top. Catal.* **2009**, *52*, 1261–1271.
- (5) Townsend, R. Ion Exchange in Zeolites. In *Introduction to Zeolite Science and Practice*; van Bekkum, H., Flanigen, E. M., Jansen, J., Eds.; Elsevier: Amsterdam, The Netherlands, 1991; Vol. 58, pp 359–390.
- (6) Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. *J. Chem. Theory Comput.* **2005**, *1*, 694–715.
- (7) McCreery, J. H.; Christoffersen, R. E.; Hall, G. G. *J. Am. Chem. Soc.* **1976**, *98*, 7191–7197.
- (8) Declerck, R.; De Sterck, B.; Verstraelen, T.; Verniest, G.; Mangelinckx, S.; Jacobs, J.; De Kimpe, N.; Waroquier, M.; Van Speybroeck, V. *Chem.—Eur. J.* **2009**, *15*, 580–584.
- (9) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. *J. Phys. Chem. A* **2005**, *109*, 3957–3959.
- (10) Matta, C. F.; Bader, R. F. W. *J. Phys. Chem. A* **2006**, *110*, 6365–6371.
- (11) Bultinck, P.; Popelier, P. Atoms in Molecules and Population Analysis. In *Chemical Reactivity Theory*; Chattaraj, P. K., Ed.; Taylor and Francis: Oxon, U. K., 2009; pp 215–227.
- (12) Politzer, P.; Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- (13) Lowdin, P. *J. Chem. Phys.* **1950**, *18*, 365–375.

- (14) Hirshfeld, F. L. *Theor. Chem. Acta*. **1977**, *44*, 129–138.
- (15) Reed, A. E.; Weinstock, R. B.; Weinhold, F. J. *Chem. Phys.* **1985**, *83*, 735–746.
- (16) Bader, R. F. W. *Phys. Rev. B* **1994**, *49*, 13348.
- (17) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbo-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.
- (18) Franci, M. M.; Chirlian, L. E. The Pluses and Minuses of Mapping Atomic Charges to Electrostatic Potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; John Wiley & Sons, Inc.: New York, 2000; Vol. 14, pp 1–31.
- (19) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (20) Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (21) Van Damme, S.; Bultinck, P.; Fias, S. *J. Chem. Theory Comput.* **2009**, *5*, 334–340.
- (22) Catak, S.; D'hooghe, M.; Verstraelen, T.; Hemelsoet, K.; Van Nieuwenhove, A.; Ha, H.; Waroquier, M.; De Kimpe, N.; Van Speybroeck, V. *J. Org. Chem.* **2010**, *75*, 4530–4541.
- (23) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. *Chem. Phys. Lett.* **2007**, *444*, 205–208.
- (24) Mortier, W. Electronegativity equalization and its applications. In *Structure & Bonding*; Kali, S., Jørgensen, C., Eds.; Springer: Germany, 1987; Vol. 66, pp 125–143.
- (25) Parr, R. G.; Yang, W. Aspects of atoms and molecules. In *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1994; pp 218–236.
- (26) Bultinck, P.; Vanholme, R.; Popelier, P. L. A.; De Proft, F.; Geerlings, P. *J. Phys. Chem. A* **2004**, *108*, 10359–10366.
- (27) Bultinck, P.; Langenaeker, W.; Carbo-Dorca, R.; Tollenaere, J. *P. J. Chem. Inf. Comput. Sci.* **2003**, *43*, 422–428.
- (28) York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 159–172.
- (29) Chelli, R.; Procacci, P.; Righini, R.; Califano, S. *J. Chem. Phys.* **1999**, *111*, 8569–8575.
- (30) Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
- (31) van Duin, A. C. T.; Strachan, A.; Stewman, S.; Zhang, Q.; Xu, X.; Goddard, W. A. *J. Phys. Chem. A* **2003**, *107*, 3803–3811.
- (32) Mortier, W.; Ghosh, S.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315–4320.
- (33) Nistor, R. A.; Poliironov, J. G.; Müser, M. H.; Mosey, N. J. *J. Chem. Phys.* **2006**, *125*, 094108.
- (34) Nistor, R. A.; Müser, M. H. *Phys. Rev. B* **2009**, *79*, 104303.
- (35) Rappe, A.; Goddard, W. *J. Phys. Chem.* **1991**, *95*, 3358–3363.
- (36) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Phys.* **2009**, *131*, 044127.
- (37) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. *P. J. Phys. Chem. A* **2002**, *106*, 7887–7894.
- (38) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. *P. J. Phys. Chem. A* **2002**, *106*, 7895–7901.
- (39) Baekelandt, B.; Mortier, W.; Lievens, J.; Schoonheydt, R. *J. Am. Chem. Soc.* **1991**, *113*, 6730–6734.
- (40) Menegon, G.; Shimizu, K.; Farah, J. P. S.; Dias, L. G.; Chaimovich, H. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5933–5936.
- (41) Chelli, R.; Procacci, P. *J. Chem. Phys.* **2002**, *117*, 9175–9189.
- (42) De Proft, F.; Langenaeker, W.; Geerlings, P. *THEOCHEM* **1995**, *339*, 45–55.
- (43) Warren, L. G.; Davis, J. E.; Patel, S. *J. Chem. Phys.* **2008**, *128*, 144110.
- (44) Parr, R.; Pearson, R. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.
- (45) Gilson, M. K.; Gilson, H. S.; Potter, M. J. *J. Chem. Inf. Model.* **2003**, *43*, 1982–1997.
- (46) Varekova, R. S.; Jirouskova, Z.; Vanek, J.; Suchomel, S.; Koca, J. *I. J. Mol. Sci.* **2007**, *8*, 572–582.
- (47) Berente, I.; Czink, E.; Náray-Szabó, G. *J. Comput. Chem.* **2007**, *28*, 1936–1942.
- (48) Njo, S. L.; Fan, J.; van de Graaf, B. *J. Mol. Cat. A* **1998**, *134*, 79–88.
- (49) Mulliken, R. S. *J. Chem. Phys.* **1934**, *2*, 782–793.
- (50) Smalø, H. S.; Åstrand, P.; Jensen, L. *J. Chem. Phys.* **2009**, *131*, 044101.
- (51) Chaves, J.; Barroso, J. M.; Bultinck, P.; Carbo-Dorca, R. *J. Chem. Inf. Model.* **2006**, *46*, 1657–1665.
- (52) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. General Linear Least Squares. In *Numerical Recipes in C: The Art of Scientific Computing*; Cambridge University Press: New York, 1992; pp 671–680.
- (53) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Model.* **2008**, *48*, 1530–1541.
- (54) Verstraelen, T.; Van Houteghem, M.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Model.* **2008**, *48*, 2414–2424.
- (55) Ghysels, A.; Verstraelen, T.; Hemelsoet, K.; Waroquier, M.; Van Speybroeck, V. *J. Chem. Inf. Model.* **2010**, *50*, 1736–1750.
- (56) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- (57) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (58) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian* 03, revision B.03; Gaussian, Inc.: Wallingford, CT, 2004.
- (59) Koopmans, T. *Physica* **1934**, *1*, 104–113.
- (60) Cordero, B.; Gomez, V.; Platero-Prats, A. E.; Reves, M.; Echeverria, J.; Cremades, E.; Barragan, F.; Alvarez, S. *Dalton Trans.* **2008**, 2832–2838.
- (61) Bultinck, P.; Carbo-Dorca, R. *J. Math. Chem.* **2003**, *34*, 67–74.
- (62) Bultinck, P.; Carbo-Dorca, R.; Langenaeker, W. *J. Chem. Phys.* **2003**, *118*, 4349–4356.
- (63) Simons, J.; Joergensen, P.; Taylor, H.; Ozment, J. *J. Phys. Chem.* **1983**, *87*, 2745–2753.
- (64) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995–4999.
- (65) Nicolay, S.; Sanejouand, Y. *Phys. Rev. Lett.* **2006**, *96*, 078104.
- (66) Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Catak, S.; Brooks, B. R.; Van Neck, D.; Waroquier, M. *J. Comput. Chem.* **2010**, *31*, 994–1007.
- (67) Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Van Neck, D.; Brooks, B. R.; Waroquier, M. *J. Chem. Theory Comput.* **2009**, *5*, 1203–1215.
- (68) De Proft, F.; Martin, J. M. L.; Geerlings, P. *Chem. Phys. Lett.* **1996**, *250*, 393–401.
- (69) Bultinck, P.; Carbo-Dorca, R. *Chem. Phys. Lett.* **2002**, *364*, 357–362.
- (70) Baekelandt, B.; Mortier, W.; Schoonheydt, R. The EEM approach to chemical hardness in molecules and solids: Fundamentals and applications. In *Chemical Hardness; Structure & Bonding*; Springer: Berlin, 1993; Vol. 80, pp 187–227.
- (71) Acke, G. *Benaderende Elektrostatische Potentialen via Atoms in Molecules en Elektronegativiteitsequalisatie*; Ghent University: Gent, Belgium, 2010.