

X-ray constrained wavefunctions based on Hirshfeld atoms. II. Reproducibility of electron densities in crystals of α -oxalic acid dihydrate

Max L. Davidson,^{a*} Simon Grabowsky^{b*} and Dylan Jayatilaka^{a*}

^aSchool of Molecular Sciences, The University of Western Australia, 35 Stirling Highway, Crawley 6009, Western Australia, Australia, and ^bDepartement für Chemie, Biochemie und Pharmazie, Universität Bern, Freiestrasse 3, 3012 Bern, Switzerland. *Correspondence e-mail: max.davidson@research.uwa.edu.au, simon.grabowsky@unibe.ch, dylan.jayatilaka@uwa.edu.au

Received 20 January 2022

Accepted 18 April 2022

Edited by M. Spackman, University of Western Australia, Australia

Keywords: X-ray constrained wavefunction; Hirshfeld atom; halting problem; X-ray wavefunction refinement; α -oxalic acid dihydrate.

CCDC reference: 2144303

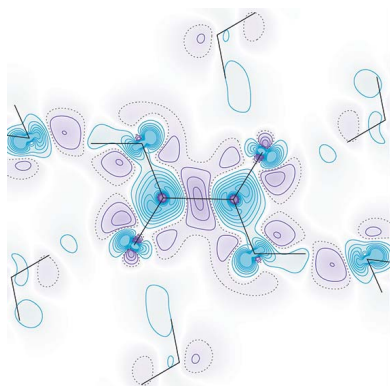
Supporting information: this article has supporting information at journals.iucr.org/b

The Hirshfeld atom-based X-ray constrained wavefunction fitting (HA-XCW) procedure is tested for its reproducibility, and the information content of the fitted wavefunction is critically assessed. Fourteen different α -oxalic acid dihydrate data sets are used for this purpose, and the first joint fitting to 12 of these data sets is reported. There are systematic features in the electron density obtained from all data sets which agree with higher level benchmark calculations, but there are also many other strong systematic features which disagree with the reference calculations, most notably those associated with the electron density near the nuclei. To enhance reproducibility, three new protocols are described and tested to address the halting problem of XCW fitting, namely: an empirical power-function method, which is useful for estimating the accuracy of the structure factor uncertainties; an asymptotic extrapolation method based on ideas from density functional theory; and a ‘conservative method’ whereby the smallest value of the regularization parameter is chosen from a series of data sets, or subsets.

1. Introduction

There has recently been much interest in the technique of X-ray constrained (or restrained) wavefunction (XCW) fitting (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001), but doubts have been raised about the accuracy, reproducibility and the meaning of the electron densities obtained from such experimental wavefunctions (Landeros-Rivera *et al.*, 2021; Macetti *et al.*, 2021; Kleemiss *et al.*, 2021; Grabowsky *et al.*, 2020; Ernst *et al.*, 2020; Genoni *et al.*, 2017). To address these concerns, and in preparation for this paper, we have therefore first reviewed the status of the XCW method, in Davidson *et al.* (2022). That review presents missing equations and missing implementation details for the widely used Hirshfeld atom X-ray constrained wavefunction (HA-XCW) method, which is also used here.

In this paper, the focus is on the original problem of testing the reproducibility of electron density features obtained from a HA-XCW procedure. We do this by applying the HA-XCW procedure to X-ray diffraction data from different experiments, that is, experiments conducted on different hardware and with different analysis protocols but using similar samples of the same compound under conditions of the same temperature and pressure. By comparing the electron densities obtained by fitting to these different data sets, the effects of errors due to different experimental set-ups can be quantified, and the current limits of reproducibility can be established.



We have chosen α -oxalic acid dihydrate in this study because 14 different data sets are available in the literature and because one of these data sets was used about 20 years ago by Grimwood & Jayatilaka (2001) when they reconstructed for the first time an experimental wavefunction for any molecular crystal (hereafter we refer to that paper as GJ2001). α -Oxalic acid dihydrate has been used for similar purposes in assessing the reproducibility of structural and atomic displacement parameters obtained from X-ray diffraction experiments (Coppens *et al.*, 1984; Sanjuan-Szklarz *et al.*, 2020).

There are good reasons to suspect that the XCW procedure may not yield experimentally reproducible features in the electron density. For example, even in the early GJ2001 work, there were doubts about whether enhanced electron density features around the two symmetry-equivalent hydrogen atoms, shown in Fig. 1, were reproducible. At the time, it was suggested that these features were due to the use of an incorrect hydrogen atom position taken from an independent atom model (IAM) refinement of only the high angle reflection data (Grimwood, 2002). In fact, later work on the ammonia crystal (Bytheway *et al.*, 2002; Capelli *et al.*, 2014) confirmed that small differences in the atomic positions and atomic displacement parameters (ADPs) do indeed strongly affect the electron densities obtained from an XCW fit; in fact, even after a HAR calculation, the experimental and theoretical structure factor magnitudes agreed well within one standard deviation, so there was actually no need to perform an XCW calculation (Bytheway *et al.*, 2002; Capelli *et al.*, 2014). Since the issue of the hydrogen atom position in α -oxalic acid dihydrate has never been addressed properly, we address it here.

Directly related to the question of reproducibility is what we have called ‘the halting problem’ in part I (Davidson *et al.*, 2022) – that is, the problem of deciding what level of agreement should be achieved between the calculated and observed

X-ray data before the fitting can be terminated. Properly addressing the halting problem requires a way to accurately characterize the errors in the estimated standard uncertainties (the so-called σ s) associated with the measured structure factor magnitudes. Errors in the σ s can be determined, statistically, if the X-ray diffraction experiment is repeated on the same sample under the same conditions, using different equipment; or by using a completely different technique that measures the same quantities. This is rarely done in practice. Therefore, unsurprisingly, the halting problem remains unsolved, but all the known attempts have been reviewed in part I (Davidson *et al.*, 2022).

It is worth clarifying that the halting problem does not arise in a least-squares refinement to model parameters, unlike in, *e.g.* a Hirshfeld atom refinement (HAR; Jayatilaka & Dittrich, 2008). However, in the XCW procedure there are usually many more parameters than data, so there is a possibility of ‘overfitting’ the model. In fact, in principle, with a sufficiently flexible basis set, an XCW should be capable of exactly reproducing the experimental data. In practice, however, convergence problems are observed if too good a fit is attempted, so that the lack of convergence is a good indicator of an attempt of overfitting the data. Lack of convergence (defined in some arbitrary way) is the most popular way to terminate the XCW procedure, but it remains an assumption. In this paper, we suggest and test three new protocols for dealing with the halting problem, which we outline in the next section.

The aims of this paper, namely to investigate the reproducibility of features in the fitted electron densities and, by implication, the halting problem, coincide with those of a project of D. Jayatilaka supported by the IUCr Commission on Charge, Spin, and Momentum Densities (now IUCr Commission on Quantum Crystallography). That project was never completed in a satisfactory way (Dacombe, 2013). The present work, all previous investigations of the halting problem, and the work of Landeros-Rivera *et al.* (2021) represent contributions to that project.

2. Three new procedures to address the halting problem

The functional minimized in XCW fitting (Jayatilaka, 1998) which determines the experimental wavefunction Ψ is

$$J[\Psi, \lambda] = E_{\text{QM}}[\Psi] - \lambda(\text{GoF}^2[\Psi] - \Delta), \quad (1)$$

where $E_{\text{QM}}[\Psi]$ is the variational energy and GoF is the goodness of fit agreement statistic between the calculated and observed experimental data [which we have previously called χ^2 , incorrectly, see part I (Davidson *et al.*, 2022)]. λ is a Lagrange multiplier parameter associated with the penalty function, the term in the parentheses. In practice, the value of λ is manually adjusted so as to achieve Δ , the desired value of GoF². It is the choice of the final value of the desired Δ (achieved via a particular value of λ) where the fitting is terminated that defines the halting problem. In what follows, we refer to λ_{opt} as a value of λ estimated by any method to halt

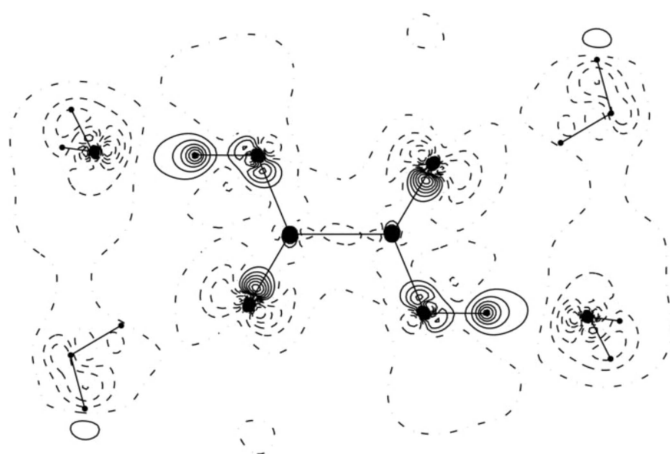


Figure 1

The change in the electron density of α -oxalic acid dihydrate due to the X-ray constrained fitting procedure. Static electron density difference map between the constrained and the unconstrained Hartree–Fock wavefunctions. Copyright IUCr, from Grimwood & Jayatilaka (2001). Contours are in increments of $0.1 \text{ e} \text{ \AA}^{-3}$, dashed lines are a decrease while solid lines are an increase.

the XCW fitting process. More details about the meaning and derivation of any of the parameters used in XCW fitting are given in part I (Davidson *et al.*, 2022).

2.1. The power function extrapolation procedure

We observed that log-log plots of GoF^2 versus λ from the XCW procedure on experimental data were linear, indicating that a good fit could be obtained by utilizing a power function form for GoF^2 :

$$\text{GoF}^2(\lambda) = A\lambda^B. \quad (2)$$

The quality of these fits (see §S.1) is remarkable, especially when one considers that for $B < 0$ (which is required in order to reproduce the observed monotonically decreasing curves) the power function goes to infinity as $\lambda \rightarrow 0$. It is also remarkable because, as λ goes to infinity, the power function goes to zero, something which is clearly *not* true based on a perusal of the actual data. The remarkable fit is significant because, by contrast, an assumed exponential form did *not* fit the data nearly as well, see for example Fig. S4 in the supporting information.

Alternatively, a [1/1]-Padé approximant

$$\text{GoF}^2(\lambda) = \frac{a + b\lambda}{1 + c\lambda} \quad (3)$$

would more accurately reflect the behaviour of the $\text{GoF}^2(\lambda)$ function at high and low λ values. The fit does indeed fit the data remarkably well as shown in Fig. S5 in the supporting information. However, in part I (Davidson *et al.*, 2022), we showed that GoF^2 is an even function, but both the power function fit and [1/1]-Padé approximant are not inherently even functions. If λ is simply squared to produce an even function [*i.e.* $\text{GoF}^2(\lambda^2)$], neither function now fits the data well. Despite this, both the Padé approximant and the power function fit the data extraordinarily well. In this paper, we only pursue the power function fit, but the Padé approximant fit certainly warrants further investigation in a future study.

Concerning the power function, we observed that the A parameter is not clearly normally distributed (see Fig. S6 in the supporting information), suggesting that it may be a scale dependent quantity. Indeed, from a theoretical perspective, if we define a ‘corrected’ GoF^2 by

$$\text{GoF}_{\text{corr}}^2(\lambda) = \frac{\text{GoF}^2(\lambda)}{A}, \quad (4)$$

then plots of $\text{GoF}_{\text{corr}}^2$ versus λ become very similar across fits for different data sets of the same compound. Indeed, they are mathematically required to intersect at $\text{GoF}_{\text{corr}}^2(1) = 1$ where $\text{GoF}^2 = A$, so that scaling all the standard uncertainties (σ s) by $(A)^{1/2}$ would mean that at $\lambda = 1$ all the different experiments have a common variance relative to the experimental data.

Based on this, the first termination method we propose to investigate is the one where λ is chosen to be this common intersection point; that is, we test terminating the fitting procedure at the value of λ equal to

$$\lambda_{\text{opt}} \equiv \lambda_{\text{pow}} = 1, \quad \text{where} \quad \text{GoF}^2 = A. \quad (5)$$

The value of $\lambda = 1$ has an added significance: at $\lambda = 1$, the effective force of the constraint due to the quantum mechanical energy is the same as that due to the minimization of the GoF^2 term, that is

$$\frac{\partial J}{\partial c_{vi}^*} = 0 \implies \lambda = \frac{\partial E'_{\text{QM}}}{\partial c_{vi}^*} \left[\frac{\partial \text{GoF}^2}{\partial c_{vi}^*} \right]^{-1} \quad (6)$$

with c_{vi} being the molecular orbital coefficients. Thus, when $\lambda = 1$ the magnitudes of these two derivatives, or effective forces, are the same, so the information coming from the quantum mechanical constraint and the least-squares fitting may be said to be similar. This argument is not a very strong one, because the units of these effective forces are different.

2.2. Modified Tozer–Ingamells–Handy (TIH) asymptotic extrapolation procedure

A second method of obtaining an optimal value of λ to stop the fitting process may be adapted from the work of Zhao & Parr (1993) and Tozer *et al.* (1996), who exhibit methods to obtain the Kohn–Sham wavefunction from theoretical electron densities defined on a set of grid points in real space. As was shown in part I (Davidson *et al.*, 2022), the constraint term C defined in equation (4) of the paper by Zhao & Parr (1993) used to obtain the constrained-fitted wavefunction is exactly the same as our GoF^2 term, if the values of the σ s are appropriately interpreted. Therefore, the XCW procedure used by us differs only in the fact that we choose to obtain constrained-fitted wavefunctions using structure factor magnitudes, which are essentially electron density values defined on a grid in reciprocal space, rather than real space.

As a matter of fact, Tozer *et al.* (1996) have shown that the convergence problems observed by Zhao & Parr (1993) when attempting to obtain the Kohn–Sham wavefunction from theoretical electron densities are due to incompleteness in the basis sets used to obtain the constrained-fitted wavefunctions. In the XCW method, we have also observed the same convergence problems. Hence, since our XCW method is essentially the same as theirs, and since we also use a finite basis set, our convergence problems must at least in part be due to basis-set incompleteness. However, in addition, the structure factor magnitudes that we fit to in the XCW procedure are subject to experimental errors, and while these errors are conceptually no different to the errors which are present in the electron density grids used by Zhao & Parr (1993) and (Tozer *et al.*, 1996), these errors are likely much larger. Therefore, we expect that these experimental errors will exacerbate any convergence problems (this is confirmed in §4.1.1 with a simple test).

Zhao & Parr (1993) used the fact that their constraint term C [which corresponds to our $\text{GoF}^2(\lambda)$] is an even-function asymptotic expansion in λ to extrapolate the total energy and kinetic energy. In part I (Davidson *et al.*, 2022), we likewise showed that $\text{GoF}^2(\lambda)$ is also an even function of λ . On the other hand, Tozer *et al.* (1996) argued that their expansion lacked a constant term representing the error term. Therefore,

following the latter authors, we also assume an asymptotic form

$$\text{GoF}^2(\lambda) = D + E\lambda^{-2} + F\lambda^{-4} + \dots \quad (7)$$

where the term D represents the error term. Like Tozer *et al.* (1996), we now desire to choose an optimum value of $\lambda = \lambda_{\text{opt}}$ which is not so large that the constant error term D dominates, *i.e.* we want

$$D < E\lambda^{-2} \text{ or } \lambda < \lambda_{\text{upper}} \equiv \left| \frac{E}{D} \right|^{1/2}, \quad (8)$$

and not so small that the smaller, higher-order asymptotic term dominates, *i.e.* we want

$$F\lambda^{-4} < E\lambda^{-2} \text{ or } \lambda > \lambda_{\text{lower}} \equiv \left| \frac{F}{E} \right|^{1/2}. \quad (9)$$

The optimum value of λ is thus determined when these two terms are equal, *i.e.* when $E = \lambda^2 D = F/\lambda^2$, or more specifically when

$$\lambda_{\text{opt}} \equiv \lambda_{\text{TIH}} = \left| \frac{F}{D} \right|^{1/4}. \quad (10)$$

If $\lambda_{\text{lower}} < \lambda < \lambda_{\text{upper}}$ then we say that we are in the asymptotic region of the $\text{GoF}^2(\lambda)$ curve. We should expect the coefficients $|D|$, $|E|$ and $|F|$ form a monotonically decreasing sequence, and this is what we indeed observe. However, in all cases $\lambda_{\text{lower}} > \lambda_{\text{upper}}$, which means we are not strictly in the asymptotic region.

We have also observed that the value of F may be negative, which is unusual for an asymptotic form. Tozer *et al.* (1996) have also seen such negative values, and say that ‘it is not surprising that $[F]$ has become negative, because for this smaller range of three λ values, more terms in the [asymptotic] series ... should be used’ and those authors use only the first (basis set error) condition, equation (8), which in our opinion is a tacit acknowledgement that this term is more important. We also do the same. In any case, the objective here is to extract the ‘best’ value for E which avoids overfitting, which in this case means ensuring that the D term is not dominant.

We have determined the terms D , E , and F by fitting to either the last 3 (λ_{TIH3}) or the last 6 (λ_{TIH6}) data points in the GoF^2 versus λ plots to check if the two GoF^2 values determined in these two ways are similar enough, say within a few standard deviations of the least squares error in the TIH6 fit to the data (there are no errors in the TIH3 method since these parameters are exactly determined by three data points). That is, a strong difference in the GoF^2 values obtained from these different fits would be another indication of problems with the asymptotic expansion itself. We always fit to the data points corresponding to the largest values of λ where convergence was possible because these are the points closest to the asymptote. These choices are subjective.

2.3. A conservative method to halt fits when using multiple data sets

We do not have any reason why the power function method and the asymptotic expansion method described in the previous two subsections should produce similar halting points for a general system. Indeed, the two proposed methods for estimating λ_{opt} are incompatible: the asymptotic method relies on the fact that $\text{GoF}^2(\lambda)$ is an even function of λ , whereas the power function method is not even, and as mentioned, does not have the correct asymptotic behaviour at $\lambda \rightarrow \infty$, or at $\lambda = 0$.

Therefore, we suggest a third method to solve the halting problem, which is, however, only applicable if multiple data sets are available as in this study, or if multiple subsets can be produced in a meaningful way. We term it the ‘conservative halting method’ for λ_{opt} because it consists in using the minimum of all λ_{opt} values,

$$\lambda_{\text{pen}}^{\text{min}} = \min_i [\lambda_{\text{pen}}(\text{oxai})], \quad (11)$$

and

$$\lambda_{\text{TIH6}}^{\text{min}} = \min_i [\lambda_{\text{TIH6}}(\text{oxai})]. \quad (12)$$

λ_{pen} is the penultimate value of λ before convergence ceases. *oxai* refers to any of the 14 different oxalic acid data sets. λ_{TIH6} was introduced in the previous subsection.

For a general application, the method described in this subsection suggests to break up a given data set into different subsets, determine $\lambda_{\text{opt}}^{\text{min}}$ for each subset, and then use the above formulae. The reflections in each subset could be selected from the set of repeated measurements for a given reflection, for example, in a kind of bootstrap procedure (Efron & Tibshirani, 1994; Vetterling *et al.*, 2002). This idea is quite similar to the one used in the protein R_{free} method (Brünger, 1992). It is also similar to suggestions by Sanjuan-Szklarz *et al.* (2020) and Landeros-Rivera *et al.* (2021), to break data sets into low and high resolution parts. These approaches are not investigated further in this paper but may be of use in the future.

3. Methods and experimental data

3.1. The Zobel (1996) data

We used the data of oxalic acid from GJ2001, which were taken from the appendix of the PhD thesis of Grimwood (2002). The description in the thesis is a little better than that from the original paper, and it states:

‘The structure factor data were supplied privately by Zobel (1996), and were taken at 15 K using 0.71069 Å radiation. This set is different to the one they deposited to the IUCr (Zobel *et al.*, 1992) ... The experimental errors for the newer data are on average 1.4 times smaller than for the published data. However, the newer data is only a subset of the published data in terms of the number of reflections’.

In the GJ2001 paper, only low-angle reflections with $\sin\theta/\lambda < 0.71$ Å were used, and weak intensity reflections with $|F| < 2\sigma$ were removed. For this paper, for consistency with the other

data sets, we further pruned reflections by rejecting weak reflections with $|F| < 4\sigma$, leaving a total of 560 reflections. Genoni *et al.* (2017) showed that weak high order reflections hamper the extraction of the electron correlation effects from the structure factors. Like GJ2001, we refined the hydrogen atoms using isotropic ADPs.

The IAM geometry present in the crystallographic information file (CIF) from Zobel *et al.* (1992) was used to initiate the HAR (Capelli *et al.*, 2014; Jayatilaka & Dittrich, 2008) in this paper. The CIF was obtained from the Cambridge Structural Database (CSD). Whereas we use the HAR-derived positions and ADPs for the XCW fitting in this study, GJ2001 used the IAM geometry in their XCW fittings. These positions and ADPs were refined from the high angle part of the original Zobel *et al.* (1992) data. In his PhD thesis, Grimwood reports that the reason for this was because these gave structure factors in better agreement with the experimental values than when using the multipole-model-refined parameters on the high-angle part of the data set, when using a Hartree–Fock wavefunction.

3.2. The Kamiński *et al.* (2014) data

The 13 data sets of Kamiński *et al.* (2014) on α -oxalic acid were all collected at 100 K on three different experimental setups:

- (i) Data sets oxa1-oxa7, and oxa11-oxa12 were collected on a multi-layer optics Bruker AXS instrument.
- (ii) Data sets oxa8-oxa10 were collected with a Bruker AXS instrument with single monochromator optics.
- (iii) Data set oxa13 was collected with a KUMA instrument.

The CIFs and F^2 structure factors are available as supporting information on the *IUCrJ* webpage where the original publication appears. According to Wozniak (2021), these data sets are not to be regarded as of the usual charge density quality, but were collected to calibrate new instruments over several years. Nevertheless, such a large set of repeated measurements of the same crystal type over many machines and many years is rare, so these data are valuable, and have been widely used for assessment purposes (Kamiński *et al.*, 2014; Sanjuan-Szklarz *et al.*, 2020).

Further comments on these data are as follows

- (i) *Resolution.* The resolution of these data sets was above 1.13 \AA^{-1} except for data sets oxa10 (1.00 \AA^{-1}), and oxa11 (1.03 \AA^{-1}).
- (ii) *Completeness.* All the data sets had completeness equal to or greater than 95% except for data sets oxa4 (91%), oxa8 (89%) and oxa10 (85%).
- (iii) *Internal agreement.* The internal agreement factors R_{int} were equal or less than 2.60% except for data sets oxa2 (4.87%), oxa5 (3.83%) and oxa9 (4.61%).

For further details, consult Table 1 in Kamiński *et al.* (2014).

3.3. Choice of system for model wavefunction

The asymmetric unit of the crystal consists of only half an oxalic acid plus one water molecule [Fig. 2(b)]. However, we chose a cluster of molecules comprising a central oxalic acid

moiety with the closest six surrounding water molecules, as shown in Fig. 2(a). Note that this model system is larger than any of those used in the GJ2001 work to ensure that crystal field effects on the central moiety are properly accounted for here and to be consistent with the work of Sanjuan-Szklarz *et al.* (2020). Point charges and dipoles surrounding this cluster were not considered initially (see §5).

3.4. Wavefunction methods

All HAR and XCW fittings were based on Restricted Hartree–Fock (RHF) wavefunctions. The effect of electron correlation was assessed with DFT calculations using B3LYP and PBE0 functionals [see *e.g.* Parr & Yang (1994)] and with coupled-cluster calculations including singles and doubles excitations (CCSD) (Helgaker *et al.*, 2014), but without the frozen core approximation. All DFT and coupled-cluster calculations were single point based on the fixed HAR geometry of the oxa11 data set.

3.5. Basis sets

Results in this paper are based on the Karlsruhe def2-SVP basis set functions (Weigend & Ahlrichs, 2005), known to be sufficient to provide HAR parameters in good agreement with those from neutron diffraction experiments (Fugel *et al.*, 2018). Cartesian Gaussian basis functions were used throughout. We did not use a triple-zeta basis set in order to limit the fitting flexibility, and thus mitigate the danger of overfitting.

3.6. Software

Reference DFT and CCSD calculations were performed using the *Gaussian09* program (Frisch, 2009).

HAR and HA-XCW procedures were performed using the *Tonto* software package (Jayatilaka & Grimwood, 2003). After the last step of HAR and equivalently at the super-

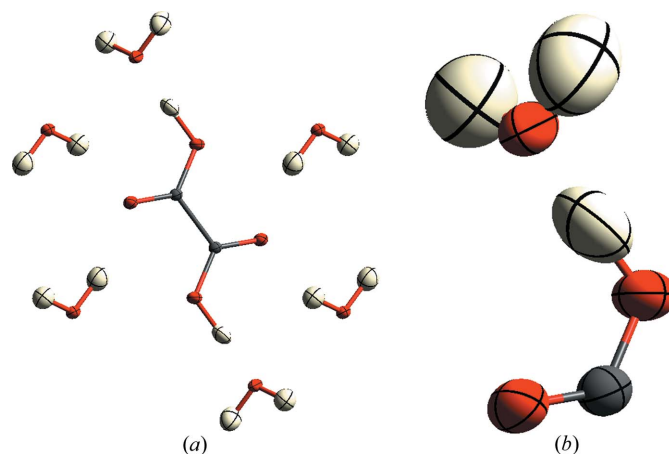


Figure 2
Structures obtained by the combined HAR on 12 Kamiński *et al.* (2014) data sets as detailed in §4.2.4. (a) The model system for α -oxalic acid used to construct the wavefunction comprising a complete central oxalic acid moiety surrounded by six water molecules. The ADPs are plotted at the 50% probability contour level. (b) Subset of (a): ORTEP plot of the symmetry unique atoms using the 99% probability contour level.

turbed $\lambda = 0$ step at the outset of XCW fitting, a HF single-point wavefunction (via MOs) was saved, for which *Tonto* calculated electron density grid files. The same was done at each λ step during the fitting procedure.

For the DFT and CCSD calculations, formatted checkpoint (FChk) files from *Gaussian09* were produced and *Tonto* was used to read in and generate electron density grid files. In the case of CCSD, natural orbitals were calculated in *Tonto* for this purpose.

Contour plots were generated with the program *gnuplot* version 5.2 (Williams *et al.*, 2019) using a script written by the *Tonto* program (Jayatilaka & Grimwood, 2003).

Plots of the oxalic acid system with ADPs were made using the *CrystalExplorer 21* software (Spackman *et al.*, 2021).

3.7. Contour plots

Electron densities are represented as (positive) number densities, throughout.

Plots represent the differences of static electron densities at a chosen maximum or optimum λ -value minus the electron density from the unfitted wavefunction ($\lambda = 0$, no physical effect beyond the HF ansatz), *i.e.* they show the change from the unfitted wavefunction due to the constraint. The difference between contour levels is always $0.025 \text{ e } \text{\AA}^{-3}$. In all plots for this paper, blue indicates a positive contour value, a relative increase in the number of electrons, while lilac represents a negative contour value, a relative decrease in the number of electrons.

To understand the effect of electron correlation in isolation, contour plots of the CCSD or DFT electron densities relative to (subtracting) the corresponding HF electron densities are shown at the fixed geometry from the oxa11 HAR calculation. (The geometric changes between different HARs are too small to affect the resulting CCSD and DFT calculations used in the qualitative comparisons that we used these data for). In the case of CCSD, we used the relaxed density matrix formalism (Frisch, 2009).

3.8. Details of HAR

$|F|^2$ reflection data was reduced to $|F|$ data by setting $|F| = (|F|^2)^{1/2}$ and at the same time $\sigma(|F|)$ was set to the larger of the two values of $[|F|^2 \pm \sigma(|F|^2)]^{1/2} - (|F|^2)^{1/2}$. HAR included all $|F| > 4\sigma(|F|)$. Extinction was not modelled. The weighting scheme used in HAR in *Tonto* is $1/\sigma(|F|)$. A scale factor was calculated by least squares by minimizing GoF^2 separately and after the refinement of the positions and ADPs in each HAR cycle, *i.e.* the scale is not optimized at the same time as all the other parameters. The reason is that the scale needs to be refined even in the XCW procedure for the molecular orbital parameters, which is not a least-squares process.

The geometrical and ADP parameters in the CIFs were used as starting parameters for HAR. Note that in our HAR CIFs, more than an asymmetric unit of atoms is provided [Fig. 2(a)], and this corresponds to the system used to calculate the wavefunction Φ . Moreover, we use the convention that the first symmetry-unique atoms listed in the CIF are those which

are actually refined. (Recall that some atoms in the system, *e.g.* extra water molecules, are used only as ‘buffer’ atoms to provide a good approximation to the crystalline environment, and they are not refined but defined by symmetry from those earlier in the list). Residual density maps for these HARs are shown in Fig. S19.

The resulting geometry and ADPs for a simultaneous HAR on all Kamiński *et al.* (2014) data sets is shown in Fig. 2(b). A single model was refined from 12 of the Kamiński *et al.* (2014) data sets. Data set oxa2 was excluded from the joint fitting because its figures of merit (Table S6) and difference density distribution (Fig. S13) showed it to be an outlier. For the other 12 data sets, each set of reflections was assigned to a different group, and a scale factor for each group was least-squares refined. The optimum scale factors were all close to unity and are reported in §S.8 of the supporting information. A new procedure was implemented in *Tonto* to calculate the residual electron density from the single calculated structure factor set and the 12 experimental data sets for the joint refinement. The corresponding residual density maps are shown in Fig. S20. The CIF of this joint refinement has also been deposited with the Cambridge Structural Database and can be obtained under CCDC-2144303 from <https://www.ccdc.cam.ac.uk/structures/>.

3.9. Outlier data from HAR

Our experience is that the GoF^2 value and the convergence characteristics of the XCW procedure are dominated by only a handful of outliers. The quantile-quantile (Q-Q) plots (see the supporting information, Figs. S8 and S9) show that the data are not normally distributed, and have ‘fat tails’. The distributions are also sometimes highly skewed. One then has to decide whether these differences are due to insufficiencies of the model, or errors in the experiment. In the former case, there would be important information in the experiment which is not being modelled.

We take the view that the errors are more likely in the experiment than in the model. One reason is that the HAR model produces positions and ADPs in agreement with independent results from neutron diffraction, even for hydrogen atoms (Woińska *et al.*, 2016). Therefore, the model is accurate and reliable enough to reproduce coordinates, ADPs and the static electron density distribution. More indirectly, the quantum chemical models have been tested against a huge range and variety of experimental data, see Goerigk *et al.* (2017) for one example. By contrast, we have few data allowing corresponding independent experimental verification for non-synchrotron X-ray diffraction experiments, such as the data at hand here. Besides, it is well known that outliers occur far more frequently in experiments than is predicted by the normal distribution, due to the fact that the rate of convergence of the experimental distribution to a normal distribution varies greatly according to how far away the data are from the mean – the convergence is much slower further away from the mean (Sherman, 1971). Therefore, outliers are expected in the experimental data.

To mitigate the effects of outliers, our approach is therefore to only remove a few of the very worst reflections from all data sets, after performing an initial HAR as described above. Outlier reflections r were decided based on the value of $(|F_r^{\text{obs}}| - |F_r^{\text{calc}}|)/\sigma(|F_r^{\text{obs}}|)$, removing only a handful while trying to maintain a symmetric distribution. $(|F_r^{\text{obs}}| - |F_r^{\text{calc}}|)/\sigma(|F_r^{\text{obs}}|)$ values of the removed outliers are reported in the supporting information in §S.11 in Table S4. After removal of the outliers, the ADPs of the hydrogen atoms in several of the data sets became far less 'flat', as discussed below.

Table 1 gives HAR agreement statistics before and after outlier removal. As expected, some GoF^2 values change drastically. Before, outlier removal data set oxa11 had the lowest R_1 value, data set oxa13 had the lowest wR_2 value, while data set oxa2 had the lowest GoF^2 value. After removal, data set oxa2 still had the lowest GoF^2 value of 0.994. Data set oxa11 had the lowest R_1 and the lowest wR_2 values.

The number of reflections removed from each data set can be seen in Table 1. Before removal of outliers, the ADPs of the hydroxyl H atoms obtained with data sets oxa2–oxa4, oxa6–oxa8, and oxa11 were unreasonably flat. No hydroxyl H atoms were unreasonably flat after removal. The removal also caused subtle changes to the hydroxyl O–H bond lengths: most changes were 0.005 Å or smaller; however, data sets oxa5, oxa7, oxa8, oxa10 and oxa13 all had decreases between 0.010–0.015 Å; data set oxa11 exhibited an increase of 0.014 Å. For the Zobel data, when hydrogen atoms were refined anisotropically (as mentioned, they were not in the later studies) then the ADP for the hydrogen-bonded H atom became non-positive definite.

3.10. HAR results for the Kamiński *et al.* (2014) data

Abrahams-Keve (or Q-Q) plots are shown for each data set in Figs. S8–S9 in §S.9 of the supporting information. Most of these show the typical non-Gaussian fat tails and outliers, even after removal of the handful of the worst ones. Particularly bad tails are observed on the data sets oxa1, oxa3, oxa5–oxa7, and oxa11–oxa12. Perhaps more important is a clearly skewed distribution (not including extreme outliers) in data sets oxa1, oxa3, oxa4, oxa6!, oxa11, and oxa12! (with exclamation marks indicating particularly strong effects). Some of the data sets have very large outliers which seem to be associated with some systematic effects.

The ADPs of the asymmetric unit atoms from the combined refinement of all the data sets (except data set oxa2) are shown in Fig. 2(b). The hydrogen ADPs are positive definite and not obviously wrong. Plots obtained from each outlier-pruned data set are separately shown in Figs. S10–S11 in §S.10 of the supporting information.

HAR on the data sets produces R_1 values between 1.2–2.1% which means at worst a 98% agreement with experimental structure factors. This means we are fitting in the absolute limit of information contents of the data with the XCW procedure after HAR is performed.

Table 1

Refinement figures of merit for HAR using a RHF/def2-SVP wavefunction for the Kamiński (2014) data sets, before and after removal (A.R.) of outlier reflections.

The removed reflections are listed in Table S4.

Data set	GoF ²	R_1	wR_2	No. of reflections removed	GoF ² A.R.	R_1 A.R.	wR_2 A.R.
oxa1	2.494	0.0171	0.0485	11	1.769	0.0161	0.0430
oxa2	1.158	0.0209	0.0436	2	0.994	0.0205	0.0422
oxa3	2.671	0.0153	0.0453	9	1.496	0.0143	0.0390
oxa4	1.426	0.0171	0.0418	3	1.215	0.0167	0.0402
oxa5	2.940	0.0196	0.0521	11	2.012	0.0184	0.0472
oxa6	2.210	0.0158	0.0426	11	1.591	0.0153	0.0385
oxa7	2.090	0.0162	0.0441	7	1.544	0.0150	0.0387
oxa8	1.274	0.0202	0.0455	3	1.223	0.0198	0.0446
oxa9	1.439	0.0209	0.0510	4	1.391	0.0208	0.0502
oxa10	1.320	0.0191	0.0449	3	1.256	0.0187	0.0438
oxa11	3.681	0.0138	0.0469	20	2.128	0.0124	0.0375
oxa12	2.230	0.0159	0.0438	11	1.762	0.0154	0.0406
oxa13	1.418	0.0172	0.0397	5	1.311	0.0163	0.0382

Table 2

λ_{max} values achievable after the standard uncertainties $\sigma_{112}^{\text{obs}}$ with $|F_{112}^{\text{obs}}| = 1613.70$ and $\sigma_{11-1}^{\text{obs}}$ with $|F_{11-1}^{\text{obs}}| = 3397.43$ are artificially scaled by factor η .

Using $\eta = 0.01$ already failed to converge in the first iteration ($\lambda = 0.01$).

Scale factor η	$\eta \sigma_{112}$	$\eta \sigma_{11-1}$	λ_{max}
1.00	3.340	17.580	0.19
0.10	0.334	1.758	0.13
0.05	0.167	0.879	0.09
0.01	0.033	0.176	–

3.11. Details of HA-XCW

XCW fittings were performed using successively larger values of λ as described below. At the first step at $\lambda = 0$, the XCW was started from the molecular orbitals (MOs) of the previous HAR. After this, the procedure continues using the MOs from the previously converged step. Direct inversion of the iterative subspace (DIIS) extrapolation was used (Pulay, 1980, 1982) keeping up to eight previous sets of MO vectors, and the procedure saving from the first iteration. If an MO vector with smaller gradient was not found by the DIIS procedure, it resets keeping the MOs with the lowest error. Density matrix damping and level shifting was used to stabilize oscillations for the first 3 iterations, with a damping factor of 0.5 (that is, keeping 50% of the density matrix from the previous iteration), and a level shift of 0.3 atomic units (au). Convergence was achieved when the DIIS error vector magnitude was less than 0.001 au which in every case resulted in successive energies being within 0.00001 au of each other or less. A calculation was deemed not to converge if it took more than 100 cycles of the above process.

3.12. HA-XCW λ scan protocols to obtain λ_{max}

The purpose of this protocol is to obtain reasonable XCWs for a given λ in the most economical way. The intent is to ensure that the XCW changes smoothly from the unfitted wavefunction to the constrained wavefunction, minimizing the

Table 3

GoF² agreement statistics for different XCW termination models, and for the IAM TCPDF result from GJ2001.

Note that GJ2001 used a model with only four surrounding water molecules, whereas this work involves six; see §3.3 for more details. Parameters for the different halting methods are also given, including halting values $\lambda = \lambda_{\text{opt}}$; power function parameters A , and B ; and asymptotic expansion coefficients D , E , and F ; see §2 for details. Missing values indicate that it was not possible to converge the XCW to the halting value. Errors in the parameters are determined by the least-squares method in cases where there are more data than parameters.

Model	λ	GoF ²	A	B	D	E	F
IAM TCPDF XCW							
λ_{max}	0.30	3.4	–	–	–	–	–
HA-XCW							
λ (HAR)	0.00	3.5	–	–	–	–	–
λ_{pow}	1.00	–	1.344 (6)	–0.159 (1)	–	–	–
λ_{TIH3}	0.03	2.36	–	–	1.64	0.0013	0.000067
λ_{TIH6}	0.08	2.02	–	–	1.53 (1)	0.0094 (6)	–0.000066 (8)
λ_{pen}	0.18	1.75	–	–	–	–	–
λ_{max}	0.19	1.74	–	–	–	–	–

possibility of spurious solutions which might be found if one used any larger value of λ directly from $\lambda = 0$. This λ scan procedure ensures a continuous set of XCWs $\Phi(\lambda)$.

To find λ_{max} , the basic idea is to start with a small value, and then to quickly increase that value by factors of 10 until convergence fails, then backtrack to the previous largest value and increase more slowly in linear increments. In more detail, start with $\lambda = 10^{-n}$, with $n = 4$. Then n was decreased in unit steps until convergence failed at n_{fail} . Every XCW procedure was started with the density matrix from the previously converged step. At this point, the procedure was restarted using the last value of λ that converged, $\lambda = 10^{-(n_{\text{fail}}+1)}$, and then increased linearly in these step sizes $[10^{-(n_{\text{fail}}+1)}]$, until convergence fails at $\lambda = k10^{-(n_{\text{fail}}+1)} = \lambda_{\text{max}}$. The penultimate converging value $\lambda_{\text{pen}} = (k-1)10^{-(n_{\text{fail}}+1)}$ was used to generate contour plots.

3.13. HA-XCW λ scan protocols to obtain λ_{opt}

To get a series of λ values for the power function and asymptotic extrapolation procedures for the halting problem, the following procedure was used

(i) For the Zobel data set, λ was increased from zero in increments of 0.01.

(ii) For the Kamiński *et al.* (2014) data sets, λ was increased from zero in increments of 0.02, 0.1, or 0.2 for λ in the ranges $0 \leq \lambda \leq 1.2$, $1.2 \leq \lambda \leq 2$, and $2 \leq \lambda \leq 4$, respectively.

4. Results

4.1. HA-XCW for the Zobel data

4.1.1. The dependence of XCW fitting on standard uncertainties. To test the sensitivity of the XCW fitting procedure to incorrect or too small standard uncertainties (σ) values, we fitted to the Zobel (1996) data set, but adjusted the σ values on the two intense reflections (1 1 2) and (1 1 –1) by scaling them by a factor η between 1 and 0.01 and investigated the value of λ_{max} in each case. The intent of this procedure is to artificially produce what we referred to as incompatible measurements in

part I (Davidson *et al.*, 2022). The results are included in Table 2. As the $\eta\sigma$ values decrease, so too does the value of λ_{max} , indicating that failure to converge occurs at lower penalty values. Indeed, using a scale factor of 0.01, no convergence could be achieved anymore for any value of λ . We conclude that the XCW fitting procedure depends on the σ values provided, as adjusting the values of only two out of over 560 reflections affects the convergence behaviour.

4.1.2. Agreement statistics. Table 3 shows the GoF² agreement statistic at the λ values for the different halting procedures. $\lambda = 0$ refers to the HAR. The result from GJ2001 is shown under the heading IAM TCPDF XCW.

HAR produces a GoF² value of 3.5, very similar to the GoF² of 3.4 from GJ2001 *after* fitting. In the original work, various molecular clusters were considered, up to and including four water molecules, and the GoF² value before constraining the orbitals were all in excess of GoF² = 15.2. Using the same reflections as in the original work (582 reflections), GoF² = 9.0 before fitting. The system used in this work involves six water molecules, representing an extra two molecules in the periphery, but it is unlikely that these additional molecules could explain such a large discrepancy. The improvement in the GoF² value is probably due instead to the optimized atomic positions and ADPs from HAR.

For all the final XCWs for which the halting procedures worked (see next subsection) a reduction of the GoF² from 3.5 to less than 2 was obtained. This corresponds to a 30% reduction of the σ -normalized deviations from the experimental structure factor magnitudes, indicating that there is some information in the experimental data which is not modelled by HAR alone.

4.1.3. Halting procedures. Fig. 3 shows the GoF² as a function of λ plot, as well as the fitted power function of equation (2), with values of $A = 1.344$ (6) and $B = -0.159$ (1).

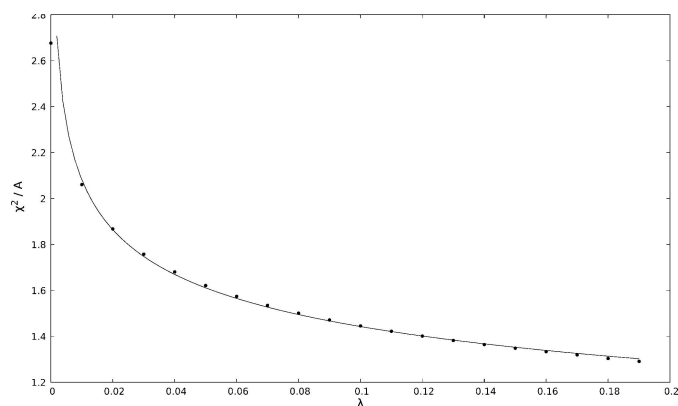


Figure 3
GoF² versus λ for a HA-XCW using a RHF/def2-SVP wavefunction for the Zobel (1996) data set, with a fitted power function $A\lambda^B$, with $A = 1.344$ (6), and $B = -0.159$ (1).

The plot and the errors in these parameters show that this function fits very well, although some of the GoF^2 values are slightly underestimated at small λ and slightly overestimated as λ increases. This behaviour is reminiscent of the observation of Genoni (2013) regarding an inflection point in the GoF^2 versus λ curve. It was not possible to reach the halting point $\lambda_{\text{pow}} = 1$ at $\text{GoF}^2 = A = 1.344$.

The 3- and 6-point TIH methods were also used to estimate λ_{opt} . The corresponding fitting parameters are also given in Table 3. We observe that $|D| > |E| > |F|$ but for the TIH6 method, the value of F is negative, and we establish that the curve is not in the asymptotic region. Therefore, we use the value $\lambda_{\text{TIH}} = \lambda_{\text{upper}}$ to halt the XCW procedure for the reasons given in §2.2. Although the values from the 3- and 6-point fits are not consistent to each other within the error bars, the values of λ_{opt} and the corresponding GoF^2 values are similar, but they are not at all close to $\lambda = 1$, the termination point for the power-function method.

4.1.4. Effect of the constraint on the electron density. Fig. 4 shows maps of the constrained minus the unconstrained electron density, for three different λ_{opt} values. Lilac regions depict a loss of electron density caused by the fitting process, e.g. in the centre of the C—C bond, whereas blue areas depict a build-up of electron density upon fitting, e.g. around the non-hydrogen atomic cores. The features in all three plots are essentially the same, but become more exaggerated (by at most one contour level) in Figs 4(a) to 4(d). Some features become more delocalized as λ increases (e.g. the depletion on the hydroxyl oxygen atom). From here on we only present plots for $\lambda = \lambda_{\text{TIH6}}$, because the key features are present and

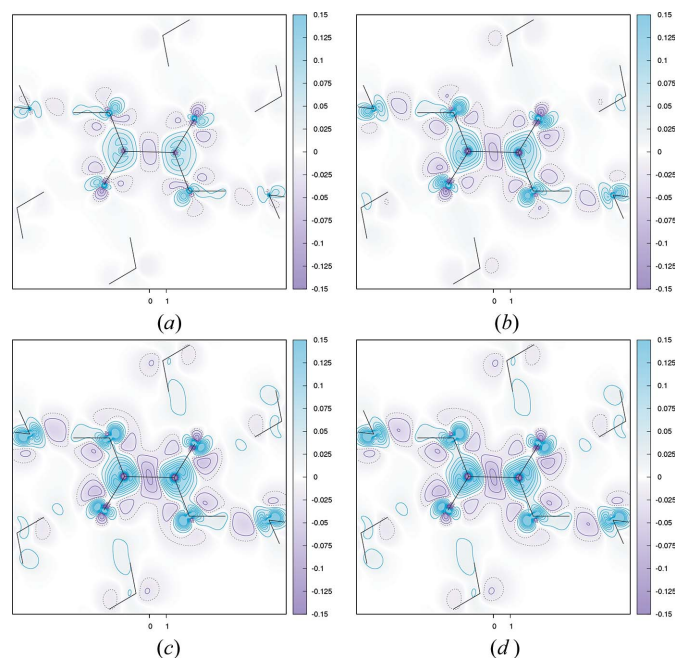


Figure 4
Comparison of changes in electron density due to the XCW procedure for the Zobel (1996) data set. The plots presented are at (a) $\lambda_{\text{opt}} = \lambda_{\text{TIH3}}$ [$\lambda_{\text{TIH3}} = 0.03$], (b) $\lambda_{\text{opt}} = \lambda_{\text{TIH6}}$ [$\lambda_{\text{TIH6}} = 0.08$], (d) $\lambda = \lambda_{\text{max}}$ [$\lambda_{\text{max}} = 0.19$], and (c) one step before at $\lambda = \lambda_{\text{pen}}$ [$\lambda_{\text{pen}} = 0.18$]. Units are in e \AA^{-3} , each contour indicates a change of 0.025 e \AA^{-3} .

Table 4

λ_{opt} values (λ_{pen} , λ_{TIH3} and λ_{TIH6}) used to halt the HA-XCW procedure, and the corresponding GoF^2 values.

'All' represents values for the simultaneous refinement of all Kaminski *et al.* data sets except oxa2. 'All_{min}' is the same but using data fitted up to $\lambda = \min_i[\lambda_{\text{pen}}(\text{oxa}i)] = 0.28$.

Data set	λ_{pen}	$\text{GoF}^2(\lambda_{\text{pen}})$	λ_{TIH3}	$\text{GoF}^2(\lambda_{\text{TIH3}})$	λ_{TIH6}	$\text{GoF}^2(\lambda_{\text{TIH6}})$
oxa1	0.74	1.263	0.20	1.411	0.21	1.401
oxa2	2.00	0.835	0.39	0.890	0.36	0.894
oxa3	1.18	1.060	0.41	1.146	0.32	1.171
oxa4	1.20	0.921	0.32	0.999	0.30	1.002
oxa5	1.18	1.415	0.36	1.549	0.30	1.572
oxa6	0.78	1.059	0.24	1.189	0.24	1.189
oxa7	0.74	1.075	0.22	1.193	0.20	1.204
oxa8	1.20	0.978	0.29	1.052	0.28	1.056
oxa9	2.00	1.153	0.40	1.240	0.38	1.243
oxa10	1.16	0.959	0.26	1.060	0.30	1.051
oxa11	0.28	1.390	0.10	1.582	0.09	1.582
oxa12	1.20	1.265	0.36	1.395	0.33	1.402
oxa13	1.40	0.981	0.39	1.074	0.36	1.082
All	1.16	1.395	0.25	1.516	0.23	1.523
All _{min}	0.28	1.509	—	—	—	—

the conservative value of λ leaves the refinements less likely to be affected by overfitting.

More interesting is the comparison with the original plot in GJ2001, reproduced in Fig. 1. Importantly, the strong increase in electron density around the hydrogen atom (by several contours of 0.1 e \AA^{-3}) has almost completely disappeared in all of the HA-XCW plots; there is now only a very small and delocalized increase of 0.025 e \AA^{-3} .

As stated in the introduction, the strong buildup of charge on the hydrogen atom might be due to a wrong O—H bond length, since this was obtained from an IAM refinement. The O—H bond length obtained from the CIF of Zobel *et al.* (1992) was 0.94 \AA . However, the bond length and ADPs used by GJ2001 were not reported; as these were from an IAM refinement, it is almost certain that the bond length was shorter than the HAR value used here, which was 1.12 \AA , *i.e.* significantly longer. The relatively large HAR value makes sense, since this hydrogen atom is part of a strong hydrogen bond with $d(\text{O} \cdots \text{O}) = 2.484 \text{ \AA}$. Therefore, the strong stretched feature near the oxalic acid hydrogen atom in Fig. 1 may be explained as a simple correction to a 'wrong' geometry, not any new many-body physics.

4.2. HA-XCW for the Kamiński *et al.* (2014) data

Tables S5–S17 in §S.12 of the supporting information present the collected GoF^2 versus λ data points for each of the 13 data sets by Kamiński *et al.* (2014). Figs. S1 and S2 present the corresponding power function fits, for the GoF^2 and $\text{GoF}_{\text{corr}}^2$ versus λ plots. Table S1 presents the A , B , D , E and F parameters for the power function and asymptotic fitting methods to decide the halting point. Some more discussion about the obtained magnitudes of these parameters is given in the supporting information in §S.1 and §S.2.

4.2.1. Comparison of the halting procedures. The halting values of λ at the penultimate value before convergence

ceases λ_{pen} and values associated with the TIH extrapolations are given in Table 4. Values of λ_{TIH3} and λ_{TIH6} and the GoF^2 corresponding to these values are very close to each other indicating that the asymptotic expansions are consistent with each other, and it does not much matter which method (3- or 6-point) is used. We observe that the GoF^2 values are very similar, differing by at most 2.2% in data set oxa3. The TIH6 method might be more preferable because it has associated error estimates that could be used to indicate a failure of reliability (if the errors are large). Values of λ_{TIH3} and λ_{TIH6} could always be reached in all of the data sets.

It is interesting to note that there is a strong correlation between the A parameter in the power function fitting and the asymptotic extrapolation D parameter (see Fig. 5 for a plot between the A parameter and the 6-point D parameter) the former of which was argued to represent the error in the σ s, while the latter was argued to represent the basis-set *and* experimental error. The Pearson correlation coefficients between A and the 3- and 6-point D values are 0.97 in both cases.

The value of A is not well correlated with λ_{TIH3} or λ_{TIH6} ; the correlation coefficients are, respectively, -0.10 and -0.19 . However, A is remarkably well correlated with the corresponding $\text{GoF}^2(\lambda_{\text{TIH3}})$ and $\text{GoF}^2(\lambda_{\text{TIH6}})$ values; the corresponding correlation coefficients are greater than 0.92 in both cases. If A is a measure of the errors in the σ s and hence a measure of the errors in the experiment, this lends support to the fact that the GoF^2 values at the TIH halting values also incorporate this same information concerning the errors in the experiment, which is very encouraging.

Note that beyond small values, λ_{opt} is rather variable compared to the corresponding $\text{GoF}^2(\lambda_{\text{opt}})$ values, due to the power functions form of the relationship between them. For example, the slope of the power function $A\lambda^B$ at $\lambda = 1$ is easily evaluated to be AB , which has a small value of about -0.1 in most cases, and it is even smaller for larger values of λ .

4.2.2. Effect of different halting procedures on fitted electron density. Fig. 6 shows the effect of the XCW fitting process on the electron density for $\lambda = \lambda_{\text{TIH6}}$ for all data sets except oxa2 (*i.e.* it shows the electron density fitted at $\lambda = \lambda_{\text{TIH6}}$ minus the electron density at $\lambda = 0$). Data set oxa2 is presented separately in Fig. S13, because it has distinctly different characteristics. Similar plots at other λ_{opt} halting values are available in §S.15 of the supporting information, Figs. S14–S18. There, the plots from the different λ_{opt} values are shown side-by-side to facilitate verification of the discussion in the next paragraphs.

We observe from these plots that the effect of XCW fitting on the electron density depends on both the data set and on the halting procedure used, however, to different extents as will be discussed now. The 3- or 6-point asymptotic methods are essentially the same, as is expected since the corresponding GoF^2 values for halting these procedures is the same. The features from the power function halting procedure maps are qualitatively similar to those obtained from the asymptotic method for each data set and the changes are bigger than between λ_{TIH3} and λ_{TIH6} since the power function halting

point at $\lambda = 1$ involves a significantly larger penalty. The method of fitting to λ_{pen} yields halting values roughly around $\lambda = 1$ (although oxa9 is exceptional) and so the electron density difference maps from this λ_{pen} method exhibit features qualitatively more similar to those from the power function method than from the asymptotic extrapolation method.

In summary, whereas the effect based on the data set shows qualitatively different features (discussed in the next subsection), the effect based on the halting procedure shows always the same qualitative features, but quantitatively more or less pronounced. The features are less pronounced with smaller λ values, and more exaggerated with higher λ values, as expected. Therefore, if only qualitative features are of interest, this means that for any regular example we can still fit to λ_{max} as has always been done.

The differences in the features of the electron density in the valence regions amount to only a few contours – essentially a few multiples of $1/40\text{th e \AA}^{-3}$. This is a substantially better agreement than was observed in GJ2001 and papers from that era, which is an important objective measure of the improvement in both the theoretical models and experimental data since that time. These improvements in the contour maps are of course consistent with the improvements in the HAR agreement statistics in Table 1.

4.2.3. Qualitative analysis of features due to HA-XCW fitting. Since the features due to the XCW fitting halting process are qualitatively identical, only differing in the magnitude of the contour levels, further comparisons

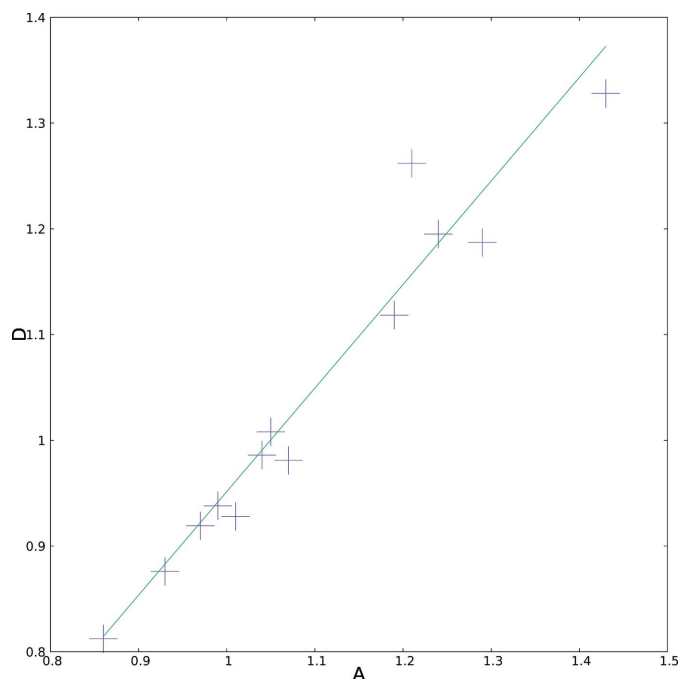


Figure 5
Plot between the A parameters from the power function fitting procedure and the corresponding D parameter from the 6-point asymptotic extrapolation procedure. These parameters have a Pearson correlation coefficient of 0.97. The line of best fit is $D = 0.98(8)A - 0.03(8)$.

are based only on an analysis of Fig. 6 from the more conservative, and theoretically more sound, asymptotic halting procedure.

In an attempt to objectively find qualitative similarities between any features, we divided the plot space into bond regions between the atoms, core regions near the nuclei, inner valence regions, and outer valence regions away from bond regions. Within these regions we tried to identify features (judged by the number, or shape, of the contours) to classify the qualitative changes. These features are described below,

and cross-referenced to ideal or model visual examples given in Fig. 7. The features are:

(i) A decrease in the ‘lone pair’ feature in the outer valence region of the oxygen atom in the hydroxyl unit, denoted by ‘-lp’ [Fig. 7(a)];

(ii) An ‘elephant ears’ – depletion of the lone pairs – feature in the outer valence on the carbonyl oxygen atom, denoted by ‘e’ [Fig. 7(b)];

(iii) A polarization in the electron density in the core *away* from the bonded atom B; this could also be called a decrease

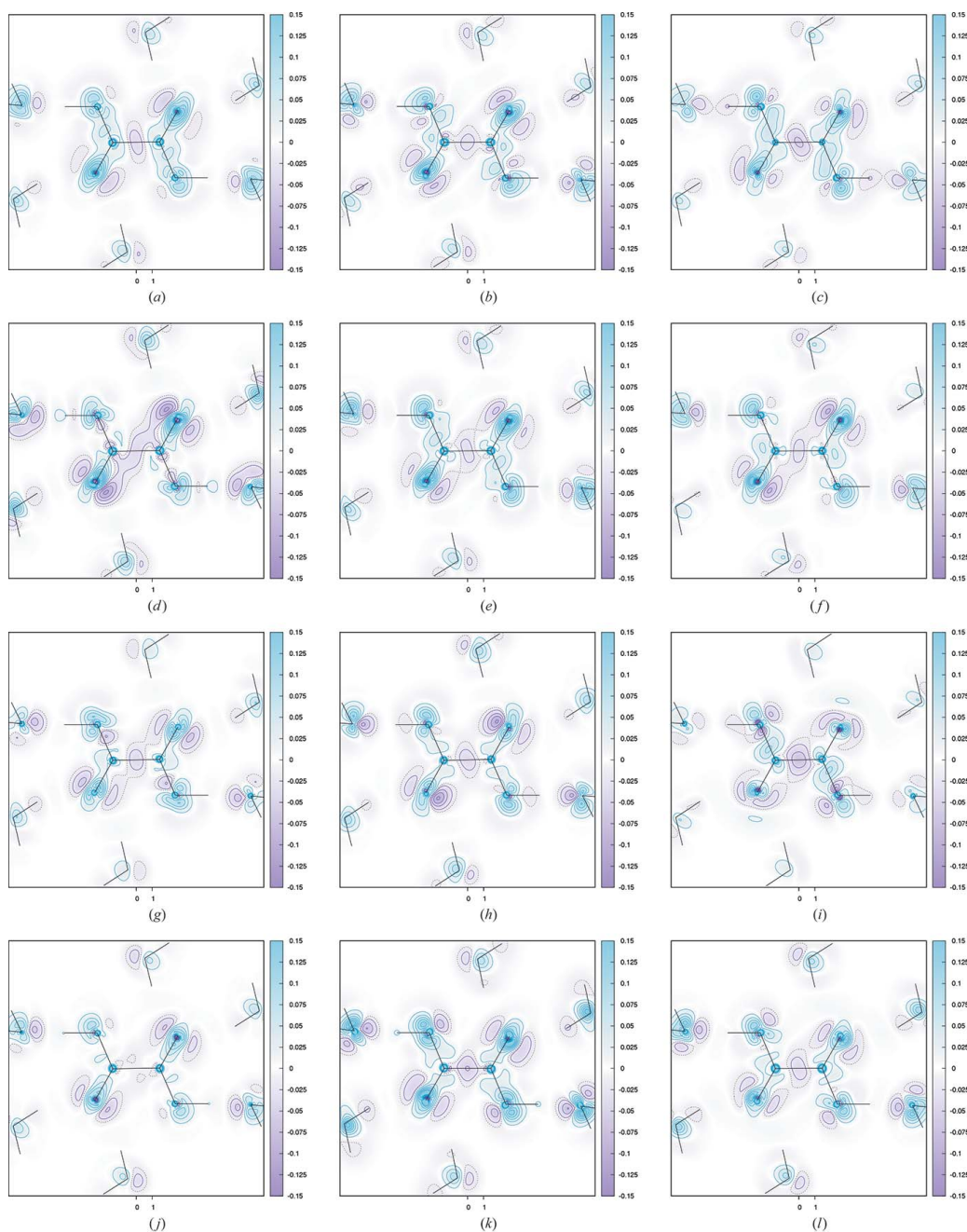


Figure 6

Difference in electron density due to the XCW fitting procedure on the Kamiński *et al.* (2014) data sets at $\lambda = \lambda_{\text{TIH6}}$ for all data sets except oxa2. Units are in $\text{e} \text{ \AA}^{-3}$, difference between contour levels is $0.025 \text{ e} \text{ \AA}^{-3}$. Lilac is a reduction of electron density, blue is an increase upon fitting.

Table 5

Effect on the electron density in the midpoint of the bond, the core, the inner-valence and outer-valence regions of the specified atoms due to the HA-XCW fitting process at $\lambda = \lambda_{\text{TIH6}}$ for each of the Kamiński *et al.* (2014) data sets.

Symbol + (–) indicates an increase (decrease) in electron density relative to the unfitted wavefunction; \rightarrow A indicates a decrease in electron density towards atom A in the core region; 0 indicates the change was less than $0.025 \text{ e } \text{\AA}^{-3}$; \perp AB indicates a crescent-shaped decrease in electron density centred on the atom, towards atoms A and B; and ee indicates elephant ears a decrease parallel to the oxygen atom; and -lp indicates a decrease in the lone pair of an atom. See Fig. 7 for model examples of these features. ‘All’ refers to the HA-XCW simultaneously fitting all data sets; and B3LYP, PBE0 and CCSD indicate results for the corresponding theoretical wavefunctions.

Data	Bond region				Core region			Inner valence region				Outer valence region		
	C–C	C–O	C=O	O–H	C	O in OH	O in C=O	H	C	O in OH	O in C=O	C	O in OH	O in C=O
oxa1	–	+	+	+	+	+	\rightarrow C	0	+	\perp CH	+	+	-lp	ee
oxa3	–	+	+	+	+	+	\rightarrow C	0	–	\perp CH	+	+	-lp	ee
oxa4	–	+	+	+	+	+	\rightarrow C	–	+	\perp CH	+	+	-lp	ee
oxa5	–	–	+	–	+	+	\rightarrow C	+	–	\perp CH	+	0	0	ee
oxa6	–	+	+	+	+	+	\rightarrow C	0	+	\perp CH	+	+	-lp	ee
oxa7	–	+	+	+	+	+	\rightarrow C	0	0	\perp CH	+	+	-lp	ee
oxa8	–	–	+	0	+	+	+	0	0	\perp CH	+	+	0	ee
oxa9	–	+	+	–	+	+	\rightarrow C	0	–	\perp CH	+	0	+	ee
oxa10	–	+	0	–	+	+	\rightarrow C	–	+	\perp CH	+	+	-lp	ee
oxa11	0	0	+	+	+	+	\rightarrow C	0	0	\perp CH	+	0	-lp	ee
oxa12	–	+	+	+	+	+	\rightarrow C	+	–	\perp CH	+	+	-lp	ee
oxa13	–	+	+	+	+	+	\rightarrow C	0	0	\perp CH	+	+	-lp	ee
All	–	+	+	+	+	+	\rightarrow C	0	–	\perp CH	+	+	-lp	ee
B3LYP	–	–	–	–	–	–	–	+	+	+	+	+	-lp	ee
PBE0	–	–	–	–	–	–	–	+	+	+	+	+	-lp	ee
CCSD	–	–	–	–	–	+	+	+	+	+	+	+	-lp	ee

or pink colour *towards* the bonded atom, denoted by \rightarrow B [Fig. 7(c)];

(iv) A crescent-shaped decrease in electron density in the inner valence region, centred on the bisector of the smaller (obtuse) C–O–H bond angle, and corresponding to a kind of partial shell structure, denoted by \perp CH [Fig. 7(d)]; and finally

(v) An increase in electron density in the outer-valence region of a given atom, away from any bond [Fig. 7(e)], or inside the bond [Fig. 7(f)], denoted by +. Likewise, a decrease is denoted by –.

Table 5 quantifies the appearance of these features for each data set, making use of the shorthand symbols associated with these features, listed above.

Bond regions. In all data sets, there is a decrease in the C–C bond density (except in oxa11), mostly an increase in the C–O bond density (except in oxa5, oxa8 and oxa11), and an increase in the C=O bond density (except in oxa10). The behaviour in the O–H bond is less clear, with a clear increase in the bond density in 8 of the 13 cases.

Core regions. Across all data sets, there is a consistent increase in the electron density very close to the centre of the C atom, and the same is true for the O atom in the O–H bond. For the O atom in the C=O bond, in all cases there is a decrease in the core region polarized towards the C atom (except oxa8). For the H atoms, the behaviour is less clear: there is generally no change (eight cases), with an increase in two cases, and a decrease in three cases.

Inner valence regions. For the C atom, there is no clear consistent behaviour: there is a weak decrease in four cases, a weak increase in five cases, and no change in four cases. The behaviour is difficult to quantify, since there are varying directional polarizations associated with this effect, which is reminiscent of a shell or crescent-like structure. For the O atom in the O–H moiety, there is a generally crescent-shaped decrease in electron density in all cases, but again, in some cases the feature is nearly absent or highly degenerated from the ideal form used for classification. For the O atom in C=O, the situation is much more clear: there is always an increase in the inner valence region in all cases, but the associated shape is somewhat variable and lopsided.

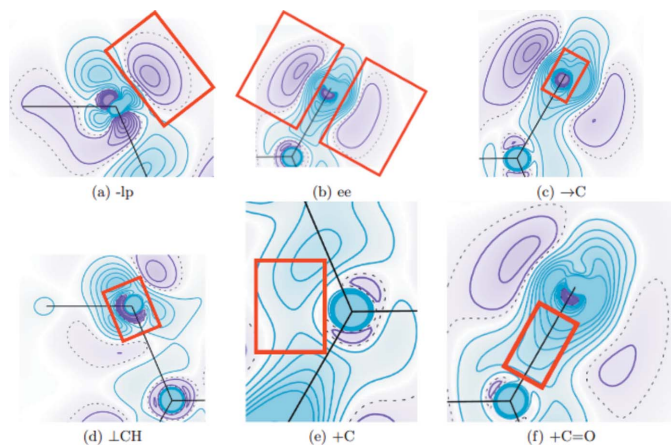


Figure 7

Model examples for key features which appear exclusively in regions of the HA-XCW electron density due to the fitting process (red boxes), and which were used to identify similarities and differences in the electron density difference maps. (a) The reduction in lone pair electrons on the hydroxyl oxygen atom (-lp), (b) the elephant ears feature (ee) on the carbonyl oxygen atom, associated with the two lone pairs, (c) the strong decrease in electron density towards the bonded atom C (\rightarrow C) in the core region, (d) the crescent-shaped decrease in the inner valence electron density in the bisector of the obtuse part of the \angle COH, perpendicular to the line joining the C and H atoms (\perp CH), (e) a build up in electron density in the outer-valence region of the atom C (+C) away from the bonds, and (f) a build up in electron density in the bonding region between the C atom and the carbonyl O atom (+C=O).

Outer valence regions. For the C atom, there is an increase in the valence electron density upon fitting on the non-bonded side in ten cases, the remaining three cases showing no change. For the O atom in OH, there is a depression in the lone pair in the outer region in nine cases (the exceptions are for oxa5, oxa8 and oxa9, though in the last cases there is a very weak depression). For the O atom in the C=O group, there is a very clear pair of elephant-ears-shaped depletions for all the cases, but the shape is neither symmetric or uniformly asymmetric.

In summary, the only clearly consistent features observed in the electron density over all data sets after fitting are:

(i) A decrease in the electron density of the C–C bond (except for oxa11, but it appears in the $\lambda = \lambda_{\text{pen}}$ plot, *i.e.* at a higher λ value);

(ii) In the core region, an increase in the electron density on the carbon atoms;

(iii) In the inner-valence region, a crescent-shaped decrease in the electron density in the hydroxyl oxygen atom; and

(iv) In the outer valence region, a decrease in the electron density in the carbonyl oxygen atom lone pairs, the so-called elephant ears structure, and an increase in electron density around the carbon atoms;

The water molecule is not discussed in the tables. For the water molecule in the hydrogen bond (the one whose electron density is constrained, see §3.8), there is a consistent increase in electron density towards the core of the oxygen atom, and a decrease in the lone pair on the oxygen atom in all data sets.

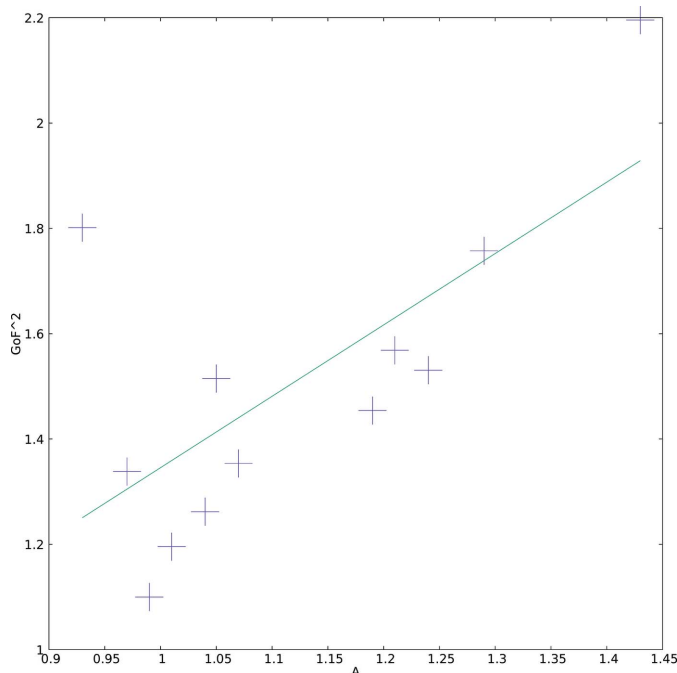


Figure 8

Plot between the individual A parameters from the power function fitting procedure (see Table S1) and the GoF^2 values for these individual data sets when they are used in the joint refinement. These parameters have a Pearson correlation coefficient of 0.68, which increased to 0.91 with the removal of the outlier at (0.93, 1.80). Without the outlier the line of best fit, shown in green, is $\text{GoF}^2 = 1.9(3)A - 0.7(3)$.

4.2.4. Simultaneous HA-XCW fitting of all Kamiński *et al.* (2014) data sets. To gather a consensus between all the data sets, we performed a HAR and then a HA-XCW fitting, *i.e.* a full X-ray wavefunction refinement procedure based exclusively on Hirshfeld atoms [called XWR(HA), see part I (Davidson *et al.*, 2022)], on all data sets but oxa2 simultaneously. Each data set was given its own separately least-squares optimized scale factor. The GoF^2 versus λ plot behaves as expected and is included in Fig. S3. A power function fit yielded fitted parameters $A = 1.4053(6)$ and $B = -0.0554(3)$, while asymptotic model fits produced 3-point parameters of $D = 1.346$, $E = 0.084$ and $F = -0.026$ and 6-point parameters of $D = 1.350(4)$, $E = 0.073(9)$ and $F = -0.018(6)$. These values are reported with those from the individual data sets in Tables 4 and S1 in the ‘All’ row. There is little of note about a comparison of all fitting parameters, other than that they are distinctly *not* the average of the parameters of the individual refinements.

If A represents a scaling associated with the σ s, as argued in §2, then one might expect that this value is dominated by the largest individual value of A in the combined data set; and indeed, the value for A in oxa5 is 1.43, compared to 1.41 when refining all data sets simultaneously. A somewhat similar alternative explanation might be that the σ s from the individual data sets are underestimated, and when combined together the A value for the combined data set is larger. The latter is a more conservative hypothesis than the former because it does *not* posit that the A value from the combined refinement is due to the largest of the individually refined A values.

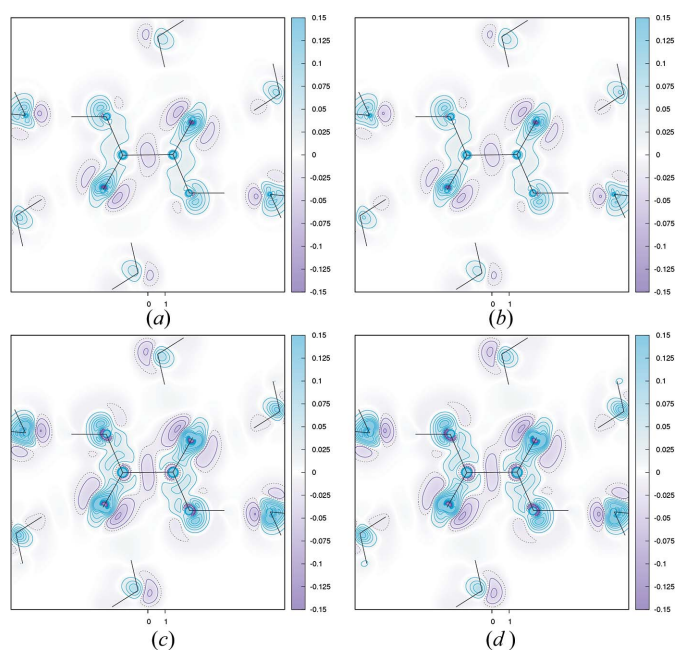


Figure 9

Difference in electron density due to the XCW fitting procedure simultaneously on all the Kamiński *et al.* (2014) data sets (except oxa2). Units are in $\text{e} \text{ \AA}^{-3}$. (a) At $\lambda = \lambda_{\text{min}} = 0.28$, the smallest value of λ_{pen} across all refinements except oxa2. (b) At $\lambda = \lambda_{\text{TiH6}} = 0.24$, (c) at $\lambda = \lambda_{\text{pow}} = 1$, and (d) at $\lambda = \lambda_{\text{pen}} = 1.16$.

Further evidence that the A parameters from the fittings against the individual data sets may represent an overall scaling factor related to the errors in the standard uncertainties (σ) can be tested by plotting those individual A values against the GoF^2 values for these individual data sets when they are used in the joint refinement. In Fig. 8, we show the result at $\lambda = \lambda_{\text{TIH6}}$. The graph is linear with one strong outlier at (0.93, 1.80) from the oxa4 data set. Removing this improves the Pearson correlation coefficient from 0.68 to 0.91. From this we conclude that in 11 out of 12 cases, the A parameter from the power function extrapolation method appears to provide a reasonable estimate of the scaling factor in the σ s. We could not find any obvious reason why oxa4 was an outlier.

In any case, if the A value is associated with a uniform scaling of the σ s, then a larger value implies that the σ s were underestimated, which in turn implies a larger value of GoF^2 than necessary (for a given λ) if the (presumed) correct σ s had been used. With this in mind, we observe that in the simultaneous refinement it was possible to fit values up to $\lambda_{\text{pen}} = 1.16$, which is in excess of what was possible for the individual refinements of data sets oxa1, oxa6, oxa7, and oxa11, which had λ_{max} values of 0.74, 0.78, 0.74 and 0.28, respectively. Therefore, effectively, more values from these data sets are ‘included’ in the simultaneous XCW procedure at $\lambda_{\text{pen}} = 1.16$ than was possible for the corresponding XCW procedure on these individual data sets, so that one might expect overfitting to these particular data sets in the simultaneous fit. For this reason, we suggest a conservative solution to the halting problem as introduced in §2.3.

We find that $\lambda_{\text{pen}}^{\text{min}} = \lambda_{\text{pen}}(\text{oxa11}) = 0.28$, and that $\lambda_{\text{TIH6}}^{\text{min}} = \lambda_{\text{TIH6}}(\text{oxa11}) = 0.09$. The former value is similar to the λ_{TIH6} value from the simultaneous refinement, which was 0.23, and of course, the features exhibited in the plots at λ_{min} [see Fig. 9(a)] and λ_{TIH6} [see Fig. 9(b)] are almost identical. Therefore, the λ_{TIH6} halting point is indeed also a conservative one, compared to $\lambda_{\text{pow}} = 1$. (Note that the power function fit using points only up to $\lambda = \lambda_{\text{min}} = 0.28$ yields parameters $A = 1.417$ (4) and $B = -0.052$ (1) which are very similar to those using all data up to λ_{pen} , which might have been guessed since the power function furnishes such a good fit; so the power-function halting value is very robustly estimated). On the other hand, the $\lambda_{\text{TIH6}}(\text{oxa11}) = 0.09$ is quite different to the conservative estimate $\lambda_{\text{pen}}(\text{oxa11}) = 0.28$, nearly a third of it. Nevertheless, if we look at the corresponding $\text{GoF}^2(0.28)$ and

$\text{GoF}^2(0.10)$ values from the simultaneous refinement, they are, respectively 1.509 and 1.60 ± 0.02 , where an estimate of the error in the second value is obtained using the difference of this $\text{GoF}^2(\text{TIH6})$ value with the corresponding $\text{GoF}^2(\text{TIH3})$ value. These values are relatively close, and both are close to the $\text{GoF}^2(\lambda_{\text{TIH6}}) = \text{GoF}^2(0.23) = 1.523$.

The effect on the electron density due to this XCW fitting procedure at different selected values of λ is shown in Fig. S12. Clearly, the largest changes in the electron density occur in the first four figures, as might be expected as the power function decay of the GoF^2 statistic (which measures the agreement of the calculated and observed structure factor magnitudes which are electron densities in reciprocal space) also displays the most rapid change when λ is small. This also supports the earlier decision to consider the more conservative $\lambda = \lambda_{\text{TIH6}}$ halting point when discussing qualitative features in §4.2.3.

The effect of the XCW procedure on the electron density is shown for $\lambda = \lambda_{\text{TIH6}}$ [Fig. 9(b)], $\lambda = \lambda_{\text{pow}}$ [Fig. 9(c)] and λ_{pen} [Fig. 9(d)]. The plots produced at the two higher values (λ_{pow} and λ_{pen}) are almost identical. As expected, the effect of the XCW procedure is smaller (*i.e.* fewer features, features smaller in magnitude) for λ_{TIH6} compared to the plots at the higher value of λ . Specific examples include a reduction in the asymmetry in the elephant ears (one contour of difference versus two contours of difference), and the delocalization of the loss of electron density in the C–C bond region (localized versus delocalized).

Overall, features become more prominent when λ increases, as expected, and we do not have a statistical criterion to judge which of the λ values gives rise to the physically most meaningful electron density. Importantly, it is worth noting that the features from the simultaneous refinement are the average of those found from the individual refinements, as can be verified from the qualitative analysis (Table 5) in the row marked ‘All’. This means that the features in Fig. 9 can be considered to be the consistent features that are due to systematic effects (systematic physical effects or systematic error) captured by the XCW fitting procedure beyond noise.

4.2.5. Comparison of HA-XCW results with high-level theoretical calculations. The effect of electron correlation on the electron density of oxalic acid surrounded by six water molecules is seen in Fig. 10, where the electron density from a Hartree–Fock calculation is subtracted from the ones calculated by the B3LYP, and PBE0 density functional theory calculations, and from a CCSD calculation. There is very little difference between the different theoretical models: the CCSD calculation has one fewer contour level in the C–C bond region, two fewer contours around the C atoms in the outer valence region, and one more contour in the outer valence of the O atom of the hydroxyl group.

Boyd & Wang (1989) have seen similar features on the carbonyl group in formaldehyde, compared to the carbonyl groups shown in Fig. 10. For example, the strong negative regions in the inner core of the C and O atoms are the same, as is a kind of trefoil structure in the inner valence region of the C atom. Boyd & Wang (1989) explained the blue positive

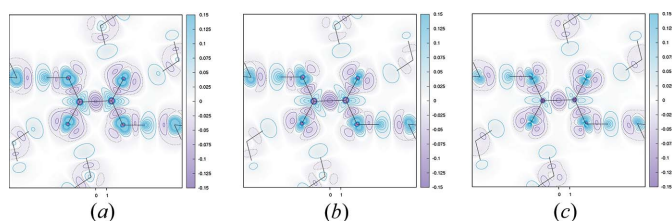


Figure 10

The change in the electron density due to electron correlation; modelled as the difference between a theoretical model wavefunction and HF/def2-SVP electron density at the HAR geometry from oxa11 data set. (a) is B3LYP, (b) is PBE0 and (c) is CCSD.

electron density features in the inner valence region on the O atoms as being due to a transfer of electron density to the electronegative atom, correcting a well known deficiency of the Hartree–Fock model, which tends to overestimate the ionic contribution in polar bonds (because it gives equal weight to covalent and ionic valence bond structures).

Assuming that charge-transfer effects, and polarization effects of the crystal field, are adequately modelled by the six-water cluster around the oxalic acid unit, it is expected that the features in Fig. 10 should be comparable to those from the consensus XCW procedure seen in Fig. 9. The features in the theoretical densities are compared with those in the experimental ones in Table 5 in the rows with headings B3LYP, PBE0 and CCSD. Since the B3LYP, PBE0 and CCSD plots are so similar, we use the CCSD plot as a reference map to compare to the consensus plots. Also, since the λ_{\min} and λ_{TlH6} plots are so similar, and likewise for the λ_{pow} and λ_{pen} plots, *i.e.* they are pairwise the same, in the following we compare only the λ_{\min} XCW plot, Fig. 9(a), with the CCSD map.

We observe several consistencies in the features; the following features are qualitatively the same but differ in the number of contour lines: there are two more contours in the C–C bond region, two more contours in the outer valence region] of the O atom on the hydroxyl group, one less contour in the more depleted of the lone pairs (or elephant ears) of the outer valence region of the O atom on the carbonyl group, two more contours in the outer valence region of the C atom [as in

Fig. 7(e)] (and, as mentioned above, even more in the DFT calculations).

The buildup of charge in the inner valence region of the O atoms is interesting in line of the work of Boyd & Wang (1989): the XCW map is clearly producing the same feature, *i.e.* the experimental data appears to be correcting the overestimation of the ionicity in the Hartree–Fock model.

We also observe some inconsistencies: the trefoil structure on the C atoms which was clear in the theoretical maps is much less well defined, if not absent, in the XCW plots. Likewise, the depletion of charge in the middle of the carbonyl bond, and the O–H bond seen in the theoretical maps is completely absent in the XCW plots.

Also interesting for this work is the significant build-up of electrons on the hydrogen-bonded hydrogen atom, which was completely absent in the consensus plot. Although this feature was also seen in the original GJ2001 work (Fig. 1), and was one of the motivations for this work, we established in §4.1.4 that it resulted from a poor choice of the hydrogen atom coordinates. This build-up at the H atom is consistently present in the theoretical electron densities but missing in the individual and consensus XCW plots (see Table 5). However, recall that the scale is four times finer than in the GJ2001 paper.

5. Discussion: possible reasons for discrepancies between experimental and theoretical densities

The differences in the qualitative features between the consensus XCW map and the theoretical maps in the previous section might be due to several reasons, but can be classified as either errors in the theoretical model used to perform XCW fitting or the experimental data. Some of the reasons were already discussed at the point they were relevant, but for completeness we mention them again here.

On the theoretical side, the following factors are of concern.

(i) *Use of the Hartree–Fock wavefunction ansatz.* The Hartree–Fock model used in this study is known to have defects, *e.g.* the overestimation of ionic bonding effects in polar bonds, as discussed in §4.2.5. Therefore, it might not be the best starting point for an XCW fitting.

(ii) *Use of a one-centre probability density function (OCPDF) model.* The work of Bučinský *et al.* (2019) indicates that the use of a OCPDF model has a large effect, albeit for systems involving heavy atoms where such effects are exacerbated. It is therefore important to establish whether such effects are also important for light-atom systems, such as α -oxalic acid dihydrate considered here.

(iii) *Insufficient treatment of crystal effect.* The model we have used is fragment-based and non-periodic, so the effects of a (nearly) infinite crystal field are not correctly modelled. We tested the effect of not modelling the crystal field properly by introducing point charges representing atomic charges and atomic dipoles on all atoms of all molecules with any one atom within 8 Å of our six-water cluster. Atomic charges and dipoles were calculated self-consistently from the symmetry-unique asymmetric unit atoms in the central moiety using Hirshfeld's method, and the charges were updated

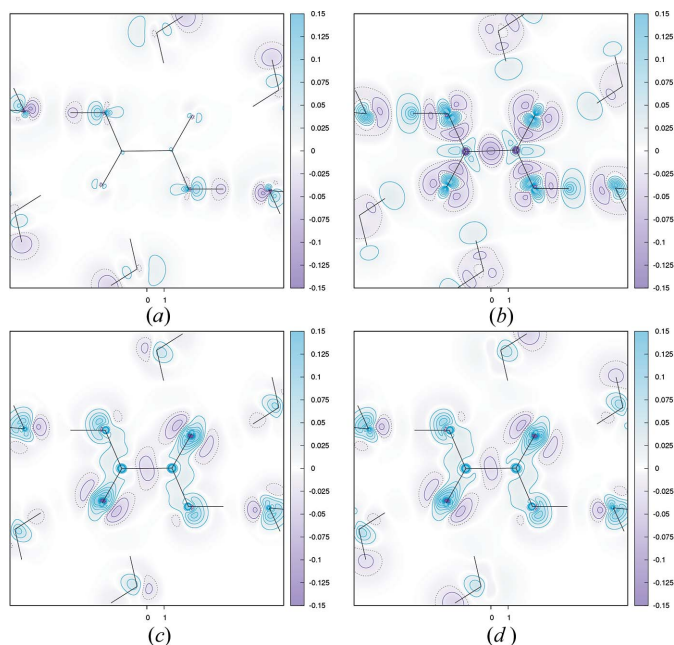


Figure 11

Change in the electron density from the usual RHF ansatz for various procedures due to: (a) the inclusion of point charges reproducing the Hirshfeld atom charges and dipoles, placed on all atoms on all molecules with one atom within a radius of 8 Å of the six-water cluster moiety [Point charges were calculated and updated self-consistently from the atoms in the central moiety during the SCF cycles.]; (b) the inclusion of correlation effects via the CCSD calculation [reproduced Fig. 10(c)]; (c) the XCW fitting procedure on all data sets at λ_{\min} [reproduced Fig. 9(d)]; (d) the XCW fitting procedure on all data sets at λ_{\min} , with point charges included [as in (a)].

self-consistently during the self-consistent field procedure used to solve the constrained restricted Hartree–Fock equations (Jayatilaka & Dittrich, 2008; Wieduwilt *et al.*, 2021).

The change in the electron density due to the addition of these point charges (at the consensus HAR geometry) is shown in Fig. 11(a). A similar figure using charges which were not self-consistently updated was indistinguishable from this figure, and is not presented. We observe that the change in the electron RHF wavefunction is non-negligible: there are three contours-worth of a dipolar-like depletion in electron density on the O atom in the hydrogen bonded water molecule, a feature which seems to be more perpendicular to the plane of the water molecule since it is not seen in other water molecules which are more in the plot plane. There are a further three contours-worth of an increase in electron density on the hydroxyl O atom toward the hydrogen-bonded H atom, and a similar sharper feature behind it. These features are located in the core or inner valence region, so they would require relatively higher resolution data to discern. Nevertheless, this indicates that there are indeed effects not properly modelled, without cluster charges, even with six water molecules surrounding the oxalic acid moiety. Features on the H atoms on the outer water molecules are understandable since they are close to the point charges representing the missing closest molecules, but the electron density on these molecules is not used to calculate the structure factors. Nevertheless, it seems that the point charges around these water molecules (representing oxalic acid molecules) are sufficient to affect the closest hydrogen bonded water molecule, and also the O atom of the hydroxyl group in the central oxalic acid moiety via a through-bond effect.

If we assume the crystal field effects in Fig. 11(a) are additive onto those representing the effects due to electron correlation, say as represented by the CCSD wavefunction, shown again in Fig. 11(b) for ease of comparison, then the net effect would be to reduce the number of contours on the hydrogen-bonded H atom and to reduce the depletion in the middle of the hydroxyl bond, bringing the CCSD plot into better agreement with the XCW model at $\lambda_{\text{pen}}^{\text{min}} = 0.28$, shown again in Fig. 11(c). The very small asymmetric crystal-field effect on the carbonyl oxygen atom due to the addition of the point charges seems too small to explain the asymmetry in the elephant ears of the XCW plots.

So far, we have only considered a Hartree–Fock wavefunction without surrounding point charges mainly to establish if the crystal field effects and electron correlation effects could be incorporated via the experimental structure factors. We were curious to test what effect using these point charges in the XCW model would have. The results at $\lambda_{\text{min}} = 0.28$ are presented in Fig. 11(d). (We did not redo the individual or consensus XCWs again to get λ_{TIH3} and λ_{TIH6} values). On the central oxalic acid moiety, we see hardly any discernible changes in the same regions as Fig. 11(a). The greatest difference between the XCWs with and without charges is exhibited in the most distant water molecules [top and bottom water molecules in Figs. 11(c) and 11(d)], and they do seem to

be roughly the addition of features in Figs. 11(a) and 11(b). In any case, these molecules are not used when calculating structure factors. Thus, based on these plots, the crystal field effect seems to be overwhelmed by the use of the experimental data, regardless of the inclusion of point charges in the wavefunction model.

(iv) *Failure of the XCW model to deal with high resolution data.* It is known that, when using theoretical structure factor data, the XCW procedure completely fails to reproduce features in the core region even with data resolved to 2.5 \AA^{-1} , much higher than in this experiment (Podhorský *et al.*, 2021). Indeed, the features in the core and inner valence regions of the atoms between the XCW and theoretical maps were mostly completely different, indicating that the resolution problem plays a decisive role for these regions in this study, too.

On the experimental side, the following concerns are raised.

(i) *Data resolution and accuracy.* This issue concerns the fact that, on the one hand, any experiment has finite resolution (see also previous point), and, on the other hand, that the high angle reflections are much weaker in intensity and therefore harder to observe, *i.e.* are associated with larger relative error. Any errors in these high angle reflections will manifest in a lack of resolution of sharp features in the map. Indeed some sharp features are systematically entirely missing compared to the theoretical maps (see core region entries in Table 5).

(ii) *Systematic experimental errors.* There may also be a range of other experimental effects which we only briefly mention, including but not limited to: systematic errors in the experimental data coming from absorption corrections; inappropriate merging of experimental data; artificial strengthening of weak reflections and weakening of strong reflections due to dynamical effects, *i.e.* rescattering of X-rays; and thermal diffuse scattering, which affects the background assumed for the X-ray reflections. Normally, these effects are not treated in a very exhaustive way, and it is clear that, given the ‘fickle’ visual agreement between the XCW plots, and the level of effort expended on the theoretical analyses, it becomes urgent to reinvestigate the treatment of the experimental data, especially in view of the improvements in detector technology.

6. Summary and conclusions

In order to assess the reproducibility of different features in the electron density reconstructed from experimental X-ray diffraction structure factor magnitudes, in this paper, we have applied the Hirshfeld-atom X-ray constrained wavefunction (HA-XCW) procedure to the historically important α -oxalic acid dihydrate system, using the same data as was used in Grimwood & Jayatilaka (2001), and also using thirteen data sets from Kamiński *et al.* (2014). In the case of the latter, more modern data, we have reported the first constrained wavefunction simultaneously fitted to multiple data sets.

Since the XCW procedure is a not a least-squares procedure but a reconstruction or regularization method, in order to

assess the reliability of our results, we have also developed three new protocols to address the so-called halting problem — namely, the problem of deciding the level of agreement between the model-calculated and experimental structure factor magnitudes at which to stop the fitting procedure. The three protocols for addressing the halting problem were: the ‘power function method’; the ‘asymptotic extrapolation procedure (TIH)’; and the ‘conservative method’.

The halting problem

The *power function method* is based on the surprising empirical observation that the goodness-of-fit GoF^2 as a function of the Lagrange multiplier λ (which determines the desired level of fit in the XCW model) could be well represented by a power function. Although theoretical arguments show that the scale parameter A associated with the power function fit may be associated with a measure of the accuracy of the estimated standard uncertainties in measurements, there is little justification for the termination point from this procedure ($\lambda = 1$) unless the scale and units of the GoF^2 term are clarified.

The *asymptotic extrapolation procedure* is a modification of one used in density functional theory developed by Tozer *et al.* (1996) (we also refer to this as the TIH method) and as such, it is theoretically well justified. Notwithstanding that it yields termination points for the fitting procedure which are practically achievable, it is disappointing that application to our data does not clearly satisfy the asymptotic conditions on which it is based. Encouraging, however, is that the D parameter in this method, which is theoretically associated with errors in the measurement, is remarkably highly correlated with the A scale parameter in the power function method, with a correlation coefficient of 0.97, obtained using the Kamiński *et al.* (2014) data sets; and even the scale factor relating these two quantities is nearly unity [0.98 (8)]. Except for one outlier, the A values are also well correlated with the GoF^2 values from each individual data set in the joint refinement (at the TIH6 halting point).

The third *conservative method* emerged from considerations to avoid overfitting with the availability of multiple data sets, or possibly a single data set split into several subsets. In this case, the fitting process is stopped at the largest (or worst value) of the GoF^2 obtained from any of the individual (sub) data sets.

Reproducibility of changes in the electron density features due to fitting

For α -oxalic acid using the Kamiński *et al.* (2014) data sets, we find that the reconstructed electron-density from any of the halting problem protocols (including the method to halt at the best possible GoF^2 before convergence ceases) are very similar to each other. In the valence electron regions differences only amount to about $0.025 \text{ e } \text{\AA}^{-3}$. For this system, and for electron density features, this means that fitting to the best possible GoF^2 before convergence ceases as normally done is justified. However, for other features such as bond critical points which are more sensitive to the model (Landeros-Rivera *et al.*, 2021) or in cases where it is known that overfitting occurs this may not be the case so

that the halting protocols introduced here could be reinvestigated.

Regarding the specific features in the change of the electron density compared to previous work and high-level quantum mechanical calculations, we found that:

(i) The strong build-up in charge on the H atom in the strong hydrogen bond of oxalic acid seen in the original paper (Grimwood & Jayatilaka, 2001) was indeed due to the use of an incorrect and too-short O—H bond length; when using HAR geometric and atomic displacement parameters, this feature completely disappeared.

(ii) High-level calculations on a six-water cluster moiety nevertheless showed a small enhancement of the electron density on the hydrogen-bonded H atom (see Fig. 10), consistent with a correction to the over-estimation of ionicity, so that fitting to the experimental data appears to correct a known problem associated with the Hartree–Fock model wavefunctions.

(iii) An estimate of the crystal field effects from a point charge model surrounding the six-water cluster of molecules found a depletion of electron density on the hydrogen-bonded H atom and an increase in electron density in the corresponding O—H bond; but this was still not enough to bring the high-level calculations into agreement with the electron densities from the fitted XCW wavefunction.

In relation to other qualitative features in the changes in the electron density due to the XCW fitting process, we found that:

(i) The features in the electron density from the XCW fitting of the simultaneous refinement of all data sets is a consensus of the features of those from the individual refinements.

(ii) The electron density differences being modelled by the XCW are very small in the valence region, amounting to a few multiples of $0.025 \text{ e } \text{\AA}^{-3}$, which is consistent with the fact that the theoretical and experimental data agree before fitting to no worse than 2% difference, and on average within 1.5 estimated standard uncertainties; that is, the XCW procedure is fitting at the limits of experimental detectability.

(iii) Notwithstanding the previous point, some consistent systematic effects were modelled across all 14 data sets: a consistent decrease in the bonding density in the C—C bond, a depletion of charge on the lone pair of the O atom in the hydroxyl group, a decrease in the lone pairs on the O atom in the carbonyl group, and an increase in the electron density in the outer valence region of the C atoms.

(iv) The increase in the electron density in the outer valence region of the C atom clearly present in the high-level theoretical calculations is much less defined in the XCW plots; the depletion of electron density in the middle of both the O—H bond and the carbonyl bond all present in the theoretical calculations are completely absent in the XCW plots; and the strong depletion in electron density in the core of the C atom in the theoretical calculations is an increase in electron density in the XCW plots. These effects point to unmodelled systematic features and errors in the data or the procedure, which must be studied in more detail in future investigations.

(v) There are a plethora of other features in the plots from the high level quantum mechanical calculations, mostly in the small regions near the core of the atoms (which would require high resolution to see) or in low electron density regions (which would not produce a strong scattering signal).

Overall conclusions

One clear shortcoming of this work is the lack of robust techniques for estimating errors associated with parameters and properties derived from the XCW regularization-fitting procedure (Bürgi & Genoni, 2022). Indeed, it is for this reason we have been careful to describe the methods developed here as ‘protocols’, which, although well defined, clearly require further verification and fine-tuning.

On the other hand, notwithstanding that there have been very significant improvements in the quality and accuracy of the experimental data (as judged from the system studied here, and the fact that the theoretical HAR electron density already agrees extremely well with the experimental data even before XCW fitting) the presence of unmodelled or incorrectly modelled systematic features in the electron density derived from the XCW is a now pressing concern that should spur a review of data-treatment protocols – especially those concerned with accuracy and error estimation. If not, we fear that there will be very little experimental information in the X-ray data for the XCW procedure to fit to! This would be a pity so soon after the proclamation by Coppens (2005) that the X-ray charge density field has ‘come of age’.

Acknowledgements

We thank Professor Hans-Beat Bürgi for his critical assessment of this manuscript and for his idea to perform a joint refinement of all data sets. Above and beyond that, we thank him for very stimulating and enlightening discussions over the past two decades. We salute him for his career. DJ in particular acknowledges HBB’s contribution in inducting him into the dark arts of crystallography. We would like to acknowledge the critical and insightful comments of all referees. We would like to acknowledge referee 1 for their suggestion of using a Padé approximant. MLD was supported by an Australian Government Research Training Program (RTP) Scholarship.

Funding information

Funding for this research was provided by: Australian Government Research Training Program (RTP) Scholarship.

References

Boyd, R. J. & Wang, L.-C. (1989). *J. Comput. Chem.* **10**, 367–375.
Brünger, A. T. (1992). *Nature*, **355**, 472–475.
Bučinský, L., Jayatilaka, D. & Grabowsky, S. (2019). *Acta Cryst.* **A75**, 705–717.
Bürgi, H.-B. & Genoni, A. (2022). *Acta Cryst.* **B78**, 298–304.
Bytheway, I., Grimwood, D. J., Figgis, B. N., Chandler, G. S. & Jayatilaka, D. (2002). *Acta Cryst.* **A58**, 244–251.
Capelli, S. C., Bürgi, H.-B., Dittrich, B., Grabowsky, S. & Jayatilaka, D. (2014). *IUCrJ*, **1**, 361–379.
Coppens, P. (2005). *Angew. Chem. Int. Ed.* **44**, 6810–6811.

Coppens, P., Dam, J., Harkema, S., Feil, D., Feld, R., Lehmann, M. S., Goddard, R., Krüger, C., Hellner, E., Johansen, H., Larsen, F. K., Koetzle, T. F., McMullan, R. K., Maslen, E. N. & Stevens, E. D. (1984). *Acta Cryst.* **A40**, 184–195.
Dacombe, M. (2013). *Acta Cryst.* **A69**, 210–239.
Davidson, M. L., Grabowsky, S. & Jayatilaka, D. (2022). *Acta Cryst.* **B78**, 312–332.
Efron, B. & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.
Ernst, M., Genoni, A. & Macchi, P. (2020). *J. Mol. Struct.* **1209**, 127975.
Frisch, M. *et al.* (2009). *Gaussian 09*, Revision D. 01. Gaussian Inc., Wallingford, CT, USA.
Fugel, M., Jayatilaka, D., Hupf, E., Overgaard, J., Hathwar, V. R., Macchi, P., Turner, M. J., Howard, J. A. K., Dolomanov, O. V., Puschmann, H., Iversen, B. B., Bürgi, H.-B. & Grabowsky, S. (2018). *IUCrJ*, **5**, 32–44.
Genoni, A. (2013). *J. Chem. Theory Comput.* **9**, 3004–3019.
Genoni, A., Dos Santos, L. H. R., Meyer, B. & Macchi, P. (2017). *IUCrJ*, **4**, 136–146.
Goerigk, L., Hansen, A., Bauer, C., Ehrlich, S., Najibi, A. & Grimme, S. (2017). *Phys. Chem. Chem. Phys.* **19**, 32184–32215.
Grabowsky, S., Genoni, A., Thomas, S. P. & Jayatilaka, D. (2020). *21st Century Challenges in Chemical Crystallography II: Structural Correlations and Data Interpretation*, pp. 65–144. Springer.
Grimwood, D. J. (2002). PhD thesis, The University of Western Australia.
Grimwood, D. J. & Jayatilaka, D. (2001). *Acta Cryst.* **A57**, 87–100.
Helgaker, T., Jorgensen, P. & Olsen, J. (2014). *Molecular Electronic Structure Theory*. John Wiley & Sons.
Jayatilaka, D. (1998). *Phys. Rev. Lett.* **80**, 798–801.
Jayatilaka, D. & Dittrich, B. (2008). *Acta Cryst.* **A64**, 383–393.
Jayatilaka, D. & Grimwood, D. J. (2001). *Acta Cryst.* **A57**, 76–86.
Jayatilaka, D. & Grimwood, D. J. (2003). In *Int. Conf. Comput. Sci.* pp. 142–151 Springer.
Kamiński, R., Domagała, S., Jarzemska, K. N., Hoser, A. A., Sanjuan-Szklarz, W. F., Gutmann, M. J., Makal, A., Malińska, M., Bąk, J. M. & Woźniak, K. (2014). *Acta Cryst.* **A70**, 72–91.
Kleemiss, F., Dolomanov, O. V., Bodensteiner, M., Peyerimhoff, N., Midgley, L., Bourhis, L. J., Genoni, A., Malaspina, L. A., Jayatilaka, D., Spencer, J. L., White, F., Grundkötter-Stock, B., Steinhauer, S., Lentz, D., Puschmann, H. & Grabowsky, S. (2021). *Chem. Sci.* **12**, 1675–1692.
Landeros-Rivera, B., Contreras-García, J. & Dominiak, P. M. (2021). *Acta Cryst.* **B77**, 715–727.
Macetti, G., Macchi, P. & Genoni, A. (2021). *Acta Cryst.* **B77**, 695–705.
Parr, R. G. & Yang, W. (1994). *Density-Functional Theory of Atoms and Molecules*. Oxford University Press.
Podhorský, M., Bučinský, L., Jayatilaka, D. & Grabowsky, S. (2021). *Acta Cryst.* **A77**, 54–66.
Pulay, P. (1980). *Chem. Phys. Lett.* **73**, 393–398.
Pulay, P. (1982). *J. Comput. Chem.* **3**, 556–560.
Sanjuan-Szklarz, W. F., Woińska, M., Domagała, S., Dominiak, P. M., Grabowsky, S., Jayatilaka, D., Gutmann, M. & Woźniak, K. (2020). *IUCrJ*, **7**, 920–933.
Sherman, R. (1971). *Biometrika*, **58**, 396–398.
Spackman, P. R., Turner, M. J., McKinnon, J. J., Wolff, S. K., Grimwood, D. J., Jayatilaka, D. & Spackman, M. A. (2021). *J. Appl. Cryst.* **54**, 1006–1011.
Tozer, D. J., Ingamells, V. E. & Handy, N. C. (1996). *J. Chem. Phys.* **105**, 9200–9213.
Vetterling, W. T., Press, W. H., Teukolsky, S. A. & Flannery, B. P. (2002). *Numerical Recipes Example Book (C++): the Art of Scientific Computing*. Cambridge University Press.

- Weigend, F. & Ahlrichs, R. (2005). *Phys. Chem. Chem. Phys.* **7**, 3297–3305.
- Wieduwilt, E. K., Macetti, G. & Genoni, A. (2021). *J. Phys. Chem. Lett.* **12**, 463–471.
- Williams, T., Kelley, C., Bersch, C., Bröker, H.-B., Campbell, J., Cunningham, R., Denholm, D., Elber, G., Fearick, R., Grammes, C., *et al.* (2019). http://www.gnuplot.info/docs_5.
- Woińska, M., Grabowsky, S., Dominiak, P. M., Woźniak, K. & Jayatilaka, D. (2016). *Sci. Adv.* **2**, e1600192.
- Wozniak, K. (2021). Private Communication.
- Zhao, Q. & Parr, R. G. (1993). *J. Chem. Phys.* **98**, 543–548.
- Zobel, D. (1996). Private communication.
- Zobel, D., Luger, P., Dreissig, W. & Koritsanzsky, T. (1992). *Acta Cryst.* **B48**, 837–848.