Fitting Elephants in the Density Functionals Zoo: Statistical Criteria for the Evaluation of DFT methods as a Suitable Replacement for Counting Parameters

Roberto Peverati[1]

[1]IJQC Special Issue

June 12, 2020

Abstract

Counting parameters has become customary in the density functional theory community as a way to infer the transferability of popular approximations to the exchange–correlation functionals. Recent work in data science, however, has demonstrated that the number of parameters of a fitted model is not related to the complexity of the model itself, nor to its eventual overfitting. Using similar arguments, we show here that it is possible to represent every modern exchange–correlation functional approximation using just one single parameter. This procedure proves the futility of the number of parameters as a measure of transferability. To counteract this shortcoming, we introduce and analyze the performance of three statistical criteria for the evaluation of the transferability of exchange–correlation functionals. The three criteria are called Akaike information criterion (AIC), Vapnik–Chervonenkis criterion (VCC), and cross-validation criterion (CVC) and are used in a preliminary assessment to rank 60 exchange–correlation functional approximations using the ASCDB database of chemical data.

1. Introduction

The success of density functional theory (DFT) as the method of choice for the calculation of the electronic structure of molecules is undeniable, and is intertwined with the development of improved approximations for the description of the exchange–correlation functional (*xc* functional, or just simply functional) [1]. Such success is reflected by an indiscriminate proliferation of approximations, calling for a rugged safari across the "zoo of functionals" [2,3,4] to select an appropriate one [5].

Two philosophies are at odds in the world of functional development: the first one originated mainly within the chemistry community from the pioneering work of Becke [6,7], which took the approach of using flexible parametrized mathematical forms that are fitted to chemical data, exact constraints, or a mix of both. The second philosophy originated primarily within the physics community from the ground-breaking work of Perdew [8,9,10], which expanded the knowledge and application of exact conditions and advocated for DFT to remain a purely *ab initio* method. These two philosophies have been largely constructive with each other, sharing ideas, providing criticisms, and validating results [1,11,12,3]. A frequent question used to navigate the zoo of density functionals—perhaps guided by the famous John von Neumann's quote: "with four parameters I can fit an elephant, and with five I can make him wiggle his trunk" [13]—is: "how many parameters does this functional have?". This question, in fact, underlies the more fundamental assumption that the number of parameters is a reliable criterion to evaluate the transferability of the results—but is it really? As pointed out in several occasions [14,15,11,16], counting the number of parameters is not always as straightforward as it might initially appear, especially for functionals that are not directly fitted to data. In fact, there is no such thing as a truly parameter-free or "zero-parameter" *xc* functional approximation,

since even functionals that are usually considered as such have mathematical forms that contain parameters that are then determined based on theoretical arguments. Since the true functional is still unknown, and potentially unknowable,[17] it seems clear that every *xc* functional approximation must contain an empirical element [11].

Instead of counting fitted parameters in "parametrized functionals" and compare them to hidden parameters in "zero-parameter" functionals, the first portion of this article explores the somehow opposite scenario where every functional—regardless of its development philosophy—is represented using a simple function containing one single parameter, as presented in section . This new representation is a direct adaptation of the recent works of Piantadosi [18] and Boué [19], where any distribution of points in any dimension is represented by a well-behaved scalar function with a single real-valued parameter. In other words, quoting Piantadosi's paper title: "One parameter is always enough", even for *xc* functionals. The result of this procedure is that every single functional on the first three rungs of Perdew and Schmidt's "Jacob's ladder" [20] (corresponding to LDA, GGA, and meta-GGA approximations) can be represented by just one single parameter. Famous "zero-parameter" functionals, such as PBE [21] and SCAN [22], as well as popular "parametrized functionals", such as the Minnesota family [23,24,25,26,27,28,29,30,31,32,33,34], are all defined by one number.

Having proven the inadequacy of the "number of parameters" as a measure of transferability of *xc* functionals, the focus of this article shifts to develop a set of statistical criteria that can be appropriately used for this task. Since the exact functional is still unknown, these criteria must rely on statistical analysis of data across as many different chemical and physical properties as possible. Luckily, several benchmark results with hundreds of functionals are already available in the literature [35,36,2,11,12,37,38], but their analysis is not unequivocal, and might even produce contrasting recommendations. This is because the large number of data in these studies can be in principle sliced and grouped into any number of *ad hoc* subsets, that can then be used to statistically validate pretty much any hypothesis. Recent work from the Author's lab has introduced a new unbiased subdivision of some of the most popular DFT databases generated without human intervention by means of data-science algorithms [39]. Interestingly enough, concepts that can be derived using simple chemical intuition have been also recovered by *a posteriori* analysis of the machine-generated groups. This reassuring fact validates the chemical-intuition–based approach that was used by DFT developers to group and analyze the data, but the data-science approach offer several other advantages nonetheless. One of this advantages is demonstrated in Section , where the unbiased subsets are used as the basis for three new statistical criteria obtained adapting the Akaike information criterion (AIC), the Vapnik–Chervonenkis criterion (VCC), and a new cross-validation criterion (CVC) to the DFT results. Preliminary rankings of 60 popular *xc* functionals are also presented and briefly discussed.

2. Fitting elephants: One-parameter fit of exchange–correlation functionals.

This section briefly discusses the application of Piantadosi's encoding procedure [18] to describe any local *xc* functional with a single real-valued parameter $\alpha \in [0,\ 1]$. The simplest case of a generalized gradient approximation (GGA) exchange functional is illustrated first, since it just requires a straightforward monodimensional fit. The more complex case of meta-GGA exchange functionals and GGA exchange–correlation functionals are also presented next. Jupyter notebooks with the code developed for each of these cases accompany the electronic version of this article and are available for download using the interactive features of this special issue and on the Author's github page. These programs allow to obtain single-parameter representations for the majority of the more than 300 *xc* functionals that are included in the LibXC library [40,41].

2.1 GGA exchange functionals

The first step to encode a functional into a single parameter using Piantadosi's procedure is to represent the functional as a series of points. This task is straightforward for GGA functionals, since they depend only

on two variables: the electron density, ρ, and its gradient, $\nabla\rho$. Restricting the discussion to the exchange portion of a general GGA functional, a further simplification can be introduced by decoupling the two variables. The resulting general formula for every GGA exchange functional is thus a simple product of the density-dependent local spin density approximation energy density, $\varepsilon_x^{\mathrm{LSDA}}$, and a gradient-dependent enhancement factor, $F_x^{\mathrm{GGA}}(s)$:

$$E_x^{\mathrm{GGA}} = \int \rho \varepsilon_x^{\mathrm{LSDA}}(\rho) F_x^{\mathrm{GGA}}(s)\, d\mathbf{r}, \quad (1)$$

with the first term simply obtained from the exchange energy density per particle of the uniform electron gas (UEG):

$$\varepsilon_x^{\mathrm{LSDA}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \rho^{\frac{1}{3}}, \quad (2)$$

and the second term usually expressed using the dimensionless reduced variable, s :

$$s = \frac{|\nabla\rho|}{2(3\pi^2)^{\frac{1}{3}}\rho^{\frac{4}{3}}}. \quad (3)$$

Therefore, the shape of every GGA exchange functional is uniquely determined by its enhancement factor, which can then be represented as a set of N equidistant points on a grid in the finite variable $u \in [0,1]$, obtained from $s \in [0, \mathrm{inf})$ using Becke's transformation [7]:

$$u = \frac{s^2}{1+s^2}. \quad (4)$$

This numerical representation becomes exact in the limit of infinite number of points, $N \to \infty$. As previously demonstrated [42], a grid of just simply $N = 20$ points is practically sufficient to describe the enhancement factors of most exchange GGA functionals (e.g. PBE [21] and B88 [43]) with sub-milliHartrees precision, when used in conjunction with a well-behaved interpolation between the points—such as a cubic or univariate spline. For a handful of more complicated functionals (e.g. SOGGA11 [44]) a slightly finer grid of $N = 100$ points will suffice to achieve accuracies of ~10^{-6} Hartrees.

Once the functional is defined on the grid, the simple sequence of points $x \in [0, ..., N]$ can be represented using Piantadosi's formula:

$$f_\alpha(x) = \sin^2\left(2^{\beta x} \arcsin\sqrt{\alpha}\right), \quad (5)$$

which is uniquely defined by a single parameter $\alpha \in \mathbb{R}$, and a constant $\beta \in \mathbb{N}$ that controls the accuracy of the encoding procedure. It is important to notice that eq. 5 only reproduces the position of the points, while the spline interpolation is still required to obtain a continuous function over the considered interval (an exact fit would require $N \to \infty$, and therefore an infinitely long encoding parameter). The drawback of this procedure is that eq. 5 is extremely sensitive to the value of the parameter. Hence α has to be represented using a huge number of significant digits. In fact, the entire point of this exercise is to encode the full complexity of the GGA exchange enhancement into the length of the single parameter. Such length (i.e. the number of significant digits required to write α) depends on both the number of interpolation points that are used to represent the functional on the grid, and the accuracy parameter β. In general, $N = 20$ interpolation points and $\beta = 8$ can be used to represent simple GGA exchange functionals—such as PBE [21]—with relative errors in the description of the enhancement factor smaller than 1%, resulting in parameters that require ~ 60 digits. For functionals that have some oscillation over the entire interval of u—such as SOGGA11 [44]—$N = 100$ interpolation points and a value of $\beta = 12$ are required for similar accuracies, resulting in parameters with ~ 350 digits. The single parameters for both the PBE and SOGGA11 functionals are reported in Fig. 1, together with the corresponding plots of the enhancement factors, F_x, as a function of u and s. A Jupyter notebook with the details of the encoding procedure—as well as an algorithm to evaluate the errors for both the spline implementation and the encoding procedure—is also associated with the Figure and is available on github.

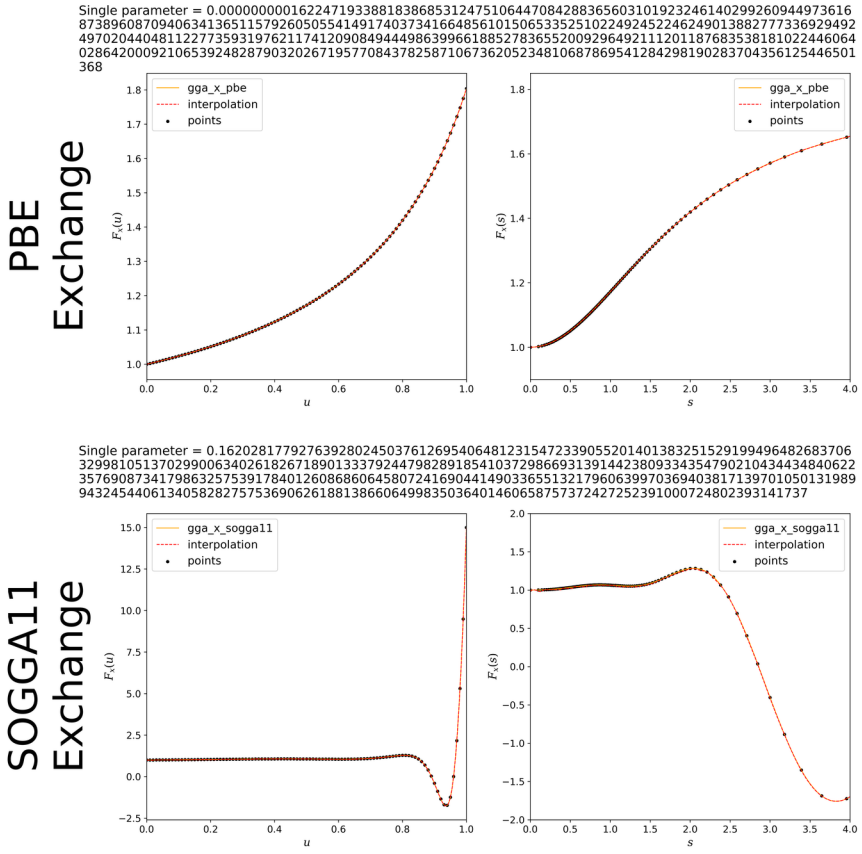


Figure 1: Single parameter representation for the PBE (upper panel) and SOGGA11 (lower panel) GGA exchange functionals as a function of the reduced density variables u, eq. **??** (left plots) and s, eq. **??** (right plots). For both functionals, the black dots are the decoded points, the orange solid curve is the original enhancement factor as obtained directly from LibXC, and the dashed red curve is the result of the decoding of the single parameter and the interpolation via univariate cubic spline. Results are obtained with $N = 100$ points and $\beta = 12$. A Jupyter notebook to encode every GGA exchange functional in LibXC, as well as to reproduce the plots and to calculate the encoding and interpolation errors is associated with the Figure.

2.2 Meta-GGA exchange functionals

The next rung in Perdew's Jacob ladder is those of meta-GGA functionals. Restricting the discussion once again to exchange functionals only, the enhancement factor for meta-GGA functionals depends only on two variables, the gradient of the density and the orbital-dependent local kinetic energy density:

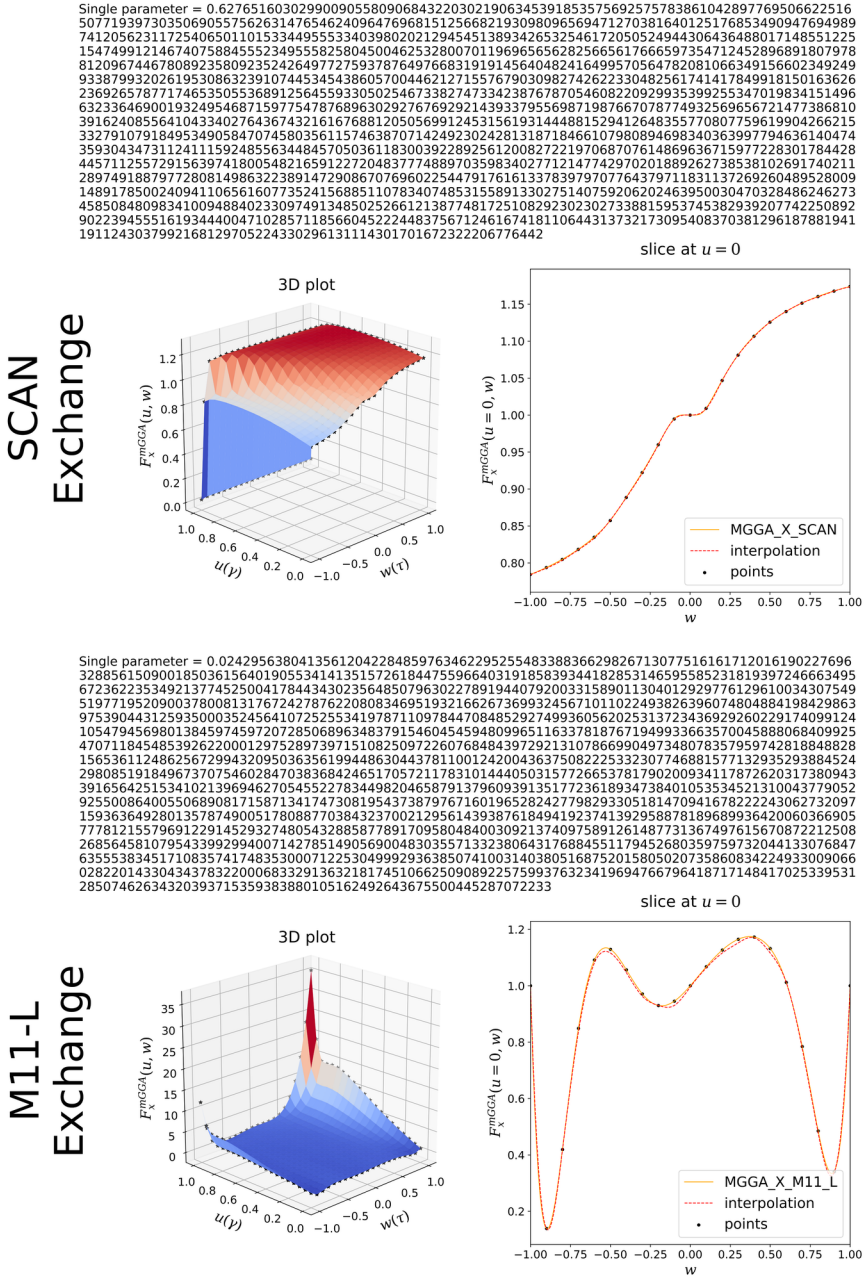
$$\tau = \frac{1}{2}\sum_i |\nabla\psi_i|^2 . \quad (6)$$

The meta-GGA enhancement factor can be easily represented by points on a two-dimensional grid using a simple extension of the code used in the previous case. The steps in this extension include using the popular transformation of τ into the finite variable $w \in [-1, 1]$ [45]:

$$w = \frac{[\frac{3}{10}\left(3\pi^2\right)^{2/3}\rho^{5/3}]\tau^{-1} - 1}{[\frac{3}{10}(3\pi^2)^{2/3}\rho^{5/3}]\tau^{-1} + 1}, \quad (7)$$

followed by the usage of a grid of $N \times N$ equidistant points on u and w. A two-dimensional spline (either

bicubic or univariate) is then used to interpolate between points on the considered interval. The implementation of Piantadosi's encoding procedure is then identical to the previous case, with the only difference that the series of points are now constructed as $x \in [(0,0),...,(0,N),(1,0),...,(1,N)]$. Once again, the accuracy of the procedure depends only on two variables, the number of points used to interpolate the enhancement factor, N^2, and the accuracy of the encoder parameter, β. The major hurdle in the procedure is that the number of digits required to represent the parameter is now much higher than for the previous case. Interpolations with $N > 20$ become computationally expensive since they require > 400 points, and result in parameters with more than 1500 digits, regardless of the value of β. For well-behaved functionals, however, $N = 20$ and $\beta = 12$ result in parameters with ˜1500 digits, and overall errors < 1 %, similarly to the GGA case. Single parameters for the exchange enhancement factors of the SCAN [22] and the M11-L [30] meta-GGA functionals are reported in Fig. 2 as a three dimensional surface and a corresponding slice at $u = s = 0$. A Jupyter notebook with the details of the encoding procedure is also associated with the Figure and is available on github.


Single parameter = 0.627651603029900905580906843220302190634539185357569257578386104289769506622516
507719397303506905575626314765462409647696815125668219309809656947127038164012517685349094769498974120562311725406501101533449555334039802021294545138934265325461720505249443064368801714855122515474991214674075884555234955582580450046253280070119696565628256656176665973547124528968918079788120967446780892358092352426497727593787649766831919145640482416499570564782081066349156602349249933879932026195308632391074453454386057004462127155767903098274262233048256174141784991815101636262369265787717465350553689125645593305025467338274733423876787054608220929935399255347019834151496632336469001932495468715977547876896302927676929214393379556987198766707877493256965672147738681039162408556410433402764367432161676881205056991245315619314448815294126483557708077596199042662153327910791849534905847074580356115746387071424923024281318718466107980894698340363997794636140474359304347311244115924855634484570503611830039228925612008272219706870761486963671597722830178442844571125572915639741800548216591227204837774889703598340277121477429702018892627385381026917402112897491887797728081498632238914729086707696022544791761613378397970776437971183113726926048958200914891785002409411065616077352415688511078340748531558913302751407592062024639500304703284862462734585084809834100948840233097491348502526612138774817251082923023027338815953745382939207742250892902239455516193444004710285711856604522244837567124616746181106443137321730954083703812968178819411911243037992168129705224330296131114301701672322206776442">

Single parameter = 0.024295638041356120422848597634622952554833883662982671307751616171201619022769632885615090018503615640190553414135157261844755966403191858393441828531465955852318193972466634956723622353492137745250041784434302356485079630227891944079200331589011304012929776129610034307549519771952090037800813176724278762208083469519321662673699324567101102249382639607480488419829486397539044312593500035245641072525534197871109784470848529274993605620253137234369292602291740991241054794569801384597459720728506896348379154604545948099651163378187671949933663570045888068409925470711845485392622000129752897397151082509722607684843972921310786699049734807835795974281884882815653611248625672994320950363561994486304437811001242004363750822253323077468815771329352938845242980851918496737075460284703836842465170572117831014440503157726653781790200934117872620317380943391656425153410213969462705455227834498204658791379609391351772361893473840105353452131004377905292550086400550689081715871341747308195437387976716019652824277982933051814709416782222430627320971593636492801357874900517808877038432370021295614393876184941923741392958878189689936420060366905777812155796912291452932748054328858778917095804840030921374097589126148773136749761567087221250826854581079543399299400714278514905690048303557133238064317688455117945268035975973204413307684763555383451710835741748353000712253049992936385074100314038051687520158050207358608342249330090660282201433043437832200068332913632181745106625090892257599376323419694766796418717148417025339531285074626343203937153593838801051624926436755004452870722333">

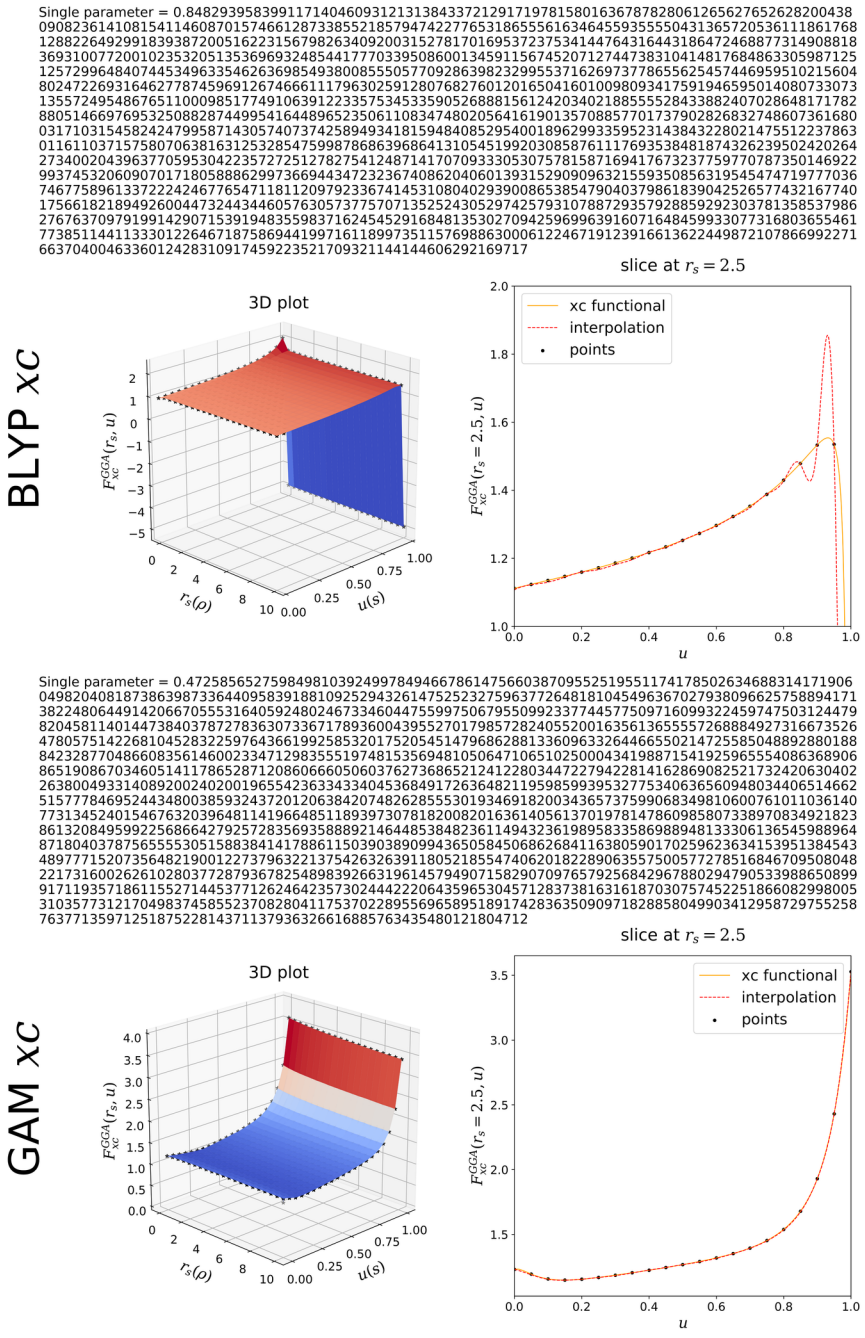
Figure 2: Single parameter representation for the SCAN (upper panel) and M11-L (lower panel) meta-GGA exchange functionals as a 3D function (left plots) of the reduced density variables u, eq. **??**, and the local kinetic energy density variable w, eq. **??**. The right plots represent slices at constant $u = 0$. For both functionals, the black dots are the decoded points, the orange solid curve is the original enhancement factor as obtained directly from LibXC, and the dashed red curve is the result of the decoding of the single parameter and the interpolation via univariate cubic spline. Results are obtained on a grid of 20×20 points and $\beta = 12$. A Jupyter notebook to encode every meta-GGA exchange functional in LibXC, as well as to reproduce the plots and to calculate the encoding and interpolation errors is associated with the Figure.

2.3 Exchange–correlation functionals

The extension to include correlation functionals is trivial, especially in the GGA case. The general shape of the enhancement factor of every GGA xc functional can in fact be represented using just two variables that depend on the density and its gradient, using the Wigner-Seitz Radius:

$$r_s = \left(\frac{3}{4\pi\rho}\right)^{\frac{1}{3}}, (8)$$

and one of the reduced density gradient variables introduced above (either s or u). The implementation of a two-dimensional interpolation and encoding procedure for GGA exchange–correlation functionals is reported in Fig. 3, using a grid of $N \times N$ points on r_s and u . Since a three-dimensional interpolation is necessary, the same numerical complication of the previous case apply. In general, most GGA xc functionals can be interpolated using $N = 20$ and encoded into single parameters with ˜1500 digits using $\beta = 12$. In Fig. 3 and related Jupyter notebook, the encoding procedure is applied to the BLYP GGA xc functional [43,46] and to the GAM NGA xc functional [47]. Single parameters of ˜1500 digits are obtained and reported. It is important to recognize that the BLYP functional diverges at $u = 1$ ($s = \infty$), hence the interpolation error for $N = 20$ grows substantially in the region where $u > 0.8$ ($s > 2$). The interpolation error can be further reduced by increasing N, pushing it to regions of s that are not very significant for chemical systems. Nevertheless, the $s \to u$ transformation is not ideal for functionals that diverge at the extremes.



BLYP *xc*

GAM *xc*

Figure 3: Single parameter representation for the BLYP (upper panel) and GAM (lower panel) exchange–correlation functionals as a 3D function (left plots) of the Wigner-Seitz radius r_s, eq. **??** and reduced density variables u, eq. **??**. The right plots represent slices at constant $r_s = 2.5$. For both functionals, the black dots are the decoded points, the orange solid curve is the original enhancement factor as obtained directly from LibXC, and the dashed red curve is the result of the decoding of the single parameter and the interpolation via univariate cubic spline. Results are obtained on a grid of 20×20 points and $\beta = 12$. A Jupyter notebook to encode every GGA or NGA exchange–correlation functional in LibXC, as well as to reproduce the plots and to calculate the encoding and interpolation errors is associated with the Figure.

Extension to meta-GGA exchange–correlation functionals, as well as to functionals with more complex forms sitting on higher rungs of Jacob's ladder, could be achieved with various degrees of difficulty. For example, meta-GGA functionals depend on at least three variables that cannot be decoupled (e.g. the density, its gradient, and the kinetic energy density), and therefore they require higher dimensional interpolations. The interpolation using multi-dimensional grids and appropriate functions is not problematic, especially using available python libraries. A slightly more complicated case is the case of hybrid functionals (e.g. functionals that include a fraction of Hartree–Fock exchange), for which the parameter that represents the fraction of HF exchange could be encoded in the procedure, for example at the beginning of the sequence. For range-separated hybrid functionals, more complicated *ad hoc* procedure must be designed. However, since representing functionals with one parameter has no inherent benefit for DFT as a method, going beyond the simple proof-of-principle described above has very little scientific merit and is not explored further in this context. A more rewarding endeavor is the search for a procedure that does not rely on counting the number of parameters to evaluate the transferability of functionals, as presented in the next section.

3. Statistical criteria of bias–variance tradeoff and analysis of 60 exchange–correlation functionals

The dispute between counting parameters and analytical fits is not a new scene in statistics and machine learning, where the problem is generally known as the bias/variance dilemma [48]. Especially in supervised learning, where a model is learned from (fitted to) some training data, this dilemma translates to the necessity to strike a balance between underfitting the data (bias error), resulting in methods that don't incorporate all the relations between the data, and overfitting them (variance), resulting in methods that are poorly transferrable. Several criteria for model selection are available in this context, and they all generally include two components, one that accounts for the performance of the model on the training data, and another that accounts for the transferability of the model to unseen data. For a good introduction, see [49]. The goal of the next section is to borrow some of the methodologies that have been developed in the context of supervised learning and apply them to the analysis of DFT approximations.

3.1 Statistical Criteria for functional evaluation

In order to introduce appropriate bias–variance criteria for *xc* functionals, well-established model validation techniques from statistical analysis must be used. Several criteria are available in statistics for model selection and validation, mostly belonging to three main classes:

- Methods based on information criteria [50].
- Methods obtained from Vapnik–Chervonenkis theory [51].
- Resampling methods [52,53].

In general, the first two classes include analytic methods that evaluate the overall uncertainty (risk) of the model by inflating the error of the fitted model calculated on the training set (or some appropriate data set) by a penalty factor that depends on the degrees of freedom (DoF) of the model and the number of data in the set. These methods usually have to rely on assumptions on both the type of function that is estimated and the statistical distribution of the data. The third class of models require external data sets for validation, and is usually more computational demanding, however it has the advantage of not relying on any assumptions on the distribution of the errors, nor the training data. Among the first class, the Akaike information criterion (AIC) [50] is the most widely used estimator of error prediction. This coefficient is constructed from maximum likelihood arguments, and it uses an additive formula to evaluate the overall risk, R , as:

$$R = R_{\text{emp}} + f\left(n, p\right), \quad (9)$$

where the empirical risk, R_{emp}, represents the error of the fitted model calculated on the training set, and should not be confused with the error associated with the comparison of DFT data and empirical (experimental) results, in a chemical sense. In order to evaluate R_{emp}, the recent ASCDB database can be used, since it was specifically created to evaluate the performance of DFT functionals. To account for the large differences in the average of the absolute reference energies of each subset of ASCDB, it is convenient to introduce here an overall weighted mean unsigned error (wMUE), calculated from the mean unsigned errors of the individual subsets, MUE_i , using:

$w\mathrm{MUE} = \sum_{i=1}^{16} w_i \mathrm{MUE}_i$ (10)

where the individual weights are calculated from the ratio between the average of the absolute reference energies for each subset, $\left|\overline{\Delta \mathrm{E}}\right|_i$, and that of the overall database (which is 6.988 kcal/mol for ASCDB, weights for this database are provided within the Jupyter notebook that accompany the electronic version of this article and on the Author's github page):

$$w_i = \left(\left|\overline{\Delta \mathrm{E}}\right|_i \sum_{i=1}^{16} \frac{1}{\left|\overline{\Delta \mathrm{E}}\right|_i} \right)^{-1} = \frac{6.988 \frac{\mathrm{kcal}}{\mathrm{mol}}}{\left|\overline{\Delta \mathrm{E}}\right|_i}. \tag{11}$$

This quantity is a slightly simplified version of the WTMAD-2 indicator introduced by Goerigk et al. [2] to rank functionals based on the performance on the GMTKN55 database. On the statistical standpoint, wMUE is the average of a slightly modified coefficient of variation for each subset $\left(m\mathrm{CV}_i = \mathrm{MUE}_i / \left|\overline{\Delta \mathrm{E}}\right|_i\right)$ where the standard deviation is replaced by the mean unsigned error of each subset

$$w\mathrm{MUE} = \frac{\sum_{i=1}^{16} m\mathrm{CV}_i}{6.988 \frac{kcal}{mol}}. \tag{12}$$

This replacement is justified by the fact that the MUE has similar content to the root mean square error (which, up to a constant in the definition, is the standard deviation of the distribution of the errors) but it is usually preferred in the DFT literature as an indicator of functional performance. While Goerigk et al. warned not to use their weighted MUE indicators as an estimation of statistical error for specific chemical problems,[2] the connection with the coefficient of variation makes wMUE useful beyond classification purposes as a balanced measure of the empirical risk of a functional, as demonstrated by the results presented below. It is important to keep in mind though that—in accordance with Goerigk et al.'s suggestion—weighted MUE values for different databases should never be compared in absolute terms, because they intrinsically depend on the molecules that are included in each database, and their main purpose is to provide a basic criterion for the ranking of functionals.

$f(n,p)$ in Eq. **??** is an additive penalty function that depends on the number of training data, n , and the degrees of freedom (number of free parameters) of the fitted function, p . Assuming a gaussian distribution of the errors, the penalty function can be calculated for regression models as:

$f\,(n,p) = \frac{2p}{n}\sigma^2,$ (13)

with:

$\sigma^2 = \frac{n}{n-p} R_{\mathrm{emp}}$, (14)

resulting in a final formula for AIC that is:

$\mathrm{AIC} = w\mathrm{MUE} \cdot \left(1 + \frac{2p}{n-p}\right)$.(15)

Among the second class of methods the Vapnik–Chervonenkis criterion (VCC) [51] can be selected. This criteria inflates the empirical risk by a multiplicative penalty function related to Vapnik's measure:

$$\mathrm{VCC} = w\mathrm{MUE} \cdot \left(1 - \sqrt{p - p\ln p + \ln(n/2n)}\right)^{-1} . (16)$$

For both AIC and VCC, $n = 200$ when evaluated using ASCDB, while p is an estimation of the degrees of freedom (DoF) that is equal to the number of fitted parameters for fitted functionals, while it is set to 1 for non-fitted ones (Table 1).

The definition of a resampling criterion for *xc* functionals is unfortunately not as straightforward, since in several cases it might be difficult to find data that can be used as an external, unbiased, validation set to be used in cross-validation methods. As such, cross-validation criteria are intrinsically dependent on the data set that is used to obtain them [52,53], and particular effort has to be devoted to creating a criterion that is representative and transferable across as many functionals and data sets as possible. The purpose of cross-validation is, in practice, to highlight inconsistencies in the treatment of external data, when compared to the data that are used for the training of the parameters. Therefore, every overfitted model present a large difference between the errors for the training set and those for the validation set. The major hurdle in the evaluation of *xc* functionals from different sources and development philosophies is to find two appropriate and independent sets of data that can function as a "training set" and as a "validation set". On the one hand, the first 12 subsets of ASCDB include chemical systems that are conventionally used to train and evaluate computational methods. While none of the existing functionals is specifically trained on all molecules of these subsets, most of the modern fitted functionals were trained on similar systems (e.g. the Minnesota and wB97 families), and even non-fitted ones (e.g. revTPSS and SCAN) have been subject to convergent evolution to provide at least reasonable results for those basic chemical systems. On the other hand, the last four subsets of ASCDB contain unconventional systems that are very far from what current functionals have been designed or trained for and represent a good dataset for validation. (The three main subsets in this category comes from the mindless benchmark database of Grimme and coworkers, while the last one includes the energies of atoms on a per-electron basis.) A simple cross-validation measurement of the overfitting of a functional can then be obtained from the ratio between the MUE of the unbiased calculation—used as a "validation set"—and the overall *w*MUE of ASCDB—used as the "training set". Interpreting this last quantity as a cross-validation estimate of the unknown noise variance of the distribution of the errors, σ^2, the cross-validation criterion (CVC) can then be calculated by inflating the empirical risk using eqs. **??** and **??**, as:

$$\mathrm{CVC} = w\mathrm{MUE} + \frac{2h}{n}\frac{\mathrm{MU}E_{\mathrm{UC}}}{w\mathrm{MUE}}, (17)$$

where the *w*MUE of the full ASCDB database is used at the denominator in place of the MUE (or weighted MUE) of the first twelve subsets of ASCDB because numerical evidence showed no significant differences in the rankings when this transformation was performed. Apart from a much simpler formula to calculate CVC, the main advantage of using the weighted MUE of the entire database is that eq. **??** also becomes extensible to other databases. For example, a straightforward extension of CVC to the GMTKN55 database is obtained by using the overall WTMAD-2 at the numerator, and the MUE (MAD using Goerigk et al. notation) for the mindless benchmark subset at the denominator:

$$\mathrm{CVC}^{\mathrm{GMTKN55}} = \mathrm{WTMAD2} + \frac{2h}{n}\frac{\mathrm{MAD}_{\mathrm{MB1643}}}{\mathrm{WTMA}D2} . (18)$$

As for the *w*MUE case discussed before, it is important to keep in mind that, despite providing very similar rankings, CVC values from different databases are difficult to compare in absolute terms because they intrinsically depend on the molecules that are included in each database.

3.2 Evaluation of 60 exchange–correlation functionals

The usefulness of the three criteria described above can be evaluated on the set of 60 popular exchange–correlation functional approximations.[54,55,35,56,57,58,59,60,61,62,63,43,46,64,65,66,67,68,69,70,22,27,71,72,23,24,33,73,74,21,75,31,76,77,78,79,28,31,80,81,8234,3283,30,25,84,85,86,87,88] The functionals is an expanded set of those that were originally selected to develop the ASCDB database, and they include a broad set of approximations across all different

rungs of Jacob's ladder, as well as several decades of functional development. The list of all used functionals, as well as detailed reference to their original publications, are reported in the last column of Table 1.

All calculations are stable broken-symmetry solutions close to the complete basis set limit, and have been performed using quadruple-ζ quality basis sets using Q-Chem 5.1 [89]. Results for the three statistical criteria, AIC, VCC, and CVC, are reported in Table 1, as well as the ranking of each functional according to each specific criterion (in parenthesis). The average ranking of each functional across the three criteria is also reported in the last column of Table 1, and is used as the final indicator for performance of a functional. It is clear that the rankings obtained using the statistical criteria align well with the Jacob's ladder picture of functional approximations. According to all three criteria, for example, the three best functionals are double-hybrid "fifth rung" approximations. "Fourth rung" hybrid meta-GGA/NGA are the second-best class, followed by "fourth rung" hybrid GGA/NGA and "third rung" local meta-GGA, with similar average performance. "First and second rung" Local GGA/NGA are on average at the bottom of the rankings. Interestingly enough, modern non-fitted functionals such as PBE and SCAN-D3(BJ) sits in the middle of the ranking, together with most of the parametrized Minnesota functionals. Even more interesting than the general trends are some of the outliers. For example, the B3LYP-D3(BJ) ranks near the top according to all three criteria, while its parent functional B3LYP is consistently ranked at the bottom, more than 23 positions below B3LYP-D3(BJ), confirming the trends observed in the literature. However, PBE-D3(BJ) is slightly more transferable (despite a slightly higher wMUE) than B3LYP-D3(BJ), confirming recent finding of transferability issues in the popular B3LYP-D3(BJ) functional.

	DoF	wMUE	AIC	VCC	CVC	AVG ranking	Ref
DSD-PBEP86-D3(BJ)	7	2.14	2.29 (1)	3.61 (1)	2.3 (1)	1	54
PWPB95-D3(BJ)	10	2.69	2.97 (2)	4.99 (2)	2.85 (2)	2	34;55
B2PLYP-D3(BJ)	5	3.33	3.5 (3)	5.21 (3)	3.47 (3)	3	55;56
wB97M-V	12	3.43	3.87 (4)	6.76 (5)	3.76 (4)	4.33	57
PW6B95-D3(BJ)	9	3.75	4.11 (5)	6.76 (4)	3.98 (5)	4.67	58;59
PW6B95	6	5.27	5.6 (6)	8.58 (9)	5.43 (8)	7.67	59
PBE0-D3(BJ)	4	5.45	5.68 (7)	8.19 (8)	5.53 (9)	8	58;60
HSE-HJS	1	5.7	5.75 (9)	7.23 (6)	5.73 (15)	10	61;62
B3LYP-D3(BJ)	6	5.42	5.76 (10)	8.82 (11)	5.63 (13)	11.33	43;46;58;63;64;65
B97M-rV	12	5.05	5.69 (8)	9.93 (16)	5.53 (10)	11.33	66;67
PBE0	1	5.74	5.79 (12)	7.28 (7)	5.76 (16)	11.67	60
B97M-V	12	5.12	5.77 (11)	10.07 (17)	5.61 (11)	13	68;69
wB97X-V	10	5.33	5.89 (13)	9.89 (15)	5.7 (14)	14	70
SCAN-D3(BJ)	2	6.32	6.45 (14)	8.58 (10)	6.35 (24)	16	21;58
M06-2X	29	4.86	6.5 (15)	14.37 (34)	5.36 (7)	18.67	26
revPBE-D3(BJ)	4	6.49	6.76 (17)	9.74 (14)	6.6 (26)	19	58;71
B97-1	10	5.95	6.58 (16)	11.05 (21)	6.34 (23)	20	72
M05	22	5.47	6.82 (19)	13.84 (29)	6.1 (18)	22	22
M05-2X	19	5.82	7.05 (20)	13.74 (28)	6.26 (19)	22.33	23
MN15	59	3.68	6.76 (18)	20.16 (44)	5.35 (6)	22.67	32
BMK	17	6.04	7.16 (23)	13.57 (27)	6.27 (20)	23.33	73
M06-2X-D3(0)	35	5.03	7.16 (22)	16.89 (38)	5.62 (12)	24	26;74
PBE	1	7.41	7.48 (25)	9.39 (12)	7.43 (35)	24	20
revTPSS-D3(BJ)	6	6.76	7.17 (24)	10.99 (20)	6.99 (31)	25	58;75
N12-SX	26	5.51	7.15 (21)	15.26 (37)	5.97 (17)	25	30
PW91	1	7.56	7.64 (28)	9.59 (13)	7.58 (37)	26	76;77
t-HCTHh	17	6.36	7.54 (26)	14.28 (32)	6.98 (30)	29.33	78
TPSSh	1	8.04	8.12 (33)	10.2 (18)	8.07 (41)	30.67	79
PBE-D3(BJ)	3	7.81	8.05 (31)	11.19 (23)	7.87 (39)	31	20;58
B97-D3(0)	9	7.23	7.91 (30)	13.02 (26)	7.61 (38)	31.33	58;75
M06-D3(0)	39	5.09	7.56 (27)	18.58 (41)	6.63 (27)	31.67	26;58
B3PW91	3	7.82	8.06 (32)	11.2 (24)	7.9 (40)	32	46;63;76;77
M06	33	5.58	7.78 (29)	17.98 (40)	6.85 (29)	32.67	26
TPSS	1	8.39	8.48 (37)	10.64 (19)	8.43 (44)	33.33	78
M11-D3(BJ)	46	5.18	8.28 (34)	21.81 (49)	6.29 (22)	35	68;79
M08-HX	47	5.24	8.46 (36)	22.5 (50)	6.29 (21)	35.67	27
revTPSS	1	8.79	8.88 (39)	11.15 (22)	8.83 (47)	36	75
M11	40	5.6	8.4 (35)	20.86 (46)	6.64 (28)	36.33	79

It is important to highlight that this study is primarily intended to establish the reliability of the statistical criteria for evaluation of performance and transferability of functionals. While the reported rankings can be used to establish some trends, the list of functionals is not comprehensive enough to provide reliable final suggestions on which functional to pick among the more than 300 available in the literature. Some conclusion on the performance and transferability of the considered functionals are still interesting to report, and are as follows:

- Best double-hybrid: DSD-PBEP86-D3(BJ) [54], alternate: PWPB95-D3(BJ)[35,55].
- Best hybrid meta-GGA: ωB97M-V [57], alternate: PW6B95-D3(BJ) [59,58].
- Best hybrid-GGA: PBE0-D3(BJ) [60,90,58], alternate: B3LYP-D3(BJ) [63,43,46,64,65,58].
- Best local meta-GGA: B97M-rV[66,67], alternate: SCAN-D3(BJ) [22,58].
- Best local GGA: PBE [21], alternate: PW91 [77].

These results are strengthened by the fact that the majority of the highlighted functionals overlap with the top performers suggested in recent reviews by Head-Gordon's [12], Goerigk's [3], and Grimme's [2] groups, obtained with larger databases and considering a broader spectrum of functionals. Finally, connecting the transferability results to the issue of counting the number of parameters presented in Section , the summary of the results plotted in Fig. 4 demonstrates a clear lack of correlation between the average ranking of each functional and its number of degrees of freedom. This lack of correlation supports the main message of this work: The number of fitted parameters does not represent an effective measure of the transferability of a functional. More reliable statistical criteria—such as those developed in this work, or alternatively, the probabilistic performance estimator recently introduced by Pernot and Savin [91,92]—should be used to evaluate the reliability of new and existing xc functionals.

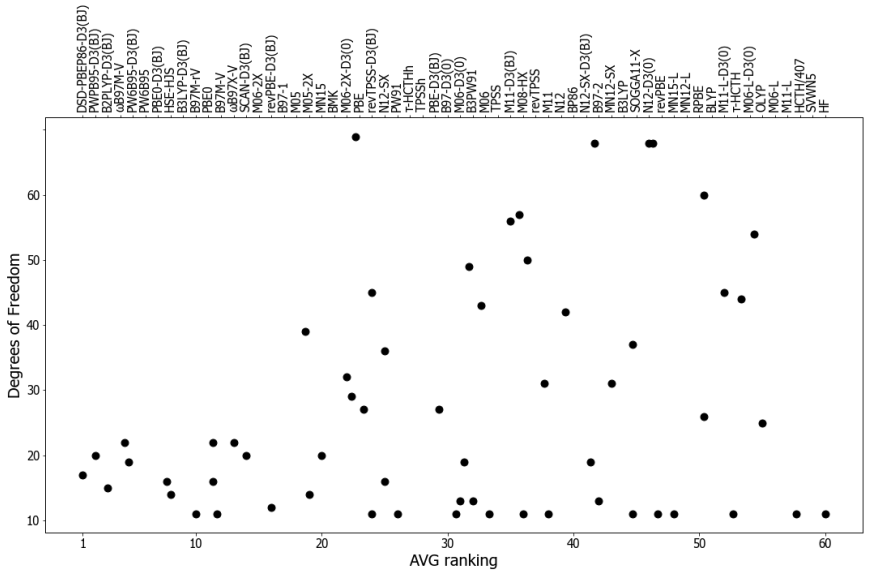


Figure 4: Scatter plot of the statistical analysis of 60 exchange–correlation functionals. The results shCiao ow no correlation between the degrees of freedom of a functional—a loose count of the number of its fitted parameters (see text)—and its transferability—measured as the average ranking across the three new statistical criteria.

Conclusions

A simple encoding procedure borrowed from data science was used to show that the number of parameters of a fitted exchange–correlation functional (or in a general sense, its degrees of freedom) are not representative of transferability across different chemical systems. In section , more than 300 functionals from the LibXC DFT library are represented using one single parameter. This exercise disentangles the arbitrary measurement "number of parameter" from the fundamental concept of transferability of the results, and validates the proposition of Yu and Truhlar [11] reading: "Counting parameters in a density functional is a little bit like evaluating the quality of a research program by counting the publications it produces—the number of publications is hardly irrelevant, but it is far from the whole story, and usually it is not the decisive measure of quality."

To compensate for this lack of a "decisive measure of quality", three new criteria based on the statistical analysis of the recently proposed ASCDB database of chemical data were developed in section for the assessment of exchange–correlation functional approximations. These criteria are the Akaike information criterion (AIC), the Vapnik–Chervonenkis criterion (VCC), and the cross-validation criterion (CVC). While the criteria mostly provide similar rankings, some differences between them do exist, and the average ranking across the three criteria is the most unambiguous measurement for the evaluation of functionals.

Preliminary results of the average ranking with 60 functionals show that the best ones are those that carefully use a flexible mathematical form with a modest number of appropriately fitted parameters (5–12). In the debate between different functional development philosophies, occupying the middle ground seems to be the current winning strategy.

References

1.A. D. Becke, *Journal of Chemical Physics*, **2014**, 140, 18A301–19.

2.L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Physical Chemistry Chemical Physics*, **2017**, 19, 32184–32215.

3.L. Goerigk and N. Mehta, *Australian Journal of Chemistry*, **2019**, 72, 563.

4.H. Jacobsen and L. Cavallo, in *Handbook of Computational Chemistry*, Springer International Publishing, Cham, **2017**, pp. 225–267.

5.G. Pacchioni, *Catalysis Letters*, **2015**, 145, 80–94.

6.A. D. Becke, *INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY*, **1983**, 23, 1915–1922.

7.A. D. Becke, *The Journal of Chemical Physics*, **1997**, 107, 8554–8560.

8.D. C. Langreth and J. P. Perdew, *Solid State Communications*, **1979**, 31, 567–571.

9.D. C. Langreth and J. P. Perdew, *Physical Review B*, **1980**, 21, 5469–5493.

10.J. P. Perdew, *Physical Review Letters*, **1985**, 55, 1665–1668.

11.H. S. Yu, S. L. Li and D. G. Truhlar, *Journal of Chemical Physics*, **2016**, 145, 130901–24.

12.N. Mardirossian and M. Head-Gordon, *Molecular Physics*, **2017**, 115, 2315–2372.

13.F. Dyson, *Nature*, **2004**, 427, 297–297.

14.G. K.-L. Chan and N. C. Handy, *Journal of Chemical Physics*, **2000**, 112, 5639–5653.

15.R. Peverati and D. G. Truhlar, *Philosophical Transactions Of The Royal Society Of London Series A-Mathematical Physical And Engineering Sciences*, **2014**, 372, 20120476.

16.B. Civalleri, D. Presti, R. Dovesi and A. Savin, in *Chemical Modelling*, **2012**, pp. 168–185.

17.N. Schuch and F. Verstraete, *Nature Physics*, **2009**, 5, 732–735.

18.S. T. Piantadosi, *AIP Advances*, **2018**, 8, 095118.

19.L. Boué, *arXiv:1904.12320 [cs, stat]*, **2019**.

20.J. P. Perdew and K. Schmidt, *AIP Conference Proceedings*, **2001**, 577, 1–20.

21.J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, **1996**, 77, 3865–3868.

22.J. Sun, A. Ruzsinszky and J. P. Perdew, *Physical Review Letters*, **2015**, 115, 036402.

23.Y. Zhao, N. E. Schultz and D. G. Truhlar, *The Journal of Chemical Physics*, **2005**, 123, 161103.

24.Y. Zhao, N. E. Schultz and D. G. Truhlar, *Journal of Chemical Theory And Computation*, **2006**, 2, 364–382.

25.Y. Zhao and D. G. Truhlar, *The Journal of Chemical Physics*, **2006**, 125, 194101.

26.Y. Zhao and D. G. Truhlar, *The Journal of Physical Chemistry A*, **2006**, 110, 13126–13130.

27.Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, **2008**, 120, 215–241.

28.Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory And Computation*, **2008**, 4, 1849–1868.

29.R. Peverati and D. G. Truhlar, *The Journal of Physical Chemistry Letters*, **2011**, 2, 2810–2817.

30.R. Peverati and D. G. Truhlar, *The Journal of Physical Chemistry Letters*, **2012**, 3, 117–124.

31.R. Peverati and D. G. Truhlar, *Physical Chemistry Chemical Physics*, **2012**, 14, 16187–16191.

32.R. Peverati and D. G. Truhlar, *Physical Chemistry Chemical Physics*, **2012**, 14, 13171–13174.

33.H. S. Yu, X. He, S. L. Li and D. G. Truhlar, *Chemical Science*, **2016**, 7, 5032–5051.

34.H. S. Yu, X. He and D. G. Truhlar, *Journal of Chemical Theory And Computation*, **2016**, 12, 1280–1293.

35.L. Goerigk and S. Grimme, *Journal of Chemical Theory And Computation*, **2010**, 6, 107–126.

36.L. Goerigk and S. Grimme, *Journal of Chemical Theory And Computation*, **2011**, 7, 291–309.

37.G. Santra, N. Sylvetsky and J. M. L. Martin, *The Journal of Physical Chemistry A*, **2019**, 123, 5129–5143.

38.J. M. L. Martin and G. Santra, *Israel Journal of Chemistry*, n/a.

39.P. Morgante and R. Peverati, *Physical Chemistry Chemical Physics*, **2019**, 21, 19092–19103.

40.M. A. L. Marques, M. J. T. Oliveira and T. Burnus, *Computer Physics Communications*, **2012**, 183, 2272–2281.

41.S. Lehtola, C. Steigemann, M. J. T. Oliveira and M. A. L. Marques, *SoftwareX*, **2018**, 7, 1–5.

42.R. Peverati and D. G. Truhlar, *Journal of Chemical Theory And Computation*, **2011**, 7, 3983–3994.

43.A. D. Becke, *Physical Review A*, **1988**, 38, 3098–3100.

44.R. Peverati, Y. Zhao and D. G. Truhlar, *The Journal of Physical Chemistry Letters*, **2011**, 2, 1991–1997.

45.A. D. Becke, *The Journal of Chemical Physics*, **2000**, 112, 4020–4026.

46.C. Lee, W. Yang and R. G. Parr, *Physical Review B*, **1988**, 37, 785–789.

47.H. S. Yu, W. Zhang, P. Verma, X. He and D. G. Truhlar, *Physical Chemistry Chemical Physics*, **2015**, 17, 12146–12160.

48.S. Geman, E. Bienenstock and R. Doursat, *Neural Computation*, **1992**, 4, 1–58.

49.V. S. Cherkassky and F. Mulier, *Learning from data: concepts, theory, and methods*, IEEE Press : Wiley-Interscience, Hoboken, N.J, 2nd ed., **2007**.

50.H. Akaike, *IEEE Transactions on Automatic Control*, **1974**, 19, 716–723.

51.V. N. Vapnik and A. Y. Chervonenkis, *Theory of Probability & Its Applications*, **1971**, 16, 264–280.

52.S. Geisser, *Predictive inference: an introduction*, Chapman & Hall, New York, **1993**.

53.P. A. Devijver and J. Kittler, *Pattern recognition: a statistical approach*, Prentice/Hall International, Englewood Cliffs, N.J, **1982**.

54.S. Kozuch and J. M. L. Martin, *Physical Chemistry Chemical Physics*, **2011**, 13, 20104.

55.L. Goerigk and S. Grimme, *Physical Chemistry Chemical Physics*, **2011**, 13, 6670–6688.

56.S. Grimme, *The Journal of Chemical Physics*, **2006**, 124, 034108.

57.N. Mardirossian and M. Head-Gordon, *The Journal of Chemical Physics*, **2016**, 144, 214110.

58.S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, **2011**, 32, 1456–1465.

59.Y. Zhao and D. G. Truhlar, *The Journal of Physical Chemistry A*, **2005**, 109, 5656–5667.

60.C. Adamo and V. Barone, *The Journal of Chemical Physics*, **1999**, 110, 6158–6170.

61.A. V. Krukau, O. A. Vydrov, A. F. Izmaylov and G. E. Scuseria, *The Journal of Chemical Physics*, **2006**, 125, 224106.

62.T. M. Henderson, B. G. Janesko and G. E. Scuseria, *The Journal of Chemical Physics*, **2008**, 128, 194105.

63.S. H. Vosko, L. Wilk and M. Nusair, *Canadian Journal of Physics*, **1980**, 58, 1200–1211.

64.A. D. Becke, *The Journal of Chemical Physics*, **1993**, 98, 5648–5652.

65.P. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal Of Physical Chemistry*, **1994**, 98, 11623–11627.

66.N. Mardirossian, L. R. Pestana, J. C. Womack, C.-K. Skylaris, T. Head-Gordon and M. Head-Gordon, *The Journal of Physical Chemistry Letters*, **2016**, 8, 35–40.

67.R. Sabatini, T. Gorni and S. de Gironcoli, *Physical Review B*, **2013**, 87, 041108.

68.N. Mardirossian and M. Head-Gordon, *Journal of Chemical Physics*, **2015**, 142, 074111–32.

69.O. A. Vydrov and T. van Voorhis, *The Journal of Chemical Physics*, **2010**, 133, 244103.

70.M. H.-G. N. Mardirossian, *Physical Chemistry Chemical Physics*, **2014**, 9904–9924.

71.Y. Zhang and W. Yang, *Physical Review Letters*, **1997**, 80, 890–890.

72.F. A. Hamprecht, A. J. Cohen, D. J. Tozer and N. C. Handy, *The Journal of Chemical Physics*, **1998**, 109, 6264–6271.

73.A. D. Boese and J. M. L. Martin, *The Journal of Chemical Physics*, **2004**, 121, 3405–3416.

74.S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, **2010**, 132, 154104.

75.J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin and J. Sun, *Physical Review Letters*, **2009**, 103, 026403.

76.J. P. Perdew, in *Electronic Structure of Solids '91*, **1991**, vol. 11.

77.J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, *Physical Review B*, **1992**, 46, 6671–6687.

78.A. D. Boese and N. C. Handy, *The Journal of Chemical Physics*, **2002**, 116, 9559–9569.

79.V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *The Journal of Chemical Physics*, **2002**, 119, 12129–12137.

80.J. P. Perdew, *Physical Review B*, **1986**, 33, 8822–8824.

81.P. J. Wilson, T. J. Bradley and D. J. Tozer, *The Journal of Chemical Physics*, **2001**, 115, 9233–9242.

82.R. Peverati and D. G. Truhlar, *The Journal of Chemical Physics*, **2011**, 135, 191102.

83.B. Hammer, L. B. Hansen and J. K. Norskov, *Physical Review B*, **1999**, 59, 7413–7421.

84.A. D. Boese and N. C. Handy, *The Journal of Chemical Physics*, **2000**, 114, 5497–5503.

85.N. C. Handy and A. J. Cohen, *Molecular Physics*, **2001**, 99, 403–412.

86.J. C. Slater, *Physical Review*, **1951**, 81, 385–390.

87.D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society*, **1928**, 24, 89–110.

88.V. Fock, *Zeitschrift für Physik*, **1930**, 61, 126–148.

89.Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kus, A. Landau, J. Liu, E. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C. M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. J. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T. C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S. P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. ONeill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stuck, Y. C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z. Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C. P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeldf, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylovb, P. M. W. Gill and M. Head-Gordon, *Molecular Physics*, **2015**, 113, 184–215.

90.M. Ernzerhof and G. E. Scuseria, *The Journal of Chemical Physics*, **1999**, 110, 5029–5036.

91.P. Pernot and A. Savin, *The Journal of Chemical Physics*, **2018**, 148, 241707.

92.P. Pernot and A. Savin, *The Journal of Chemical Physics*, **2020**, 152, 164108.