nature reviews physics

https://doi.org/10.1038/s42254-023-00655-3

Expert recommendation

Check for updates

How to verify the precision of density-functional-theory implementations via reproducible and universal workflows

Emanuele Bosoni[1], Louis Beal[2], Marnik Bercx[3], Peter Blaha[4], Stefan Blügel[5], Jens Bröder[5,6], Martin Callsen[7,8,9], Stefaan Cottenier[7,8], Augustin Degomme[2], Vladimir Dikan[1], Kristjan Eimre[3], Espen Flage-Larsen[10,11], Marco Fornari[12], Alberto Garcia[1], Luigi Genovese[2], Matteo Giantomassi[13], Sebastiaan P. Huber[3,14], Henning Janssen[5], Georg Kastlunger[15], Matthias Krack[16], Georg Kresse[17,18], Thomas D. Kühne[19,20], Kurt Lejaeghere[8,21], Georg K. H. Madsen[4], Martijn Marsman[17,18], Nicola Marzari[3,16], Gregor Michalicek[5], Hossein Mirhosseini[22], Tiziano M. A. Müller[23], Guido Petretto[13], Chris J. Pickard[24,25], Samuel Poncé[13], Gian-Marco Rignanese[13], Oleg Rubel[26], Thomas Ruh[4,8,27], Michael Sluydts[7,8,28], Danny E. P. Vanpoucke[7,29], Sudarshan Vijay[15], Michael Wolloch[17,18], Daniel Wortmann[5], Aliaksandr V. Yakutovich[30], Jusong Yu[3,16], Austin Zadoks[3], Bonan Zhu[31,32] & Giovanni Pizzi[3,16] ✉

Abstract

Density-functional theory methods and codes adopting periodic boundary conditions are extensively used in condensed matter physics and materials science research. In 2016, their precision (how well properties computed with different codes agree among each other) was systematically assessed on elemental crystals: a first crucial step to evaluate the reliability of such computations. In this Expert Recommendation, we discuss recommendations for verification studies aiming at further testing precision and transferability of density-functional-theory computational approaches and codes. We illustrate such recommendations using a greatly expanded protocol covering the whole periodic table from $Z = 1$ to 96 and characterizing 10 prototypical cubic compounds for each element: four unaries and six oxides, spanning a wide range of coordination numbers and oxidation states. The primary outcome is a reference dataset of 960 equations of state cross-checked between two all-electron codes, then used to verify and improve nine pseudopotential-based approaches. Finally, we discuss the extent to which the current results for total energies can be reused for different goals.

Sections

A full list of affiliations appears at the end of the paper. ✉e-mail: giovanni.pizzi@psi.ch

Key points

- Verification efforts are critical to assess the reliability of density-functional theory (DFT) simulations and provide results with properly quantified uncertainties.
- Developing standard computation protocols to perform verification studies and publishing curated and FAIR reference datasets can greatly aid their use to improve codes and computational approaches.
- The use of fully automated workflows with common interfaces between codes can guarantee uniformity, transferability and reproducibility of results.
- A careful description of the numerical and methodological details needed to compare with the reference datasets is essential; we discuss and illustrate this point with a dataset of 960 all-electron equations of state.
- Reference datasets should always include an explanation of the target property for which they were generated, and a discussion of their limits of applicability.
- Further extensions of DFT verification efforts are needed to cover more functionals, more computational approaches and the treatment of magnetic and relativistic (spin–orbit) effects. They should also aim at concurrently delivering optimized protocols that not only target ultimate precision, but also optimize the computational cost for a target accuracy.

Introduction

The fast improvement of hardware, methods and tools for density-functional theory (DFT) calculations in periodic boundary conditions has greatly advanced the field of condensed matter physics and computational materials science, paving the way for an effective use of the materials design process that accelerates the discovery, development and deployment of new materials, thanks to the aid of simulations[1,2]. Efficient software infrastructures[3–13] now enable high-throughput calculations of a panoply of material properties, which are often made available to the public in large repositories[14–21]. Most datasets aspire to be findable, accessible, interoperable and reusable (FAIR)[22] in order to accelerate materials discovery, possibly with the aid of machine learning. They can be queried with ad hoc application programming interfaces (APIs) or, for many of them, via a single common API, thanks to the recent efforts of the OPTIMADE[23] consortium. Full integration of different data is, however, often limited by considerations related to uncertainty quantification[1,24–27]. In this work, we discuss recommendations for quantifying to what extent properties (total energies and derived quantities) obtained by different DFT codes agree with each other.

In principle, DFT applies the fundamental laws of quantum physics to predict properties of a material, with no other inputs than the chemical composition and the crystal structure. In reality, the electronic-structure calculations involve a variety of choices to solve the equations prescribed by DFT and introduce several levels of approximation. Those choices, reflected in the resulting data, range from the specific flavour of DFT (such as the approach used for the exchange–correlation functional) to the discretization assumptions (like the basis set), and to the specific computational parameters needed by the codes. Some approaches are more reliable, and therefore often slower, whereas others make more substantial approximations to gain computational speed and enable the study of systems with more atoms. Furthermore, even when formally the same choices have been made in different codes, these may provide slightly different results owing to the details of their implementations. We thus encourage users to report in their publications the complete details of the codes and parameters used and, whenever possible, to refer to quantitative sources on the reliability, accuracy and precision of the chosen computational approach (Box 1).

The importance of verifying the precision of codes has been long recognized[28]. Despite this, when considering DFT codes that adopt periodic boundary conditions, a first systematic assessment of their precision was only performed in 2016, where the consistency of 40 computational approaches was assessed by calculating the equation of state (EOS; the energy-versus-volume curve) for a test set of 71 elemental crystals[29] (benchmark results from ref. 29, plus data collected later on, were stored on https://molmod.ugent.be/deltacodesdft (obsolete) and are now archived on Materials Cloud Archive[30]). This 'Δ-project' led to the conclusion that the mainstream codes were in very good agreement with each other, which was not the case a decade before. Despite already being a large project by itself, the Δ-project was only the first step towards a careful verification of DFT calculations, which requires a much larger diversity of structural and chemical variables[29].

In this Expert Recommendation, we list a set of guiding principles to perform new verification studies of DFT calculations, and we illustrate examples of verification by using a curated reference set of highly converged results for the EOS of 960 crystals, with two independent state-of-the-art all-electron (AE) DFT codes (FLEUR[31,32] and WIEN2k[33,34]). These 960 crystals cover all elements and a wide variety of structural and chemical environments in the form of four unary compounds and six oxides. The resulting data are shared on the Materials Cloud[19] according to the FAIR[22] principles. A key feature of our work is that the thousands of computations performed are implemented within a reproducible and automatic infrastructure. Specifically, the launching and management of all the DFT calculations is carried out using AiiDA[3,4,35]. The choice of code-specific inputs and numerical parameters (called 'protocols' in the following) are implemented in the publicly available `aiida-common-workflows` (ACWF) package[36,37] together with a number of error handlers to recover automatically from typical failure modes of each code. This setup enables new datasets to be to easily generated and allows the current work to be extended for the verification of other computational approaches.

To make correct use of the dataset, choices of numerical parameters (such as the smearing type and size, or the k-point integration mesh) must be performed consistently. One of our recommendations for verification efforts is to develop metrics to quantify discrepancies between codes that depend on physically measurable quantities. Here, we define two new metrics (in addition to the Δ metric introduced in ref. 38) to aid quantitative comparison of EOS results for pairs of codes or computational approaches, and we discuss their benefits. Using these metrics, we compare the EOS results of our reference dataset to the results obtained by several pseudopotential codes. The latter are designed to enhance computational efficiency by considering explicitly only valence electrons, which contribute to bonding[39–41]. The codes considered here are: ABINIT[13,42,43], BigDFT[44], CASTEP[45],

CP2K[46,47], GPAW[48,49], Quantum ESPRESSO[50,51], SIESTA[52,53], the SIRIUS[54] library (via its CP2K interface) and VASP[55,56].

We stress that the aim of this study is not to provide a ranking or to evaluate the quality of different codes, but to illustrate with a few examples the value of curated datasets generated following our recommendations. In particular, we illustrate its use to improve existing pseudopotentials and to assess the consistency of results of several computational approaches to compute the EOS within DFT. Looking forwards, we suggest covering more exchange–correlation functionals, computational approaches, and treatment of magnetic and relativistic (spin–orbit) effects. However, future studies should not only target ultimate precision, but also aim at delivering protocols that optimize the computational cost for a target accuracy. Note that we limit all discussions to verification efforts: how precise codes are in reproducing the ideal theoretical results given by DFT (for a given choice of exchange–correlation functional). We do not discuss validation: the accuracy of these simulations with respect to the experimental results. Although this is also a highly relevant topic, it is beyond the scope of this Expert Recommendation.

AE reference dataset for EOS parameters

By using a reference dataset of EOS calculations, we illustrate, with a practical example, our recommendations for performing verification studies of DFT calculations (Box 2). The results are obtained with the AE codes FLEUR and WIEN2k, using the PBE exchange–correlation functional[57], and considering the scalar-relativistic approximation (no spin–orbit coupling) for the orbitals treated as valence states (including the heavy elements). The two codes use the linearized augmented plane waves plus local orbitals method, but differ in details of the basis set and some computational setup parameters.

Crystal-structures dataset

We compute the EOS on a dataset of 960 cubic crystal structures. To provide a chemically comprehensive dataset, we consider all elements in the periodic table from $Z = 1$ (hydrogen) to $Z = 96$ (curium). Furthermore, we systematically scan structural diversity and investigate the transferability to more complex chemical environments by examining, for each element, four mono-elemental cubic crystals ('unaries dataset') and six cubic oxides ('oxides dataset').

Specifically, the unaries dataset considers all elements in the face-centred-cubic (FCC), body-centred-cubic (BCC), simple-cubic (SC) and diamond crystal structure, thus covering a wide range of coordination numbers (12, 8, 6 and 4, respectively); a total of 384 systems (Supplementary Table 1.1). The oxides dataset is composed of six cubic oxides for each of the 96 elements X, with chemical formula X_2O, XO, X_2O_3, XO_2, X_2O_5 and XO_3, thus totalling 576 additional structures, whose crystal structures are detailed in Supplementary Table 1.2. The oxide stoichiometries are chosen such that the formal oxidation state of the element considered varies from +1 to +6. The actual oxidation state, as determined from calculated atomic charges, is typically different (see the discussion with a Hirshfeld-I[58–60] analysis in Supplementary Section 2), but the two quantities show a good correlation, with the actual oxidation state (or the calculated charges) being on average about half the formal oxidation state. The X–O distance varies rather systematically over these six oxides (Supplementary Fig. 1.3), typically with the smallest distance for XO_3 and the largest for X_2O. This indicates that XO_3 could be a proxy for systems with very short bond lengths, such as in high-pressure studies. The two datasets of unaries and oxides (jointly called 'full dataset'; a total of 960 systems) complement each other in covering chemical and structural variety for each element.

In addition to the criteria above, all structures have been chosen to be cubic and such that forces on all atoms are zero by symmetry. Therefore, the only free parameter is the unit cell volume V or, equivalently, the lattice parameter. As a consequence, the EOS results can be compared with any code able to compute total energies, with no requirement on the capability of computing forces and stresses. It is important to note that most structures are not stable in nature (in particular under our constraint of cubic spacegroup symmetry). Still, they can be used to assess that all codes reproduce the same DFT result, with the advantage of providing a consistent set across the whole periodic table.

Computation of EOS parameters and comparison between AE codes

The EOS has been traditionally used to determine the computational parameters and study convergence of DFT calculations. By fitting the DFT energy $E(V)$ versus cell volume V to an EOS, it is possible to extract the theoretical predictions of the equilibrium volume V_0, the bulk modulus B_0 and its derivative with respect to the pressure, B_1. The Birch–Murnaghan EOS[61]

$$E(V) = E_0 + \frac{9V_0B_0}{16}\left\{\left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^3 B_1 + \left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^2\left[6 - 4\left(\frac{V_0}{V}\right)^{\frac{2}{3}}\right]\right\} \quad (1)$$

(where, in addition to V_0, B_0 and B_1, E_0 is also a fitting parameter denoting the bottom of the EOS curve) was used in the Δ-project[29,30], and we

Box 1

Recommendations for users of DFT codes

- When publishing research that makes use of DFT codes, always report all the necessary information that guarantees the reproducibility of the simulations. This includes the code version and all the essential numerical parameters of the calculations (such as k-point integration mesh and smearing, basis-set type and size, or plane-wave cutoffs and so on).
- Always cite the exact pseudopotentials that are used in published simulations, including the exchange–correlation functional, the library from which they were obtained and the exact library version. Lack of this information results in essentially non-reproducible simulations.
- Refer to quantitative sources that document the precision of the numerical implementation (all-electron versus pseudopotential, basis-set type and size, and so on).
- Include a validation statement that refers to the accuracy of the chosen exchange–correlation functional to correctly and accurately address the physics at hand. Note that, however, this is beyond the scope of the current Expert Recommendation focusing on the precision of numerical implementations.

follow the same approach by performing a fit of $E(V)$ of equation (1) using calculations of the total energy corresponding to seven equidistant constant volumes between 94% and 106% of a reference central volume $\tilde{V}_0$ (for each structure). We emphasize that the results are sensitive to the precise choice of volume range, of reference central volume, and even of fitting algorithm (Supplementary Section 3). In this work, the reference central volumes $\tilde{V}_0$ for each of the 960 crystals have been chosen after an iterative process of performing more and more accurate simulations with the two AE codes considered here, until the difference between the reference central volume and the equilibrium volume of the EOS fit was smaller than the 2% volume spacing between total-energy calculations. These central reference volumes are tabulated in Supplementary Section 1, and the corresponding crystal-structure files are available in the data entry associated to this Recommendation[62]. These volumes have no physical significance, but for precise comparison between computational approaches, each of them should use the same reference volumes.

The results obtained with the AE codes FLEUR and WIEN2k constitute our reference data. Figure 1 shows the distributions of the percentage difference between FLEUR and WIEN2k for V_0, B_0 and B_1 with respect to their average. For instance, the V_0 difference (in %) is given by:

$$100 \times \frac{V_0^{\mathrm{WIEN2k}} - V_0^{\mathrm{FLEUR}}}{(V_0^{\mathrm{WIEN2k}} + V_0^{\mathrm{FLEUR}})/2}. \quad (2)$$

Box 2

How to perform verification studies of DFT calculations

- Quantitatively estimate the precision of DFT computational approaches and implementations with respect to exact numerical results. Provide adequate details of the verification protocols to ensure reproducibility of the results and correct reuse in data-driven research, for example, clarifying their range of applicability and specifying which parameters need to be fixed — independent of the approach — to ensure comparable results.
- Develop fully automated workflows to guarantee uniformity and transferability of parameters between computational approaches. This includes the definition and use of 'standard protocols', such as automated selection of numerical parameters — often specific to each computational approach — that ensure numerically precise results.
- Publish curated reference datasets from systematic verification studies. Help them to be used to improve other codes by making the datasets FAIR: findable and accessible on open repositories, interoperable by using standard formats and clear annotations, and reusable by specifying all parameters needed to reproduce the results.
- Organize the reference data in appropriate subsets by recognizing the diversity of focus and the non-uniform capabilities of available computational approaches (some systems may require additional effort to be supported by all codes).

Although the histograms do not carry material-specific information, they clearly highlight the agreement between the two AE codes. The relative difference in the equilibrium volume is below 0.3% for all the materials except for five oxides (Supplementary Section 4; raw data available in ref. 62). The discrepancies for B_0 and B_1 are larger; this is not surprising, because they originate from higher derivatives of the EOS curves (Supplementary Section 3). We emphasize that these values, obtained after careful convergence of all numerical parameters related to the basis-set choices in the two codes, are of extremely high precision, with a spread that can even be an order of magnitude smaller than the typical discrepancies that we observe between pseudopotential codes.

The complete list of numerical parameters used for the AE calculations is presented in Supplementary Section 5. We highlight here that the exact choices of the electronic-state smearing and of the k-point integration mesh, as well as the specific quantity considered as the energy $E(V)$ (internal energy, or free energy including the entropic smearing contribution as we do here), are of crucial importance for a reliable comparison among codes and must be fully consistent (Box 3).

Average all-electron dataset as the reference for further studies

In addition to the data for each of the two codes and in order to provide a single comparison reference, we also provide a reference 'average AE dataset' obtained by averaging the values of V_0, B_0 and B_1 for each of the 960 systems in the full dataset. The corresponding values are in Supplementary Section 4 and published according to the FAIR principles in ref. 62. Considering the very good agreement between the two codes, this average dataset constitutes an excellent reference, and we use this average to compare with the pseudopotential codes later. In addition, if error bars are desired, the spread between the results of the two AE codes can be used as an estimate of our dataset precision.

Metrics for equation-of-state comparison

In the Δ-project[29,30], the Δ metric was used to compare the EOS computed with two different DFT computational approaches a and b. There, $\Delta = \Delta(a,b)$ was defined as:

$$\Delta(a,b) = \sqrt{\frac{1}{V_{\mathrm{M}} - V_{\mathrm{m}}} \int_{V_{\mathrm{m}}}^{V_{\mathrm{M}}} [E_a(V) - E_b(V)]^2 \, \mathrm{d}V}, \quad (3)$$

where $E_a(V)$ and $E_b(V)$ are the Birch–Murnaghan fits of the datapoints obtained from approaches a and b, respectively, the two EOS curves have been lined up with respect to their minimum energy, and as discussed earlier the integral spans a ±6% volume range centred at a central volume $\tilde{V}_0$ (with $\tilde{V}_0$ values tabulated in Supplementary Section 1), such that the minimum volume is $V_{\mathrm{m}} = 0.94\tilde{V}_0$ and the maximum volume is $V_{\mathrm{M}} = 1.06\tilde{V}_0$.

The use of a single metric to compare two EOS curves simplifies the data analysis, as it can be used instead of the difference of the Birch–Murnaghan parameters V_0, B_0 and B_1, as we did in Fig. 1. However, the value of $\Delta(a,b)$, with units of energy, has the shortcoming of being too sensitive to the value of the bulk modulus of the material: visually similar discrepancies between two curves result in larger Δ values for materials with larger B_0. This was already recognized in ref. 63, in which a modified metric Δ_1 was suggested, renormalized to a reference value of V_0 and B_0. In addition, the $E_a(V)$ and $E_b(V)$ quantities in equation (3) are typically renormalized by the number of atoms in the unit cell, to provide a 'Δ/atom' metric, independent of the choice of

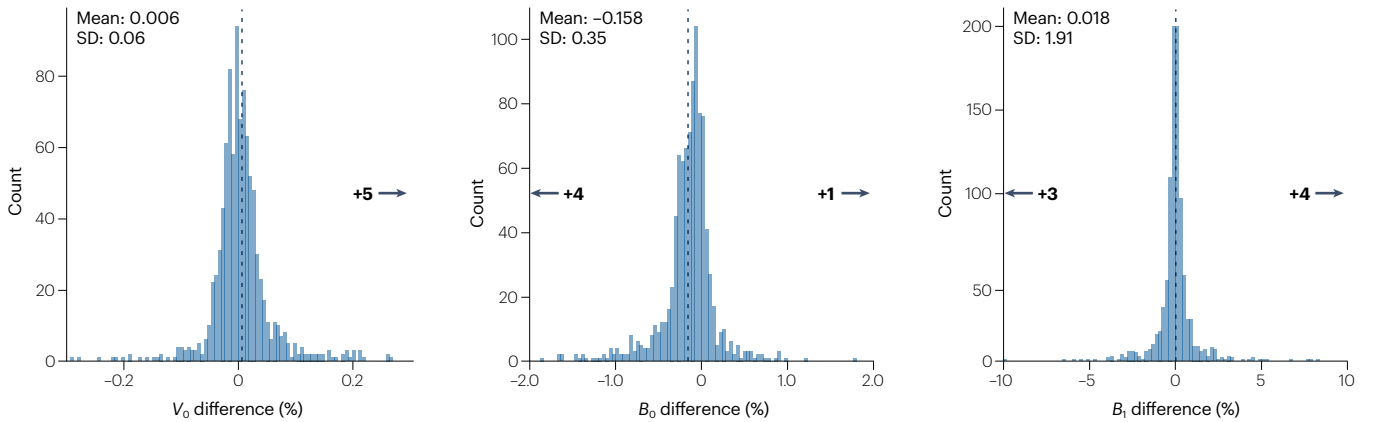

Fig. 1 | Differences in results for codes FLEUR and WIEN2k. Histograms show the percentage difference between the results of the two all-electron codes (FLEUR and WIEN2k) with respect to their average for the three parameters of the EOS: equilibrium volume V_0, bulk modulus B_0, and its derivative with respect to the pressure, B_1, for the full dataset of unaries and oxides. Positive values indicate larger values for WIEN2k than FLEUR. Mean and standard deviation (SD) of the distributions are reported at the top left of each panel. The numbers near the arrows indicate the number of outliers outside the *x*-axis range. The relative difference in V_0 is below 0.1% for 93% of all structures in our dataset; the relative difference in B_0 is below 1% for 97% of the structures; and the relative difference in B_1 is below 2% for 92% of the structures. The five outliers for V_0 are NeO_3 (0.302%), RbO_3 (0.343%), Cs_2O_5 (0.323%), Fr_2O_5 (0.645%) and Ra_2O_5 (0.333%), corresponding to lattice parameters for FLEUR/WIEN2k of 4.320/4.324 Å, 4.783/4.789 Å, 6.247/6.254 Å, 6.120/6.133 Å and 6.238/6.244 Å, respectively.

the simulation cell size. As we expand our analysis to two-component oxides, generalizations might be required (such as by normalizing instead per formula unit).

Here, we propose two new metrics that we label ε and v, and we discuss their pros and cons. We first define the following shorthand notation for the integral average of a quantity $f(V)$ over the volume range $[V_m, V_M]$:

$$\langle f \rangle = \frac{1}{V_M - V_m} \int_{V_m}^{V_M} f(V)\, \mathrm{d}V. \tag{4}$$

Using this notation, we can simply write $\Delta(a,b) = \sqrt{\langle [E_a(V) - E_b(V)]^2 \rangle}$. The first metric $\varepsilon(a,b)$ that we define is a renormalized dimensionless version of Δ:

$$\varepsilon(a,b) = \sqrt{\frac{\langle [E_a(V) - E_b(V)]^2 \rangle}{\sqrt{\langle [E_a(V) - \langle E_a \rangle]^2 \rangle \langle [E_b(V) - \langle E_b \rangle]^2 \rangle}}}. \tag{5}$$

This metric, similar to the Δ_1 of ref. 63 or the subsequently defined Δ_{rel} available in the DeltaCodesDFT package[30], is insensitive to the magnitude of the bulk modulus (Supplementary Section 6). In addition, it is independent of the use of a 'per formula unit' or 'per atom' definition of the EOS (Supplementary Section 6). Therefore, $\varepsilon(a,b)$ provides a uniform metric across the variety of structural and chemical environments under investigation, given the requirement that it must be calculated with the same relative volume range for every material. As the list of central reference volumes has been fixed (Supplementary Section 1), and as we use the same ±6% volume range as in refs. 29,30, the 960 intervals $[V_m, V_M]$ are unambiguously defined. We highlight, in passing, that the discrete form of equation (5),

$$\varepsilon(a,b) = \sqrt{\frac{\sum_i [E_a(V_i) - E_b(V_i)]^2}{\sqrt{\sum_i [E_a(V_i) - \langle E_a \rangle]^2 \sum_i [E_b(V_i) - \langle E_b \rangle]^2}}} \tag{6}$$

where the index i runs over the explicit calculations of $E(V)$ from DFT, provides a reasonably good approximation to the value of equation (5) as long as the minima of the $E_a(V_i)$ and $E_b(V_i)$ datapoints are aligned on the energy scale, with the advantage that it can be used to directly compare raw DFT total-energy data without requiring an EOS fitting. Nevertheless, we stress that in the rest of this work we use the expression of equation (5) and not its discrete version, equation (6). Equation (6) is grounded in the definition of the coefficient of determination (or R^2) in statistics[64] as a fraction of variance unexplained. We can interpret the value of $1 - \varepsilon^2$ as the coefficient of determination (that is, $1 - \varepsilon^2 \approx R^2$) in a situation when one EOS $E_a(V)$ is treated as a fit for the other EOS $E_b(V)$. More precisely, because we want to define a symmetric metric $\varepsilon(a,b) = \varepsilon(b,a)$, our ε^2 is the $1 - R^2$ value that one would obtain treating $E_a(V)$ as a fit for $E_b(V)$ and vice versa with the geometric mean of both data variances. We note that the interpretation $R^2 \approx 1 - \varepsilon^2$ holds in very good approximation when the value of ε is much smaller than 1 (which applies for very similar $E(V)$ curves). We discuss the sensitivity of ε to perturbations of the Birch–Murnaghan parameters in Supplementary Section 7. The main outcome is that ε is mostly sensitive to the variations of V_0, and much less sensitive to B_0 and B_1. For some applications, however, some of the parameters are more relevant than others (for example, if one is mostly interested in accurate bulk moduli). For these cases, we recommend defining metrics of discrepancy that depend directly on physically measurable quantities.

As an EOS is very well described by the three parameters V_0, B_0 and B_1, we thus suggest a second metric v that directly captures the relative difference of these three parameters between two computational approaches a and b, using appropriate weights $w_{V_0}, w_{B_0}, w_{B_1}$,

$$v_{w_{V_0}, w_{B_0}, w_{B_1}}(a,b) = 100 \sqrt{\sum_{Y = V_0, B_0, B_1} \left[w_Y \frac{Y_a - Y_b}{(Y_a + Y_b)/2} \right]^2}, \tag{7}$$

where, for instance, $(V_0)_a$ indicates the value of V_0 obtained by fitting the data of approach a, and so on. The (arbitrary) prefactor 100 is chosen to obtain values with similar order of magnitude to those of ε. It also helps in interpreting the value of v as an estimate of percentage errors (rather than relative errors) on the fit parameters. We highlight

Box 3

How to perform comparisons with the reference dataset

- Use the PBE exchange–correlation functional, do not include spin-polarization effects, and consider a scalar-relativistic treatment of (valence) electrons.
- Use Fermi–Dirac smearing with a value of 0.0045 Ry. Although this choice does not lead to zero-smearing results (which would require extrapolation and extremely dense k-point integration meshes), using the same values ensures that results are comparable. In extreme cases, using a different smearing may significantly affect the equilibrium volume and the overall shape of the EOS.
- Compute the equations of state (EOS) using as the proper variational functional the free energy $E-TS$, where E is the internal energy and $-TS$ is the smearing-energy entropic contribution. Other choices, such as the internal energy E, or the extrapolated energy for zero smearing (such as the expression $E-TS/2$, valid for Fermi–Dirac or Gaussian smearings[82,83]) can result in significant changes of the EOS, including large variations of the minimum-energy volume.
- Use the same protocol to fit the EOS curves: seven equally spaced points in a volume range of ±6% around the specified central volume. With these choices, values $\varepsilon \lesssim 0.06$ or $v \lesssim 0.1$ can be considered to indicate an excellent agreement, and $\varepsilon \lesssim 0.2$ or $v \lesssim 0.33$ a good agreement (with a noticeable but still relatively small discrepancy between them). A different volume range will affect these thresholds and require a different choice of weights for v to capture differences that are not purely statistical in nature. In addition, a different volume range will affect the k-point integration mesh (see next point).
- Use the exact same choice of the k-point integration mesh: regular grid including the Γ point and the smallest mesh guaranteeing a spacing between points of at most 0.06 $Å^{-1}$ along each of the three reciprocal-space directions for the smallest volume, and the same set of k-points (in scaled units) for all other volumes. This is typically converged for most systems and ensures that results can be compared even in the rare case of an unconverged k-point integration mesh.
- Do not transfer the choices performed for this reference dataset to a different context, as this might lead to incorrect conclusions. For instance, extracting formation energies from our reference dataset can provide inaccurate results, as the parameters used in our simulations are guaranteed to be consistent only for different volumes of the same material, not necessarily among different materials.

that v depends on the weights (w_y), which in turn could be chosen to satisfy particular applications. Here, because we aim to be application-agnostic, we choose weights based only on the sensitivity of our fitting procedure to random numerical noise applied to the energy values of the EOS datapoints. The detailed procedure to determine the weights is described in Supplementary Section 3; we just report here the final choice $w_{V_0} = 1, w_{B_0} = 1/20$ and $w_{B_1} = 1/400$. Intuitively, the reduced weights are consistent with the expected increase of numerical uncertainty propagated in the fit when estimating higher-order derivatives of the EOS. Similarly to ε, these weights also depend on the volume range of the datapoints used for the EOS fit, as well as the specific choice of the fitting algorithm (Supplementary Section 3). In the rest of the article, we refer to v assuming this specific choice of weights, that is, $v \equiv v_{1,1/20,1/400}$. In Supplementary Section 7 we also discuss an intuitive interpretation of the v metric: it is the percentage error on the equilibrium volume between the two approaches a and b, when B_0 and B_1 are the same in the two approaches; otherwise, when B_0 and B_1 differ, it corresponds to an equivalent percentage error on V_0 that would result in quantitatively similar changes to the EOS curve, within the ±6% volume range considered here.

The metrics ε and v allow us to compare a pair of codes for each material in the dataset. Figure 2 reports the results for the pair (FLEUR, WIEN2k) across the entire set of structures under investigation in the form of periodic tables, enabling a quick identification of the most problematic elements in each set. For instance, as one might expect, the agreement is generally worse for noble gases, which are weakly bonded systems with a very small bulk modulus and thus more susceptible to numerical errors due to the choice of the basis set and of other computational parameters.

We emphasize that with our choice of the volume range and weights for v, the two metrics provide very consistent information, highlighting the importance of properly defining metrics based on physically measurable quantities, and on careful analysis of the error propagation of the fitting procedure (Supplementary Section 7, where we discuss quantitatively the effect of perturbations on the EOS parameters to the values of ε, v and Δ). Indeed, although ε and v are constructed according to quite different principles, they turn out to contain nearly identical information (Supplementary Section 8, where we show that they are to a good extent linearly correlated for $v \lesssim 1$, with $v \approx 1.65\varepsilon$). This has the consequence that periodic tables for ε or v will be almost identical if the range of the colour scale is taken according to this linear correlation (as it is the case in Fig. 2 and is discussed in more detail in Supplementary Section 9).

Finally, we identify indicative thresholds on ε and v to represent an excellent agreement between two EOS curves if $\varepsilon \lesssim 0.06$ or $v \lesssim 0.1$, or a good agreement (noticeable, but still relatively small) if $\varepsilon \lesssim 0.2$ or $v \lesssim 0.33$ (Supplementary Section 7). As discussed earlier, we can interpret the two thresholds $\varepsilon = 0.06$ ($\varepsilon = 0.2$) approximately as a determination coefficient $R^2 \approx 1 - 0.06^2 = 0.9964$ ($R^2 \approx 1 - 0.2^2 = 0.96$) if one EOS is treated as a fit for the other. The data from the two AE codes shows an overall excellent agreement: only four systems out of 960 have one or both metrics outside the 'good agreement' range (Cs_2O_5, Fr_2O_5, Ra_2O_5 and RbO_3) when comparing the two AE codes of our reference dataset, and 883 out of 960 systems have an excellent agreement for both ε and v according to the thresholds discussed above.

Using the all-electron reference dataset

Here, we discuss several aspects that must be carefully considered when using the reference dataset, and we then show how the dataset

has been used to evaluate the precision of several computational approaches based on pseudopotentials and to improve a number of pseudopotential libraries.

Recommendations on how to use the dataset

When comparing our reference dataset with results from other codes, either for verification purposes or as a reference to improve basis sets and pseudopotentials, it is essential to use the same approximations (such as the exchange–correlation functional or the treatment of spin) and numerical choices (smearing and k-point integration mesh), as these parameters significantly affect the EOS results (Box 3).

All calculations are performed in periodic boundary conditions using the PBE[57] exchange–correlation functional, without including spin-polarization effects (non-magnetic calculations) and within a scalar-relativistic approximation (no spin–orbit coupling) for the orbitals treated as valence states. The reciprocal-space integration is performed with a Monkhorst–Pack uniform k-point grid including the Γ point, chosen as the smallest integration mesh guaranteeing a linear spacing of at most 0.06 $Å^{-1}$ in each of the three reciprocal-space directions for the smallest volume, and the same set of k-points (in scaled units) for all other volumes. For instance, this corresponds to a grid of 21 × 21 × 21 k-points for a simple-cubic primitive cell with lattice parameter of 5 Å. A Fermi–Dirac smearing of electronic states with a broadening of 0.0045 Ry (≈ 61.2 meV) is used in all cases, requiring the high-density k-points sampling mentioned above. In addition, the quantity $E(V)$ that is fitted with the Birch–Murnaghan expression of equation (1) is not the internal energy, but the free energy that includes the entropic contribution $-TS$ introduced by the smearing (where T is the effective temperature given by the smearing broadening). This corresponds to considering isothermal quantities for B_0 and B_1 (note that the temperature effect in these simulations do not mimic a physical temperature, but a fictitious temperature due to the electronic smearing).

We stress that, in general, two codes using a different smearing distribution are expected to return comparable results only in the limit of an infinite number of k-points and an infinitesimal smearing. However, for the purpose of verification, we do not need to reach this computationally expensive limit, provided that the same parameters

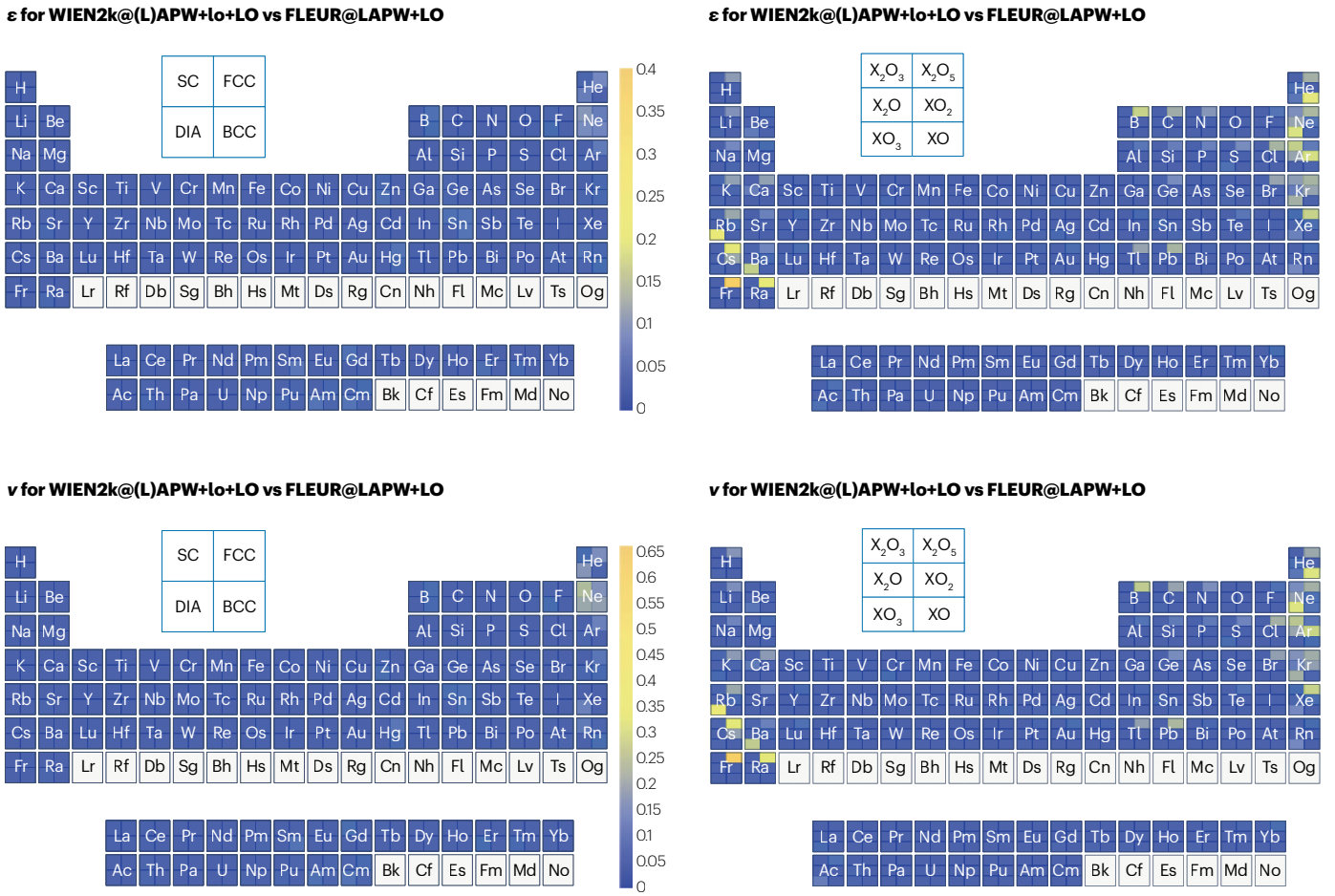

Fig. 2 | Discrepancy between the all-electron results obtained with WIEN2k and FLEUR in our reference dataset. Results are measured either with the ε metric (top panels) or the v metric (bottom panels), for all 96 elements considered. Each square for a given element is subdivided into four (six) in the left (right) panel, each referring to one of the unary (oxide) structures, as indicated in the key shown in each panel. The colour scale is the same for each pair (unaries and oxides) of periodic tables for the same metric (ε or v). All structures are within our threshold for good agreement except Cs_2O_5 ($\varepsilon = 0.20$, $v = 0.33$), Fr_2O_5 ($\varepsilon = 0.40$, $v = 0.66$), Ra_2O_5 ($\varepsilon = 0.21$, $v = 0.33$) and RbO_3 ($\varepsilon = 0.21$, $v = 0.37$). The AE codes are labelled with their code name, followed by an indication of the basis set they use. SC, simple cubic; FCC, face-centred cubic, BCC, body-centred cubic; DIA, diamond.

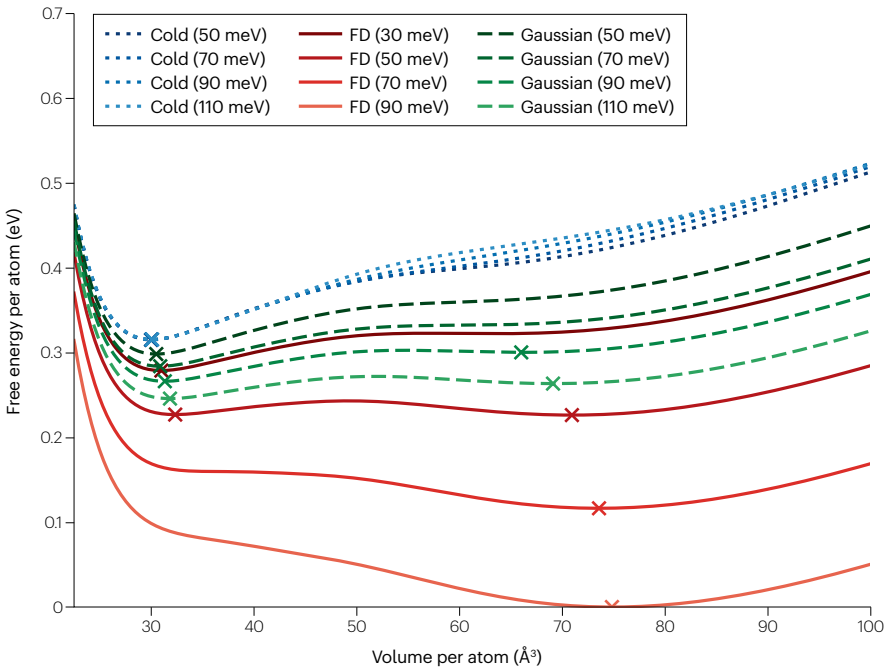

Fig. 3 | Effect of different choices of smearing on the equation of state of the artificial diamond structure of erbium computed with the Quantum ESPRESSO code. Erbium in the diamond structure is one of the systems in which the effect of smearing is most pronounced. In the legend, 'Cold', 'FD' and 'Gaussian' indicate, respectively, cold smearing[81], Fermi–Dirac smearing and Gaussian smearing. Two alternative minimum-energy solutions can exist for the FD and Gaussian smearings at very different volumes (as indicated by the crosses, where local minima exist in the curves); which one is selected depends on the choice of smearing and broadening. Note also the much-reduced sensitivity of the cold smearing to the broadening temperature, and how FD and Gaussian smearings are essentially equivalent after a renormalization of the FD broadening by a factor of about 2.565, as discussed in ref. 82.

among codes are chosen. As a consequence, our results should not be considered a prediction of the zero-temperature (no smearing) limit. We still highlight, however, that our choice of the *k*-point density results in a very dense and almost converged integration mesh (for fixed broadening): all values of V_0 computed with WIEN2k change by less than 0.07% when comparing with a denser *k*-point integration mesh with linear spacing of 0.045 $Å^{-1}$, except in two cases (RbO_3, 3.7% change; and HeO, 0.16% change). More details are reported in Supplementary Section 10.

To emphasize the sensitivity of the EOS to the choice of smearing, we show one of the most pathological cases of our dataset (Fig. 3). In erbium in the diamond structure, the EOS does not have a simple shape but displays instead, for the case of Fermi–Dirac and Gaussian smearing, two minima at very different volumes. Which one is favoured in energy depends on the type of smearing and the value of the broadening. This behaviour can be explained by the presence of a set of narrow *f* bands close to the Fermi level, whose filling strongly depends on the smearing (Supplementary Section 11). If we are after an improved erbium pseudopotential, trying to optimize it with a different smearing (and thus possibly for a different minimum) will result in an incorrect pseudopotential. A similar reasoning holds for the choice of using the free energy instead of the internal energy for the EOS (Supplementary Section 12). We also highlight that we adopted a scalar-relativistic treatment of valence electrons for our dataset. In most pseudopotential codes, this is obtained by simply using scalar-relativistic pseudopotentials, and the treatment is applied only for the valence electrons that are considered explicitly, while the treatment for the core electrons is implicitly included in the pseudopotential used. Even for AE codes, electrons are typically partitioned into a core (treated fully relativistically) and a valence set (treated scalar-relativistically). We highlight that the two AE codes used in this work do not adopt the same core/valence assignment for all crystals (Supplementary Section 5), yet they agree very well, illustrating that the core/valence assignment might not lead to ambiguities in the calculated results, provided that all other numerical parameters are chosen consistently.

Finally, we note that many additional code-specific parameters exist, such as the type and size of the basis set or the pseudopotential family. These choices are implemented in our automated common workflows[36] and can be selected using a new protocol defined for this work and named `verification-PBE-v1`. Specific details for each code are reported in Supplementary Section 5 for the AE codes, and in Supplementary Section 13 for all the other codes.

Before showing an example of the comparison of our reference AE dataset with nine computational approaches based on pseudopotentials, we discuss an additional recommendation. Our dataset was generated to provide reference EOS for each of the 960 structures. One might be tempted to reuse our dataset for different purposes. For instance, as the values of the minimum energy of the EOS curves are also available from the fits, one could imagine using them to compare total energies of various oxides of the same element X, estimating their relative stability and the corresponding formation energies. However, although this approach often results in sensible values, some notable cases lead to significantly off results, even by 1 eV per atom (Supplementary Section 18). The reason is that we designed our workflows and protocols for the EOS, to guarantee that simulation parameters are chosen consistently for all volumes of a given material, but this is not necessarily true among different materials. As an example, because oxides of the same element might have very different interatomic spacings, the choices of atomic radii (and the corresponding core/valence separation) for the AE codes might be different in different systems, which precludes direct comparison between total energies. From a more general point of view, one needs to be aware of the context in which data were produced and consider implications and limitations when using them for different applications.

Comparison with pseudopotential-based computational approaches

We compare our reference dataset with the results obtained with nine computational approaches based on pseudopotentials. Each approach is defined not only by the choice of the code, but also by the pseudopotentials used (and, where applicable, by the type of basis set). The exact versions of the codes and libraries, together with the other code-specific choices implemented in the `verification-PBE-v1` protocol, are detailed in Supplementary Section 13. The choices of computational approaches (for each code) have been selected by the workflow developers of each code, trying to identify converged parameters and limiting to choices commonly available to users. Here, we summarize briefly the meaning of the labels used for every computational approach, which, for the reasons discussed above, do not simply include the code name, but also indicate a few additional relevant parameters to better specify the details of the computational approach.

The two AE codes, FLEUR and WIEN2k, are labelled with their code name, followed by an indication of the basis set they use: `FLEUR@LAPW+LO` and `WIEN2k@(L)APW+lo+LO`, respectively (Supplementary Section 5). All other labels also include, at the end, the name of the pseudopotential library that was used. In particular: `ABINIT@PW|PseudoDojo-v0.5` indicates the ABINIT code, adopting a plane-wave (PW) basis set, using norm-conserving pseudopotentials from the PseudoDojo standard library version 0.5[65,66]; `BigDFT@DW|HGH-K(Valence)` indicates a (partial) set of structures with valence-only Hartwigsen–Goedecker–Hutter pseudopotentials[67] calculated with the BigDFT code, adopting a basis set of Daubechies wavelets (DW); `CASTEP@PW|C19MK2` indicates the CASTEP code using on-the-fly generated core-corrected ultrasoft pseudopotentials from the `C19` library with updated settings for the *f*-block elements; `CP2K/Quickstep@TZV2P|GTH` indicates the CP2K Quickstep code using Goedecker–Teter–Hutter pseudopotentials[68,69] and a molecularly optimized TZV2P-type basis set[70]; `GPAW@PW|PAW-v0.9.20000` indicates the GPAW[48,49] code used in its plane-wave mode using GPAW's PAW pseudopotentials included in the setup release 0.9.20000[71]; `Quantum ESPRESSO@PW|SSSP-prec-v1.3` indicates the Quantum ESPRESSO code using the Standard Solid-State Pseudopotentials (SSSP) library (PBE precision version 1.3)[72,73]; `SIESTA@AtOrOptDiamond|PseudoDojo-v0.4` indicates the SIESTA code using norm-conserving pseudopotentials from the PseudoDojo standard library version 0.4 in psml format[65,66,74] and localized basis sets in which the orbitals for each element are taken from a partial optimization, considering just the unary diamond structure for that element (therefore no optimization for the chemical environment of each material has been performed); `SIRIUS/CP2K@PW|SSSP-prec-v1.2` indicates the SIRIUS library code (run via its interface to CP2K) using the SSSP pseudopotential library (PBE precision version 1.2)[72]; `VASP@PW|GW-PAW54*` indicates the VASP code (v6.3) using the PAW GW PBE pseudopotentials released in the dataset `potpaw_PBE.54`, except for the lanthanides.

Using all these codes, we calculate the percentage error of V_0, B_0 and B_1, with respect to the reference average AE dataset, in the form of box-and-whisker plots (Fig. 4). We partition our results in three groups (considering rare earths and/or heavy elements separately), in order to highlight the non-uniform capabilities of the various computational approaches. Indeed, the narrow bands originating from the localized *f* electrons are very challenging to describe accurately with plain DFT[75]. Therefore, pseudopotentials for these elements are often not available, and thus several approaches cannot produce data for rare earths. Even when available, those pseudopotentials might be less tested and thus deliver a lower precision. By separating the results, we also enable a fairer comparison of approaches for the common set of elements (from H to Bi excluding the lanthanide elements from La to Lu).

Our results show that different numerical approaches have different precision; in general, the spreads of the parameters of pseudopotential approaches are significantly larger than those between our two AE codes. In addition, the results indicate that more work is required to obtain a high precision with approaches using localized basis sets (which, on the other hand, are typically faster and scale better with system size) with respect to those using a plane-wave basis set (in the present case, the approaches listed above using the ABINIT, CASTEP, GPAW, Quantum ESPRESSO, VASP and SIRIUS/CP2K codes). Indeed, while a plane-wave basis set can be tuned with a single numerical parameter (the energy cutoff), systematically improving localized basis sets (and the associated pseudopotentials) requires dedicated efforts (Box 4). Verification projects such as this will aid these efforts by providing appropriate benchmarks. In another example of verification, we discuss the agreement of different codes adopting the same computational approach, in particular with the same plane-wave basis set and the same pseudopotential library (Supplementary Section 14). In this case, the results show an agreement that is similar in precision to that between the two AE codes. We finally stress that the choice of the energy cutoff is specific to plane-wave codes, and in addition it depends on the pseudopotentials adopted. Therefore, in this Expert Recommendation, we do not provide reference values for the energy cutoffs (as we instead do for other code-agnostic parameters such as the *k*-point integration mesh or the energy smearing). Instead, the cutoffs are determined in the protocols of each code, together with all other code-specific numerical parameters; each code is responsible for verifying that the cutoff choice is converged for the goal of computing the EOS discussed here.

Periodic tables (similar to Fig. 2) for each code are provided in Supplementary Section 9, allowing for a closer inspection of the results resolved per chemical element and crystal-structure type. These tables also show that using a larger crystal-structure set (960 systems here) than the set of 71 of ref. 29 helps in highlighting possible shortcomings of pseudopotentials (Supplementary Section 15). The results of each code are also available in ref. 62, and can be visually displayed and compared directly online on the Materials Cloud[19] at https://acwf-verification.materialscloud.org.

Pseudopotential improvement

Curated datasets can drive efforts to improve pseudopotentials, ultimately delivering more precise computational approaches. To illustrate this, we briefly summarize examples of pseudopotential enhancements that we performed to improve the comparison with our AE results (Supplementary Section 16) and used in the generation of the data of Fig. 4.

The results for `ABINIT@PW|PseudoDojo-v0.5` for elements around the 4*f* block (from Te to Ba, and from Tl to Rn) were not giving ideal agreement when using available pseudopotentials from PseudoDojo (version 0.4). In almost all cases, we found that the accuracy of the pseudopotentials is significantly improved by including a projector for the unbound *f* state, at the expense of an increase of the computational cost when applying the non-local part of the Hamiltonian V_{nl} (this can, however, be mitigated by the use of Legendre polynomials). Without this projector, the local part of the pseudopotential cannot reproduce the all-electron scattering properties of the *f* angular momentum (Supplementary Section 16.1). This led to the creation of a new PseudoDojo table (version 0.5), used here.

For CASTEP@PW | C19MK2, starting from the on-the-fly pseudopotential generation settings for the built-in C19 library, pseudopotentials for the lanthanide and actinide elements were improved by systematically changing the core radii, including additional projectors and adding fractional occupations of states that are empty in the reference atomic configurations. Although making these changes did result in improvements, we note that no iterative optimization has been carried out to fit to the AE results.

For Quantum ESPRESSO@PW | SSSP-prec-v1.3, the pseudopotentials of SSSP PBE Precision version 1.1.2[76] have been updated for elements Na, Cu, Cs, Cd, Ba, As, Te, I, Hg, Ne, Ar, Kr, Xe and Rn; these new pseudopotentials have been released in the new SSSP PBE Precision 1.2[77]. The new pseudopotentials have been selected by re-verifying the precision of pseudopotentials from various external libraries against the AE reference dataset discussed here, and replacing those displaying significant discrepancies with pseudopotentials from other libraries that displayed a better agreement (lower ε and v). Moreover, in SSSP PBE Precision version 1.3[73] (used here) new pseudopotentials have been included for actinides (Th–Lr) from ref. 78, as well as for Ac, At, Ra and Fr from PSlibrary[79].

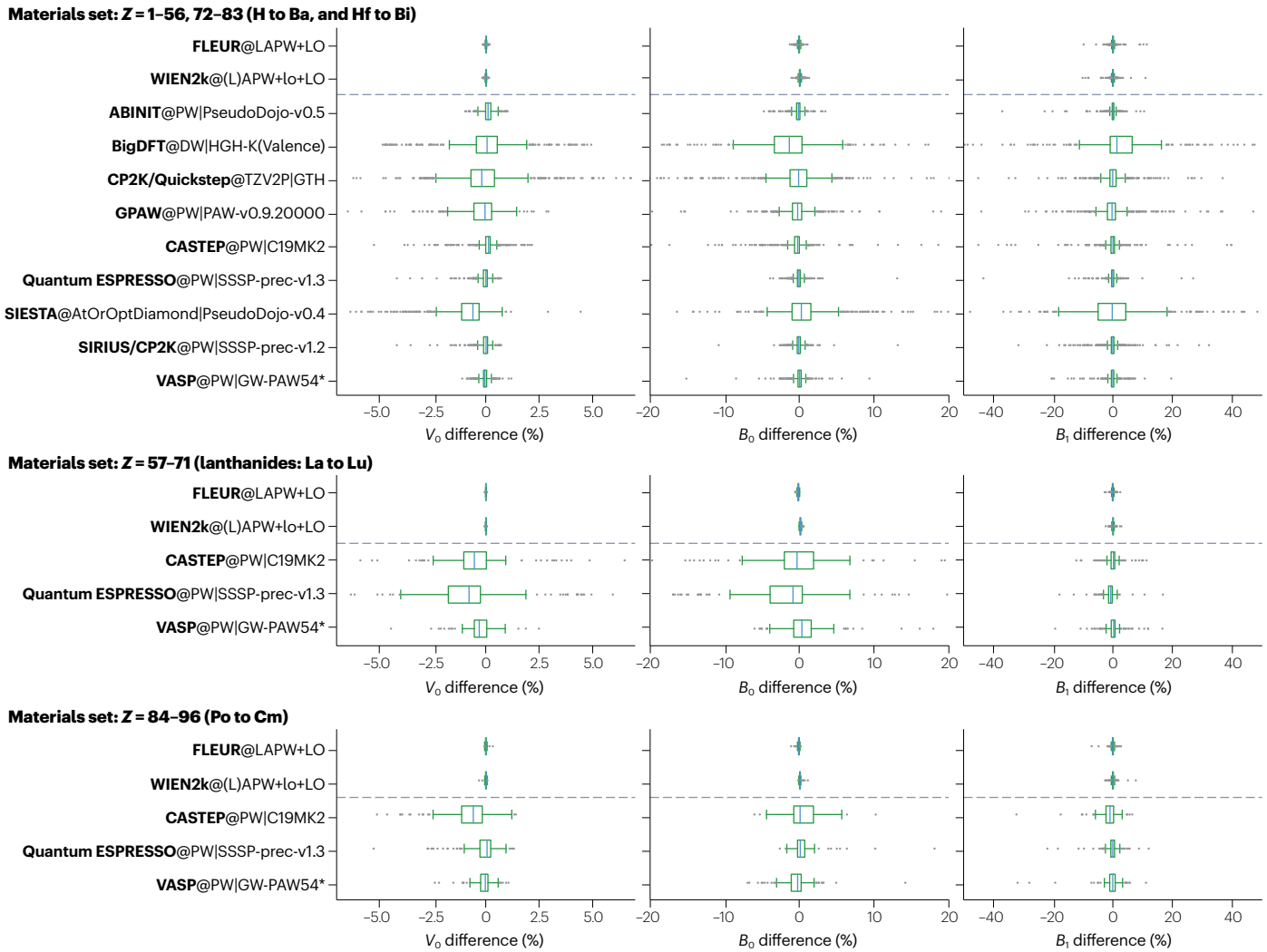

Fig. 4 | Comparing the results of different computational approaches. The box-and-whisker plots compare the discrepancy of the equilibrium volume V_0, the bulk modulus B_0, and its derivative with respect to the pressure, B_1, for each computational approach involved in this work with respect to the average all-electron (AE) reference dataset. The two AE codes are also reported at the top of the plot above the dashed line, for comparison. In the box-and-whisker plots, the central blue line represents the median and the box extends between the first quartile $Q1$ and the third quartile $Q3$. The 'whiskers' extend between the first point greater than $Q1 - 1.5$IQR and the last point smaller than $Q3 + 1.5$IQR (where IQR is the inter-quartile range $Q3 - Q1$). Outliers beyond the whiskers are represented as grey points. Note that some of the outliers are outside the visible axis range in order to aid comparison between codes on the same axis range. Each row corresponds to a different subset of materials, where only the computational approaches that could compute those materials are included (as not all approaches include pseudopotentials for rare earths). Specifically, the top row includes all materials from H to Bi excluding the lanthanide elements from La to Lu (68 elements in total). For this set, 295 crystals are missing for BigDFT@DW | HGH-K(Valence), Na (FCC) is missing for CP2K/Quickstep@TZV2P | GTH, Hg (FCC) and RbO_3 are missing for SIESTA@AtOrOptDiamond | PseudoDojo-v0.4, and all 10 crystals containing Tc are missing for GPAW@PW | PAW-v0.9.20000. The central row reports the results for lanthanides only (from La to Lu), and the bottom row for all materials from Po to Cm (heavy elements, including actinides up to Cm).

Box 4

Recommendations to extend the verification effort

- Extend the current study to more computational approaches (including codes, basis or pseudopotential sets), adopting the same reference crystal-structure set presented here.
- Investigate whether there are advantages in using a different crystal-structure set from the one presented here. For example, lower-symmetry environments (such as low-symmetry structures, vacancies or surfaces) might highlight further differences between DFT implementations; or it may be worth considering only compounds for which experimental data are available, for studies focusing instead on validation against experiments.
- Extend the current study to more properties (such as forces, phonons, Kohn–Sham band structures and formation energies). Choose properties (and materials) that maximize the number of codes that can compute them (in this case for the EOS, only single-point DFT simulations are required; forces and stresses were not used).
- Investigate the generality of optimal protocols and develop new ones for each property being computed, generalizing how to select a consistent set of parameters for multiple runs. For instance, for the EOS it is important to use the same *k*-point integration mesh at all volumes, but for a formation energy one wants a mesh that has reached a threshold accuracy for each component. In addition, use the same core/valence assignment, core radii and any other approach-specific precaution needed to compare total energies of different crystals.
- Create additional curated sets needed to generate improved pseudopotentials. For example, extend to fully relativistic simulations and consider other exchange–correlation functionals in addition to PBE, such as the local-density approximation and PBEsol, but also a selection of hybrids and meta-generalized-gradient approximations (meta-GGAs) for instance. Provide also curated sets needed to generate core-hole pseudopotentials for the simulation of core-level spectroscopies.
- Beside targeting improved pseudopotentials, develop dedicated efforts to optimize localized basis sets when these cannot be systematically improved by just tuning one or a few numerical parameters.
- Develop new protocols aiming at 'good enough' data: not only targeting ultimate numerical precision (needed for verification), but also optimizing the computational cost for a target accuracy.
- Disseminate these protocols to the broad simulation community to optimize energy and CPU time and to expand the computational feasibility of DFT computational approaches in high-throughput studies or for expensive post-DFT methods (such as many-body perturbation theory).
- For new verification protocols, define metrics (such as ε and v discussed here for the EOS) that depend on physically measurable quantities. Using such metrics, which condense into a single quantity the precision of computational approaches to a property of interest, one can easily define precision thresholds, compare approaches quantitatively, and evaluate the uniformity of results in a dataset. If fitting procedures are needed, assess the robustness of the chosen algorithms and estimate the uncertainty on the fitted parameters, using the results to define error bars.

For `VASP@PW|GW-PAW54*`, the latest available PAW potential set (version 5.4) was improved by reducing by about 20% the core radii for lanthanides (other than La, Ce and Lu). Furthermore, placing two electrons in the 6*s* shell, half an electron in the 5*d* shell and the rest in the *f* shell led to the most balanced description. For Tm, Er and Yb, three *f* projectors were required to accurately describe the *f* scattering properties. The optimization was continued until very accurate scattering properties were obtained and agreement with very small core potentials was excellent, in turn resulting in a significant improvement of the agreement with the AE reference dataset.

Outlook

This work constitutes a next step in a grand scheme of actions aiming at controlling the numerical aspects of electronic-structure calculations, where the diversity of computational approaches and codes provides an opportunity for pairwise verification. Compared with earlier work[29,30], we define here more discriminative metrics (crystals for which two approaches would agree according to Δ might agree less according to ε or v; Supplementary Fig. 8.1a,b) and consider many more crystals, leading to more stringent testing. Although major conclusions based on previous work remain valid (Supplementary Sections 15, 17), the dataset presented here — together with the clear set of recommendations on how to reuse the data — provides a more refined and valuable reference for verification, uncertainty quantification and pseudopotential optimization. Additionally, by formulating recommendations on how to perform further validation studies, and by providing and sharing universal common workflow interfaces (based on the AiiDA workflow infrastructure) to reproduce our calculations and perform new ones, we aid the community in taking new steps towards a better control of the uncertainty quantification in electronic-structure calculations (Box 4).

We recommend the creation of similar datasets for other commonly used exchange–correlation functionals (such as local-density approximation and PBEsol, but possibly also a selection of hybrid and meta-GGA functionals), as well as for fully relativistic simulations. For these studies, we recommend using the same initial set of crystal structures discussed here, possibly only adapting the central point of the volume interval $[V_m, V_M]$ if the equilibrium volume V_0 for the functional does not lie roughly in the middle of the interval. Indeed, the set is fairly complete and systematic, and using the same structures aids the comparison between different computational approaches and approximations. In addition, we recommend to test and verify codes also for magnetic materials. Once such datasets are available, efforts to further improve pseudopotentials (and basis sets) should be initiated or continued, with the aim of making the results easily available to the broad simulation community. One useful outcome could be, for instance, the generation of new, reliable, fully relativistic pseudopotential datasets for rare earths (especially of the norm-conserving type, often required by many codes computing advanced materials

properties). Another relevant example, also involving the generation of additional bespoke AE reference datasets, is the generation of pseudopotentials with a hole in the core, needed to predict the outcome of X-ray photoelectron spectroscopy or X-ray absorption spectroscopy experiments.

As a note, we highlight here that some of our structures are unrealistic. When generating a new pseudopotential for a given chemical element, one might want to accept a compromise and not precisely reproduce the EOS of all 10 unaries and oxides, in order to obtain a computationally cheaper pseudopotential (for example, with fewer projectors, more electrons in the core, or requiring a smaller energy cutoff), as long as the results are precise enough for the intended applications.

Other properties beyond the EOS are relevant to characterize materials and might benefit from tailored verification efforts; these include formation energies, electronic band structures and phonon frequencies, and the simulation protocols might be significantly different for each property. We therefore recommend that these protocols are well designed, documented and discussed, together with their limit of applicability. In particular, especially if limiting to a scalar-relativistic approach as we did here, we recommend to further investigate the relevance of the choice of which electrons are included in the core or in the valence, as this can be of higher relevance than for the EOS (for formation energies, see Supplementary Section 18). Moreover, new metrics should be designed to quantitatively compare results, ideally directly dependent on physically measurable quantities. Error propagation through any fitting procedure or data analysis should be carefully assessed to be able to define appropriate error bars (Supplementary Section 3).

Finally, we emphasize that whereas the goal here was ultimate precision in order to provide a reference dataset and obtain the best agreement possible between computational approaches, in real simulations one also needs to optimize the computational cost for a target accuracy, to obtain 'good enough' data for that scientific purpose. This is especially true for high-throughput runs or when the DFT simulations are the first step of more expensive post-DFT methods. We thus encourage developing protocols to automatically define or select optimally converged parameters that at the same time minimize energy and CPU time, and then disseminating these to the whole community, so that they become easily accessible to a broad range of users.

Code availability

The source code of the common workflows is released under the MIT open-source licence and is made available on GitHub (https://github.com/aiidateam/aiida-common-workflows). It is also distributed as an installable package through the Python Package Index (https://pypi.org/project/aiida-common-workflows). The source code of the scripts to generate the plots is released under the MIT open-source licence and is made available on GitHub (https://github.com/aiidateam/acwf-verification-scripts). All codes to generate the figures of this paper are available in the data entry of ref. 62.

Data availability

The data and the scripts used to create all the images in this work are available on the Materials Cloud Archive[62]. Moreover, the data are accessible via an interactive website, https://acwf-verification.materials-cloud.org, that offers various analysis and visualization possibilities. Note that the data include the entire AiiDA provenance graph of each workflow execution presented in the main text (therefore including all input files and output files of all simulations, as well as their logical relationship, in AiiDA format), as well as the curated data extracted from that database to produce the images.

Published online: 14 November 2023

References

1. Alberi, K. et al. The 2019 materials by design roadmap. *J. Phys. D* **52**, 013001 (2018).
2. Marzari, N., Ferretti, A. & Wolverton, C. Electronic-structure methods for materials design. *Nat. Mater.* **20**, 736–749 (2021).
3. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* **111**, 218–230 (2016).
4. Huber, S. P. et al. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* **7**, 300 (2020).
5. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source Python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
6. Jain, A. et al. Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput. Pract. Exp.* **27**, 5037–5059 (2015).
7. Mathew, K. et al. Atomate: a high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **139**, 140–152 (2017).
8. Bahn, S. R. & Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66 (2002).
9. Hjorth Larsen, A. et al. The atomic simulation environment — a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
10. Curtarolo, S. et al. Aflow: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
11. Janssen, J. et al. pyiron: an integrated development environment for computational materials science. *Comput. Mater. Sci.* **163**, 24 – 36 (2019).
12. Armiento, R. Database-driven high-throughput calculations and machine learning models for materials design. In *Machine Learning Meets Quantum Physics* (eds Schütt, K. et al.) Ch. 17, 377–395 (Springer, 2020).
13. Gonze, X. et al. The ABINIT project: impact, environment and recent developments. *Comput. Phys. Commun.* **248**, 107042 (2020).
14. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
15. Landis, D. D. et al. The Computational Materials Repository. *Comput. Sci. Eng.* **14**, 51–57 (2012).
16. Kirklin, S. et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *NPJ Comput. Mater.* **1**, 15010 (2015).
17. Merkys, A. et al. A posteriori metadata from automated provenance tracking: integration of AiiDA and TCOD. *J. Cheminform.* **9**, 56 (2017).
18. Stevanović, V., Lany, S., Zhang, X. & Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: fitted elemental-phase reference energies. *Phys. Rev. B* **85**, 115104 (2012).
19. Talirz, L. et al. Materials Cloud, a platform for open computational science. *Sci. Data* **7**, 299 (2020).
20. Draxl, C. & Scheffler, M. NOMAD: the FAIR concept for big data-driven materials science. *MRS Bull.* **43**, 676–682 (2018).
21. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
22. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9 (2016).
23. Andersen, C. W. et al. OPTIMADE, an API for exchanging materials data. *Sci. Data* **8**, 217 (2021).
24. IEEE Standard for System, Software, and Hardware Verification and Validation. IEEE Std 1012-2016 (Revision of IEEE Std 1012-2012/ Incorporates IEEE Std 1012-2016/Cor1-2017) 1–260, https://doi.org/10.1109/IEEESTD.2017.8055462 (2017).
25. Wang, A. et al. A framework for quantifying uncertainty in DFT energy corrections. *Sci. Rep.* **11**, 15496 (2021).
26. Carbogno, C. et al. Numerical quality control for DFT-based materials databases. *NPJ Comput. Mater.* **8**, 69 (2022).
27. Poncé, S. et al. Verification of first-principles codes: comparison of total energies, phonon frequencies, electron–phonon coupling and zero-point motion correction to the gap between ABINIT and QE/Yambo. *Comput. Mater. Sci.* **83**, 341–348 (2014).
28. Pople, J. Nobel Lecture: Quantum Chemical Models. https://www.nobelprize.org/prizes/chemistry/1998/pople/lecture/ (1998).
29. Lejaeghere, K. et al. Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000 (2016).
30. Cottenier, S., Delta project — archive of old website. Materials Cloud Archive 2023.133, https://doi.org/10.24435/materialscloud:5e-mv (2023).
31. The FLEUR project. https://www.flapw.de/.
32. Wortmann, D. et al. Fleur. https://doi.org/10.5281/zenodo.7576163 (2023).
33. Blaha, P. et al. WIEN2k: an augmented plane wave plus local orbitals program for calculating crystal properties. http://www.wien2k.at/.
34. Blaha, P. et al. WIEN2k: an apw+lo program for calculating the properties of solids. *J. Chem. Phys.* **152**, 074101 (2020).

35. Uhrin, M., Huber, S. P., Yu, J., Marzari, N. & Pizzi, G. Workflows in AiiDA: engineering a high-throughput, event-based engine for robust and modular computational workflows. *Comput. Mater. Sci.* **187**, 110086 (2021).
36. Huber, S. P. et al. Common workflows for computing material properties using different quantum engines. *NPJ Comput. Mater.* **7**, 136 (2021).
37. Huber, S. P. et al. AiiDA common workflows (ACWF) package, versions 1.0.1 and above (2023); https://github.com/aiidateam/aiida-common-workflows.
38. Lejaeghere, K., Speybroeck, V. V., Oost, G. V. & Cottenier, S. Error estimates for solid-state density-functional theory predictions: an overview by means of the ground-state elemental crystals. *Crit. Rev. Solid State Mater. Sci.* **39**, 1–24 (2013).
39. Grosso, G. & Pastori Parravicini, G. *Solid State Physics* 2nd edn (Academic, 2013).
40. Martin, R. M.*Electronic Structure: Basic Theory and Practical Methods* 2nd edn (Cambridge Univ. Press, 2013).
41. Cohen, M. L. & Louie, S. G. *Fundamentals of Condensed Matter Physics* (Cambridge Univ. Press, 2016).
42. Gonze, X. et al. Recent developments in the ABINIT software package. *Comput. Phys. Commun.* **205**, 106–131 (2016).
43. Romero, A. H. et al. ABINIT: overview and focus on selected capabilities. *J. Chem. Phys.* **152**, 124102 (2020).
44. Ratcliff, L. E. et al. Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *J. Chem. Phys.* **152**, 194110 (2020).
45. Clark, S. J. et al. First principles methods using CASTEP. *Z. Kristallogr. Cryst. Mater.* **220**, 567–570 (2005).
46. The CP2K simulation package. https://www.cp2k.org.
47. Kühne, T. D. et al. CP2K: an electronic structure and molecular dynamics software package — Quickstep: efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).
48. Mortensen, J. J., Hansen, L. B. & Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **71**, 035109 (2005).
49. Enkovaara, J. et al. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter* **22**, 253202 (2010).
50. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
51. Giannozzi, P. et al. Advanced capabilities for materials modelling with QUANTUM ESPRESSO. *J. Phys. Condens. Matter* **29**, 465901 (2017).
52. Soler, J. M. et al. The SIESTA method for ab initio order-*N* materials simulation. *J. Phys. Condens. Matter* **14**, 2745–2779 (2002).
53. García, A. et al. SIESTA: recent developments and applications. *J. Chem. Phys.* **152**, 204108 (2020).
54. The SIRIUS domain-specific library for electronic-structure calculations. https://github.com/electronic-structure/SIRIUS.
55. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
56. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
57. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
58. Vanpoucke, D. E. P., Bultinck, P. & Van Driessche, I. Extending Hirshfeld-I to bulk and periodic materials. *J. Comput. Chem.* **34**, 405–417 (2013).
59. Vanpoucke, D. E. P., Van Driessche, I. & Bultinck, P. Reply to "Comment on 'Extending Hirshfeld-I to bulk and periodic materials'". *J. Comput. Chem.* **34**, 422–427 (2013).
60. Bultinck, P., Van Alsenoy, C., Ayers, P. W. & Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **126**, 144111 (2007).
61. Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71**, 809–824 (1947).
62. Bosoni, E. et al. How to verify the precision of density-functional-theory implementations via reproducible and universal workflows. Materials Cloud Archive 2023.81, https://doi.org/10.24435/materialscloud:s4-3h (2023).
63. Jollet, F., Torrent, M. & Holzwarth, N. Generation of projector augmented-wave atomic data: a 71 element validated table in the XML format. *Comput. Phys. Commun.* **185**, 1246–1254 (2014).
64. Glantz, S. A., Slinker, B. K. & Neilands, T. B. in *Primer of Applied Regression and Analysis of Variance* 3rd edition, 261–298 (McGraw-Hill Education, 2017).
65. van Setten, M. et al. The PseudoDojo: training and grading a 85 element optimized norm-conserving pseudopotential table. *Comput. Phys. Commun.* **226**, 39–54 (2018).
66. The PseudoDojo website. http://www.pseudo-dojo.org/.
67. Hartwigsen, C., Goedecker, S. & Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **58**, 3641–3662 (1998).
68. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703 (1996).
69. Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **114**, 145–152 (2005).
70. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).
71. GPAW atomic PAW setups. https://wiki.fysik.dtu.dk/gpaw/setups/setups.html#atomic-paw-setups.
72. Prandini, G., Marrazzo, A., Castelli, I. E., Mounet, N. & Marzari, N. Precision and efficiency in solid-state pseudopotential calculations. *NPJ Comput. Mater.* **4**, 72 (2018).
73. Prandini, G. et al. A standard solid state pseudopotentials (SSSP) library optimized for precision and efficiency. Materials Cloud Archive 2023.65 Version v11, https://doi.org/10.24435/materialscloud:f3-ym (2023).
74. García, A., Verstraete, M. J., Pouillon, Y. N. & Junquera, J. The PSML format and library for norm-conserving pseudopotential data curation and interoperability. *Comput. Phys. Commun.* **227**, 51–71 (2018).
75. Topsakal, M. & Wentzcovitch, R. Accurate projected augmented wave (PAW) datasets for rare-earth elements (Re = La–Lu). *Comput. Mater. Sci.* **95**, 263–270 (2014).
76. Prandini, G. et al. A standard solid state pseudopotentials (SSSP) library optimized for precision and efficiency. Materials Cloud Archive 2021.76 Version v7, https://doi.org/10.24435/materialscloud:rz-77 (2021).
77. Prandini, G. et al. A standard solid state pseudopotentials (SSSP) library optimized for precision and efficiency. Materials Cloud Archive 2022.159 Version v8, https://doi.org/10.24435/materialscloud:3v-xt (2022).
78. Sachs, M. et al. DFT-guided crystal structure redetermination and lattice dynamics of the intermetallic actinoid compound UIr. *Inorg. Chem.* **60**, 16686–16699 (2021).
79. Dal Corso, A. Pseudopotentials periodic table: from H to Pu. *Comput. Mater. Sci.* **95**, 337–350 (2014).
80. Thörnig, P. JURECA: data centric and booster modules implementing the modular supercomputing architecture at Jülich supercomputing centre. *JLSRF* **7**, 182 (2021).
81. Marzari, N., Vanderbilt, D., De Vita, A. & Payne, M. C. Thermal contraction and disordering of the Al(110) surface. *Phys. Rev. Lett.* **82**, 3296–3299 (1999).
82. dos Santos, F. J. & Marzari, N. Fermi energy determination for advanced smearing techniques. *Phys. Rev. B* **107**, 195122 (2023).
83. Gillan, M. J. Calculation of the vacancy formation energy in aluminium. *J. Phys. Condens. Matter* **1**, 689 (1989).

Acknowledgements

This work was inspired and is supported in part by the European Union (EU)'s Horizon 2020 research and innovation programme under grant agreements no. 676598 and no. 824143 (European MaX Centre of Excellence 'Materials Design at the Exascale') and by NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (SNSF, grant no. 205602). For the purpose of Open Access, a CC BY public copyright licence is applied to any Author Accepted Manuscript (AAM) version arising from this submission. We thank F. J. dos Santos for discussions on the analysis of the smearing types and *k*-point convergence, and X. Gonze, M. Torrent and F. Jollet for discussions on PAW pseudopotentials. M.F. and N.M. acknowledge the contribution of S. Shankar in early discussions about the use of prototype oxides as general platform to explore the transferability of pseudopotentials. Work at ICMAB (E.B., A.G., V.D.) is supported by the Severo Ochoa Centers of Excellence Program (MCIN CEX2019-000917-S), by grant PGC2018-096955-B-C44 of MCIN/AEI/10.13039/501100011033, 'ERDF A way of making Europe', and by GenCat 2017SGR1506. We also thank the Barcelona Supercomputer Center (BSC) for computational resources. V.D. acknowledges support from DOC-FAM, EU Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 754397. O.R. acknowledges travel support from WIEN2k (Technical University of Vienna). The Jülich team (S.B., J.B., H.J., G.M., D.W.) acknowledge support by the Joint Lab Virtual Materials Design of the Forschungszentrum Jülich, the Helmholtz Platform for Research Software Engineering — Preparatory Study, the Joint Virtual Laboratory AI, Data Analytics and Scalable Simulation of the Forschungszentrum Jülich and the French Alternative Energies and Atomic Energy Commission, and the computing time granted through JARA on the supercomputers JURECA[80] at Forschungszentrum Jülich and CLAIX at RWTH Aachen University. H.M. and T.D.K. (Univ. Paderborn) acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS supercomputer JUWELS at Jülich Supercomputing Centre. S.P. and G.-M.R. (Univ. catholique de Louvain) acknowledge support from the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS). Computational resources have been provided by the PRACE-21 resources MareNostrum at the BSC-CNS and by the Consortium des Équipements de Calcul Intensif, funded by the F.R.S.-FNRS under grant no. 2.5020.11 and by the Walloon Region as well as computational resources awarded on the Belgian share of the EuroHPC LUMI supercomputer. G.Ka. and S.V. received funding from the VILLUM Centre for the Science of Sustainable Fuels and Chemicals (9455) from VILLUM FONDEN. Computational resources were provided by the Niflheim supercomputing cluster at the Technical University of Denmark. They also thank J. J. Mortensen and A. H. Larsen for discussions on optimizing the workflow for the GPAW code. S.C. acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. The computational resources and services used at Ghent University were provided by the Vienna Scientific Cluster (VSC; Flemish Supercomputer Center), funded by the Research Foundation Flanders and the Flemish Government department EWI. M.W. acknowledges computational resources provided by the VSC. This research was funded in part by the Austrian Science Fund (FWF) [P 32711]. E.F.L. acknowledges resources provided by Sigma2 — the National Infrastructure for High Performance Computing and Data Storage in Norway, and support from the Norwegian Research Infrastructure Services. B.Z. is grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1 and EP/T022213/1), and acknowledges the use of the UCL Myriad and Kathleen High Performance Computing Facility (Myriad@UCL, Kathleen@UCL), and associated support services, in the completion of this work. N.M., G.Pi. and A.G. acknowledge support from the EU Horizon 2020 research and innovation programme under grant agreement no. 957189 (BIG-MAP), also part of the BATTERY 2030+ initiative under grant agreement no. 957213. G.P., J.Y. and G.-M.R. acknowledge support by the SNSF and by the F.R.S.-FNRS through the 'FISH4DIET' Project (SNSF grant 200021E_206190 and F.R.S.-FNRS

grant T.0179.22). G.P. acknowledges support by the Open Research Data Program of the ETH Board, under the Establish project 'PREMISE'. J.Y. acknowledges support from the EU Horizon 2020 research and innovation programme under grant agreement no. 760173 (MARKETPLACE).

Author contributions

M.F. and N.M. contributed the idea of using prototype oxides to test pseudopotentials across different coordinations and chemistries. S.C. and K.L. performed an initial assessment of this idea and analysed, together with M.C., the first exploratory datasets. M.S. was responsible for job, queue and data management in the first exploratory phase. E.B. and G.Pi. contributed the idea of using the AiiDA and `aiida-common-workflows` infrastructure to carry on the thousands of DFT simulations required by the project. E.B. and G.Pi. coordinated the whole project. P.B., G.M. and O.R. performed the iterative refinement of the input parameters for the AE calculations that ultimately resulted in the generation of the central volumes of our dataset. O.R. proposed the ε metric. N.M. proposed the v metric. S.P. raised the issue of the smearing selection that ultimately led to the decision of a fixed k-point integration mesh and smearing broadening. K.E. contributed to the data analysis and the conversion of the data into a dynamic website. D.E.P.V. contributed the Hirshfeld-I calculated charges and their analysis. E.B., M.W. and G.Pi. performed the analysis of the error propagation in the fit and the estimation of the parameters of the v metric. E.B., S.C., O.R. and G.Pi. analysed in detail the dependency and sensitivity of the metrics Δ, ε and v. A.Z. and S.P. developed the ABINIT implementation of the common workflow which relies on the `aiida-abinit` plugin developed and maintained by A.Z., G.Pe. and S.P. M.G. created new pseudopotentials used for ABINIT and improved the parameter profile. A.Z. and S.P. performed the ABINIT calculations, including verification tests. The work on ABINIT was supervised by G.-M.R. and S.P. L.B., A.D. and L.G. contributed to the BigDFT-related parts of the work. A.D. developed the BigDFT implementation of the common workflows. L.B. and A.D. generated the BigDFT results under the supervision of L.G. B.Z. developed the CASTEP implementation of the common workflow, which relies on the `aiida-castep` plugin also maintained by B.Z., and performed all CASTEP simulations. C.J.P. created new on-the-fly generated pseudopotentials using the verification tests performed by B.Z. M.K., T.D.K., H.M., T.M.A.M. and A.V.Y. contributed to all CP2K-related parts of this work. A.V.Y. implemented the workflows and performed preliminary calculations. The workflows rely on the `aiida-cp2k` plugin developed by A.V.Y., T.M.A.M. and others. M.K. performed preliminary calculations, created new pseudopotentials and contributed to the design of the protocol, T.D.K. contributed to the CP2K setup, discussed the results and supervised the calculations, H.M. conducted all AiiDA calculations and analysed the results, T.M.A.M. provided implementations of CP2K input and output parsers. S.B., J.B., H.J., G.M. and D.W. contributed the FLEUR-related parts of this work. G.M. developed the parameter profile and performed the calculations, and is the contact person for the FLEUR contributions. J.B. and H.J. adapted and extended the `AiiDA-FLEUR` plugin and the related parts of the AiiDA common-workflows package. S.B. and D.W. contributed to the analysis and discussion of the FLEUR results. G.Ka. and S.V. contributed to the GPAW-related parts of the work. S.V. developed the GPAW implementation of the common workflows, which relies on the `aiida-ase` plugin, and ran the calculations. G.Ka. and S.V. analysed the GPAW calculations. M.B., S.P.H., N.M., J.Y. and G.Pi. contributed to all Quantum ESPRESSO-related parts of this work. M.B. and S.P.H. developed the Quantum ESPRESSO implementation of the common workflow, which relies on the `aiida-quantumespresso` plugin developed and maintained by M.B., S.P.H., G.Pi. and others. J.Y. generated and tested new pseudopotentials used for Quantum ESPRESSO. The work on Quantum ESPRESSO was supervised by G.Pi. and N.M. H.M. and T.D.K. contributed to all SIRIUS/CP2K-related parts of this work. The workflows rely on the `aiida-cp2k` plugin developed by A.V.Y., T.M.A.M., and others. H.M. conducted all AiiDA calculations and analysed the results. T.D.K. contributed to the SIRIUS/CP2K setup, discussed the results and supervised the calculations. E.B., V.D. and A.G. contributed to the SIESTA-related parts of the work. E.B. developed the SIESTA implementation of the common workflows, which relies on the `aiida-siesta` plugin developed by E.B., A.G., V.D. and others. E.B. generated the SIESTA results in collaboration with A.G. M.W., M.M. and E.F.-L. performed the execution and analysis of the VASP-related workflows used to generate the data for this work. G.Kr. generated updated potentials for the lanthanides. E.F.-L. maintains the VASP implementation of the common-workflows project and the `aiida-vasp` plugin (developed by a community of contributors; see full contributor list in the plugin documentation) which is used to execute the VASP calculations. P.B., G.K.H.M., O.R. and T.R. contributed the WIEN2k-related parts of this work. T.R. performed preliminary calculations, P.B. created the setup of the WIEN2k calculations and supervised and analysed the results, G.K.H.M. contributed the conversion of AiiDA structures to a WIEN2k struct file, and O.R. developed the `aiida-wien2k` plugin and performed all AiiDA-WIEN2k calculations. E.B., M.F. and G.Pi. wrote the first version of the manuscript, and all authors contributed to the editing and revision of the manuscript.

Competing interests

G.Pe. and G.-M.R. are shareholders and directors of Matgenix SRL. G.Kr. is a shareholder of the VASP Software GmbH, and M.W. and M.M. are part-time employees of the VASP Software GmbH. C.J.P. is an author of the CASTEP code and receives income from its commercial sales.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42254-023-00655-3.

Peer review information *Nature Reviews Physics* thanks Vikram Gavini, Ann Mattson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

[1]Institut de Ciència de Materials de Barcelona, ICMAB-CSIC, Bellaterra, Spain. [2]Université Grenoble-Alpes, CEA, IRIG-MEM-L_Sim, Grenoble, France. [3]Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [4]Institute for Materials Chemistry, Technical University of Vienna, Vienna, Austria. [5]Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany. [6]Institute for Advanced Simulation, Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich, Jülich, Germany. [7]Department of Electromechanical, Systems and Metal Engineering, Ghent University, Ghent, Belgium. [8]Center for Molecular Modeling (CMM), Ghent University, Ghent, Belgium. [9]Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan. [10]Norwegian EuroHPC Competence Center, Sigma2 AS, Trondheim, Norway. [11]SINTEF Industry, Materials Physics, Oslo, Norway. [12]Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA. [13]Institut de la Matière Condensée et des Nanosciences (IMCN), Université catholique de Louvain, Louvain-la-Neuve, Belgium. [14]National Centre of Competence in Research (NCCR) Catalysis, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [15]Center for Catalysis Theory (Cattheory), Department of Physics, Technical University of Denmark (DTU), Kongens Lyngby, Denmark. [16]Laboratory for Materials Simulations (LMS), Paul Scherrer Institut (PSI), Villigen PSI, Switzerland. [17]University of Vienna, Faculty of Physics and Center for Computational Materials Science, Vienna, Austria. [18]VASP Software GmbH, Vienna, Austria. [19]Center for Advanced Systems Understanding (CASUS) and Helmholtz-Zentrum Dresden-Rossendorf, Görlitz, Germany. [20]Paderborn Center for Parallel Computing (PC2) and Center for Sustainable Systems Design, University of Paderborn, Paderborn, Germany. [21]OCAS NV/ArcelorMittal Global R&D Gent, Zelzate, Belgium. [22]Dynamics of Condensed Matter, Chair of Theoretical Chemistry, University of Paderborn, Paderborn, Germany. [23]HPE HPC EMEA Research Lab, Basel, Switzerland. [24]Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, United Kingdom. [25]Advanced Institute for Materials Research, Tohoku University, Sendai, Japan. [26]Department of Materials Science and Engineering, McMaster University, Hamilton, Ontario, Canada. [27]Chair of Physical Chemistry, Montanuniversitaet Leoben, Leoben, Austria. [28]ePotentia, Deurne, Antwerp, Belgium. [29]Institute for Materials Research (IMO-IMOMEC), UHasselt — Hasselt University, Hasselt, Belgium. [30]Swiss Federal Laboratories for Materials Science and Technology (Empa), nanotech@surfaces laboratory, Dübendorf, Switzerland. [31]Department of Chemistry, University College London, London, UK. [32]The Faraday Institution, Didcot, UK.