Received: 7 September 2021 | Revised: 10 November 2021 | Accepted: 11 November 2021

DOI: 10.1002/mrc.5234

REVIEW

WILEY

Prediction of chemical shift in NMR: A review

Eric Jonas[1] | **Stefan Kuhn**[2,3] | **Nils Schlörer**[4]

[1]Department of Computer Science, University of Chicago, Chicago, Illinois, 60637, USA

[2]Cyber Technology Institute, De Montfort University, Leicester, LE1 9BH, UK

[3]Institute of Computer Science, University of Tartu, Tartu, Estonia

[4]NMR Core facility, Department of Chemistry, University of Cologne, Cologne, D-50939, Germany

Correspondence
Eric Jonas, Department of Computer Science, University of Chicago, Chicago, IL 60637, USA.
Email: ericj@uchicago.edu

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: 441958208, SCHL580/3-2

Abstract

Calculation of solution-state NMR parameters, including chemical shift values and scalar coupling constants, is often a crucial step for unambiguous structure assignment. Data-driven (sometimes called *empirical*) methods leverage databases of known parameter values to estimate parameters for unknown or novel molecules. This is in contrast to popular ab initio techniques that use detailed quantum computational chemistry calculations to arrive at parameter estimates. Data-driven methods have the potential to be considerably faster than ab inito techniques and have been the subject of renewed interest over the past decade with the rise of high-quality databases of NMR parameters and novel machine learning methods. Here, we review these methods, their strengths and pitfalls, and the databases they are built on.

KEYWORDS

chemical shift prediction, graph neural network, machine learning, NMR

1 | INTRODUCTION

Unambiguous structure assignment of complex organic molecule frequently relies on computational methods for calculation of the parameters of NMR spin systems. A typical example is a computer-aided structure elucidation system that generates candidate structures. In a second step, it then calculates, for example, chemical shift values and compares those to measured data to rank the generated structures.[1] Clearly, the correctness of the calculation is crucial for the results of the overall structure elucidation. Although ab initio techniques such as density functional theory (DFT) remain the gold standard, there has long been an interest in data-driven (*empirical*) methods that leverage existing, experimentally measured NMR parameters to predict new parameters. Such methods have the potential to overcome some of the challenges in ab initio techniques.

Here, we review both classical and more recent data-driven techniques for chemical shift prediction of small organic molecules in solution, starting from publications in 1959 up to summer 2021. We focus on historically relevant and open-source solutions—numerous commercial endeavors exist and are popular. Their exact algorithms are often not published. We mention commercial software where we think it is relevant, but do not aim for a full product evaluation—rather, we wish to give an overview of the substantial progress made and highlight opportunities for future improvement. The latest review to be published was Cobas,[2] but it has the prediction only as one of its topics. Earlier reviews were Toukach and Ananikov[3] (focusing on carbohydrates) and Lodewyk et al.[4]

In Section 2, we discuss some fundamentals of chemical shift prediction, including underlying assumptions, metrics, and datasets used. In Section 3, we review the various methods available and reference the most significant literature. The focus here is on ^{13}C and ^{1}H chemical shift prediction, but we also report work on other nuclei and on coupling constant prediction. In Section 3.2, we examine the error rates for ^{13}C and ^{1}H prediction, which were given in the literature. From this, we extract long-term trends, which we report in Section 4.

2 | FUNDAMENTALS

2.1 | Assumptions

In this review, we focus on solution-state NMR, ignoring recent work in solid-state prediction. All of the methods we mention assume a single molecule as the basis for the prediction and do not include interaction with other molecules. For solution-state NMR, this is a justified assumption and is an underlying simplication in all methods. We also restrict ourselves to small organic molecules. Therefore, proteins and other biomacromolecules are not considered. This is justified because the NMR interpretation as well as the prediction used there is very much different from small molecules (see Cavalli et al[5] for an overview). We also do not consider inorganic molecules, because the methods described cannot offer this.

Small organic molecules are most commonly represented as a graph of atoms, defined by their atom symbol and formal charge, and bonds, having a bond order. There exist several string-based canonicalizations of this representation including SMILES and InChI that are often used as input formats. These molecular graphs fundamentally convey topological information. This representation is of course not the same as the physical three-dimensional molecule that is measured in a spectrometer. Some methods do a prediction directly from the graph (most prominently HOSE [hierarchically ordered spherical description of environment] codes). This assumes that there is still enough chemical meaning in the graph to get a good approximation for the chemical shifts. Some stereochemical information is traditionally given by so-called wedge bonds and E/Z depiction. If those are given in the molecules and the method uses them, that can give a more specific prediction (e.g., the extended HOSE code in Kuhn and Johnson[6]).

A next level of information is to have a three-dimensional input. If the method encodes such information, the prediction can be even better. IMPRESSION,[7] for example, works with three-dimensional input. This will potentially give a very good result for a particular geometry, but it requires a geometry. In practice, this is not necessarily easy to determine and might not be the measured geometry.

Furthermore, for chemical shift prediction, it is assumed that peaks in a spectrum are clearly defined and can be determined and assigned. For multiplets, the center frequency is reported and predicted. In order to simulate the "real" peak, the multiplicity and the coupling constants must be taken into account as well.

2.2 | Metrics

Evaluation metrics generally involve computing the difference between predicted NMR parameter values and experimentally observed values. Metrics normally used are the mean squared error (MSE) or the mean absolute error (MAE). Computing the MAE for all values in a molecule, and then averaging across all molecules in the dataset, is fundamentally different from computing the MAE across all predicted values.

When evaluating methods' performance, care must be taken to select a representative sample of molecular structures. The molecules in this *test set* should be different from those used to train the model or fit model parameters (such as the linear scaling coefficients for DFT), and they should reflect the diversity of structures that the method will be used for. In case of data-driven methods, cross-validation is an alternative. For this, the overall data are split in a number of batches (five or 10 are typical values), and in iterations, one of them is used for testing, the rest for training. The error is then averages over all runs.

Sometimes (e.g., for increment-based methods) not all molecules from a test set are used. In contrast, methods like a neural network will always return a value, even if it is not very reliable in some cases. That makes comparing the results difficult.

For data-driven methods, the choice of the training data influences the result as well. Even if cross-validation is used (and therefore the problem of choice of test data is eliminated), the training data can still be problematic, depending on the method. For example, classes of structures that are represented only by very few example in contrast to the rest may not influence the training much, and therefore not be predicted well, even though similar examples are part of the training set.

Additionally, reporting average error for machine learning methods can give a false sense of performance, as many methods perform well when test molecules are similar to molecules in the training set. However, NMR is often used for novel molecules like natural products whose structure may differ considerably from the training set. Thus, even though some papers report performance exceeding DFT,[8] it is uncertain whether this advantage will hold on truly novel structures. One way of capturing this difference is to compare methods based on percentiles: What is the average error of (say) the worst n %-performing parameters?

2.3 | Datasets

The sensitivity of data-driven methods to their training database is well-documented in other areas of data

science and machine learning.[9] Unlike many other areas, the datasets available for data-driven NMR methods are often quite limited. Issues of chemical diversity, experimental control, and machine readability plague many datasets of NMR measurements. The FAIR standard (Findability, Accessibility, Interoperability, and Reuse)[10] defines a common understanding of what requirements data sources should fulfill. Perhaps, the largest publicly available database of NMR data is nmrshiftdb2,[11] a user-contributed database of over 44,000 molecules and 53,000 spectra (as of 23/5/2021). A purely user-contributed database has both strengths and weaknesses. Allowing users to upload their spectra increases participation and results in a larger database, but at the potential expense of control, spectra may be measured in differing conditions and crucial metadata (such as solvent) is omitted. nmrshiftdb2, while being user-contributed, is curated. In contrast, SDBS (maintained by AIST in Japan,[12] over 34,000 molecules and over 30,000 NMR spectra) and GISSMO[13] from the BMRB at Madison only allow data that comply with certain criteria. Unfortunately, SDBS is not machine readable, and it is forbidden to download the entire database. GISSMO is small, focused on proton spectra, and mostly consists of known metabolites.

Another option is to use a dataset generated by ab initio techniques. This makes it possible to generate large datasets computationally. It also guarantees uniform data quality. On the other hand, the data may suffer from limitations of the specific technique used (see Section 3.1.1 for a discussion), and bias in the selection of structures is still an issue. Overall, generated data are an option here, but not necessarily a perfect solution.

3 | METHODS AND RESULTS

3.1 | Methods

The methods used for NMR prediction can be divided into the following categories:

3.1.1 | Ab initio techniques

Ab initio techniques, especially DFT, have been the gold standard for the calculation of the parameters of NMR spin systems. We take an expansive view of the term ab initio, including not just pure techniques like Hartree–Fock SCF but also various DFT techniques where the functionals are fit to data. The rise of gauge-independent atomic orbitals (GIAOs) allowed for removal of the gauge problem and have allowed chemical shift prediction to become turnkey,[14,15] with support from popular packages like GAUSSIAN[16] and NWChem.

Briefly, calculation of NMR parameters via these techniques involves three steps: First, a conformational search is necessary, usually performed via Monte Carlo sampling techniques, dedicated low-mode search approaches, or molecular dynamics dependent on classical force fields. This is necessary as molecular motion is not insignificant in liquids, and a wide variety of conformations may be present in solution. Next, geometry optimization is performed using DFT packages to achieve quantum-mechanically accurate geometry positioning. Note that traditional force field-based geometry search methods can yield inaccurate results and thus Grimme et al[17] combine meta-dynamics with a semi-empirical approach for conformer and rotamer sampling. Finally, the actual NMR parameters of interest, including chemical shifts and scalar coupling constants, are calculated. All of these steps can be quite time consuming due to the at-least-cubic scaling of DFT methods and the need for large basis sets.

Additionally, DFT techniques often can only produce chemical shift values correct up to an affine transform. Tantillo[4] and others have done an excellent job exploring these scaling effects and their impact, which may arise due to conformational searches failing to incorporate quantum mechanical effects.[18] Some recent work[19] has suggested going beyond global linear scaling by having molecular–motif-specific scaling values, improving performance in aqueous solutions.

There are several excellent tutorials[14] and reviews[20] that further delve into ab initio methods. All of the machine learning methods below suffer from difficulties generalizing outside of their training set, and it is difficult to explain a priori where they will fail. In contrast, DFT techniques (and their limitation) are well-understood and can be extended to novel nuclei, very heavy atoms, and radicals.[21] It should be noted that there are many different functionals available in DFT, and calculation performance depends on both the functional and basis set. The results can therefore vary depending on what is used. Ermanis et al[22] discuss the trade-offs between methods, including handling of conformer examples.

Instead of computing chemical shifts directly from quantum mechanics, an alternative approach could be to compute other physical quantities from which the chemical shift can be inferred. Such an approach is used by Abraham and Mobli,[23] who calculate the chemical shift of a hydrogen from its partial atomic charge (which is calculated in turn from inductive and resonance contributions) and long-range effects from functional groups (taking 3D configuration into account). The reported error is 0.28.[24]

3.1.2 | Additive increment-based approaches

An early method to calculate chemical shifts is based on the idea that the shift of an atom can be combined of a basic shift for the atoms, corrected by (positive and negative) terms for the substituents. Using appropriate tables, it should be possible to calculate chemical shift values. The first rules were published by James Nelson Shoolery to calculate shifts of methylene groups.[25] A typical example of this is Paul and Grant,[26] giving such values for linear alkanes. Many more studies were done over time, and got compiled into books like Pretsch et al.[27] (first German edition 1976, latest English edition 2020 under changed title). The name of Ernö Pretsch became closely associated with this type of rules. Considering the large amount of data used on the one hand, and the systematic nature of the rules on the other hand, encoding this into a computer program was done in Fürst and Pretsch.[28] Using the CSEARCH database as a test case, this reports a standard deviation of 5.5 ppm for ^{13}C shift prediction. For ^{1}H shifts, a standard deviation of 0.19 was reported.[19] On the other hand, to illustrate the dependency on test sets, a later testing reports 0.30 using the same method.[24] A version of the Pretsch increments considering 3D configurations is reported to give an error of 0.21.[24] Increment rules have also been used to predict shifts of small classes of molecules, for example, ^{13}C, ^{1}H, and ^{15}N shifts of polysaccharides in the CASPER program.[29] Dubois et al. and Panaye et al[30,31] introduce an approach similar to increments. This can also deal with structures where the substituent is unknown, but the existence of a substituent is established. No overall figures for prediction quality are given.

3.1.3 | HOSE codes and other substructure codes

Moving on from increments, a general system for encoding atomic environments with the aim to predict chemical shifts seems a logical step. This was, to our knowledge, first suggested in Jezl and Dalrymple.[32] Codes were manually written and had limitations; for example, structures were treated as trees, not graphs, including small rings in an ad hoc fashion.

This lead to the development of HOSE[33] codes, which became a gold standard in the field (Bremser et al[34] published an earlier version of HOSE codes). They encode the environment of an atom in a spherical manner, where spheres are defined by bond distance. For NMR prediction, the HOSE codes of the atoms with a shift assigned in some dataset are generated with a certain number of spheres (6 is a commonly used number because of the distance effect in organic molecules). The HOSE code with the same number of spheres is then calculated for the atom for which the prediction is desired, and this code is then searched in the dataset. If it is found, the shift of the atom with that HOSE code is used as a prediction value. If several values are found, the average is typically chosen. If no match is found, the HOSE codes are compared ignoring the last sphere. This process continues down to a minimum number of spheres. Effectively, this is a nearest neighbor search, with similarity encoded by the HOSE code.

HOSE codes can be generated purely from a standard molecular description. They encode all information usually included in those, like atomic symbol, bond order, aromaticity, or formal charges. Their generation is fast, and using tools like relational databases, searching is also fast. No training process is needed. It is very easy to look up which spectra contribute to a prediction, which provides not only explainability, but makes HOSE code also a good tool for error detection.[6]

On the other hand, HOSE codes in their original form do not include stereochemistry. Conformers cannot be distinguished. Because the search is a neighbor search, relatively close examples are needed for a good prediction.

Despite these drawbacks, HOSE codes have proven not only to be an easy to use tool, but also an accurate tool. Cross-validation on large databases has given errors of 0.154[35]/0.29[6] and 1.58[36]/3.52[6] for ^{1}H respectively ^{13}C predictions. The dependence on the underlying database was shown in Meiler et al,[37] where the average error of predictions for a molecule using three different databases were 1.0, 1.7, and 2.7. Overall, HOSE codes were considered to give very good results for at least 30 years after their invention: "Programs utilizing HOSE codes provide similar or better accuracy [than neural networks]"[36] was said in 2009, while HOSE codes were originally published in 1978.

Gray et al[38] introduced a spherical code incorporating stereochemistry together with a database system. It is reported that the range of shifts assigned to the same code within a database of 12,300 shifts is 0.5 (no prediction error is reported). Magri et al[39] use spheres defined by absolute distance in force field-optimized structures, not spheres defined by bond, and codes for atoms. It reports an error of 2.4 ppm for ^{13}C shifts. Xu[40] uses an encoding scheme called generalized atom center fragment (GACF). In particular, these use a varying number of spheres depending on the structure. No error rates are reported.

An improved version of HOSE codes was published in Kuhn and Johnson.[6] This includes an encoding of stereochemistry. In line with conventional HOSE codes, it

includes everything that is depicted in standard structural formulae, via wedge bonds. This improves the results to errors of 0.25/2.82 for $^{1}H/^{13}C$ compared with 0.29/3.52 with standard HOSE codes, using the same dataset. In contrast to Gray et al.,[38] the extended HOSE codes are a superset of standard HOSe codes, and every standard HOSE code is valid extended HOSE code as well (namely, that of a nonstereochemically specified structure).

Spanton and Whittern[41] combine HOSE codes with a stereo descriptor based on IUPACInternational Chemical Identifier (InChI). It further uses a substructure match to select the shift in case there are several. It achieves an MAE of 0.1081. Castillo et al[42] use a HOSE code based system and emphasize the ability to improve the results with the addition of new data.

3.1.4 | Machine learning

Many techniques attempt to learn a parameterized function that approximates a given NMR parameter. This family of techniques often go by the term "machine learning," although the term is itself not precise.[1]

Early work focused on using artificial neural networks (ANNs) to predict shifts.[35,37,43] Those were mostly feed-forward networks, where the problem is described by a number of parameters which are fed into the network. The network is given a number of example outputs and uses those to correlate the input and the outputs. Later, the descriptors for a case can be fed into the network and the output (the prediction) calculated. Those methods have two major factors that can be tuned to achieve the result: Firstly, the type and architecture of the network (number of layers etc.) and which descriptors to use. The latter is perhaps the most crucial factor and has received much attention. The descriptors include simple atom properties (e.g., element symbol and charge), calculated properties for single atoms (e.g., partial charge), or complex descriptors involving neighboring atoms and three-dimensional configurations (e.g., radial distribution functions). The use of descriptors together with the neural networks should make the method capable of predicting molecules not well represented in the training data.

A simple machine learning technique is linear regression, which calculates a final value from a linear combinations of other values. This was used to calculate chemical shifts from atom descriptors in Jurs et al.[44,45] The authors had published a number of models for various classes of atoms before and combined these into a universal prediction program. Errors of about 1 ppm are reported for selected classes of molecules, but an exhaustive evaluation is not available. The descriptors used are similar to those mentioned for the neural networks. This distinguishes this work from increment rules, which can be considered a type of linear regression, but encode functional groups and atoms directly.

Apart from neural network, machine learning offers a wide range of other methods that try to correlate inputs (descriptors) to outputs (shifts) and can potentially be used for predictions. Some of them, like decision trees or support vector machines, can achieve similar results to neural networks. A range of methods has been tested in Kuhn et al.[35] (including linear regression mentioned above). Blinov et al.[46] combine a custom encoding with partial least squares regression to achieve a precision for ^{13}C prediction roughly in line with other AI methods, but the prediction was faster than comparable methods at that time.

There has been considerable interest in applying recent techniques from deep learning to chemical shift prediction, especially graph neural networks (GNNs).[8] GNNs learn a collection of local features centered at each atom, providing a learned "fingerprint" for each atomic neighborhood. These fingerprints are then used to compute chemical shift values and other parameters and optimized against a loss or error metric-like MSE or MAE. The first application of GNNs to chemical shift prediction[8] provided a confidence estimate for each predicted shift, allowing the user to incorporate certainty into downstream tasks. The authors found that they were able to match (and occasionally exceed) DFT-level performance. This project was trained for both ^{13}C and ^{1}H on NMRShiftDB and relied heavily on the availability of successful peak assignment. The ability to learn to predict without assigning an observed shift to a given nucleus[47] is potentially powerful and could resolve numerous dataset limitations outlined in Section 2.3. Han and Choi[48] use transfer learning with a large simulation database to be able to train models with small numbers of experimental data. It achieves comparable results as other deep learning approaches with a much smaller dataset. Guan et al[49] and Yang et al[50] use GNNs and 3D structures, achieving comparable results (Figure 1).

3.2 | Results for ^{13}C and ^{1}H prediction

In order to get a more systematic overview of the results achieved for prediction of ^{13}C and ^{1}H NMR spectra, we compiled some figures into Table 1. This gives an

[1]For example, HOSE codes can be viewed as an instance of a machine learning technique called "nearest neighbors," where a test molecule's predictions are computed by comparing with the most similar molecules in the training set.

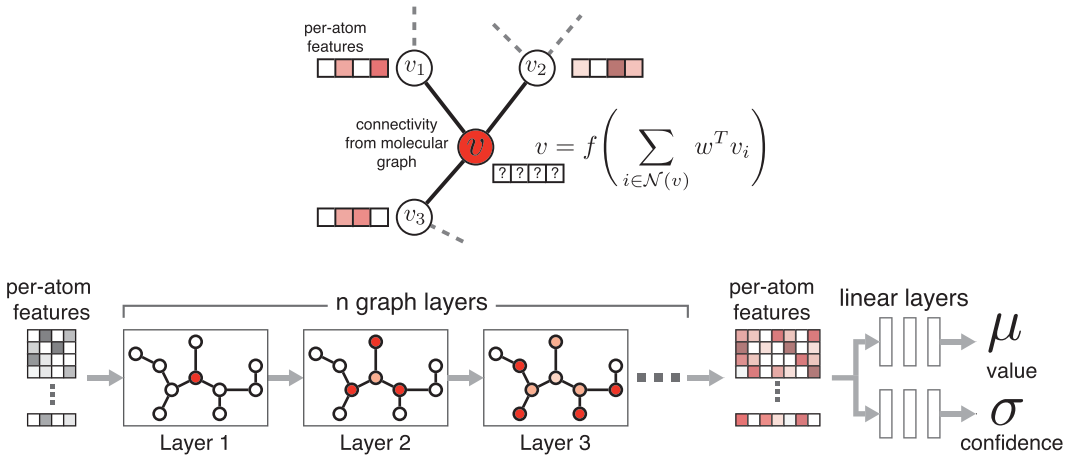

FIGURE 1 Graph convolutional networks (Section 3.1.4) learn local environmental features over larger and larger sections of the graph and then enable per-atom predictions. In Jonas and Kuhn,[8] these per-atom features were used to calculate both mean shift values as well as a confidence score. Figure adapted from Jonas and Kuhn[8]

TABLE 1 Historically achieved MAEs for various methods and datasets

Date	^1H MAE (ppm)	^{13}C MAE (ppm)	Method	Training	Test dataset	Literature
1/1990	—	5.5[b]	Pretsch increments	C13Shift	CSEARCH	[28]
4/1992	—	1.56	Optimized Increments	OPSI	3 organic molecules	[51]
5/1994	0.19[b]	—	Pretsch increments	Custom	200 molecules	[19]
12/2000	—	2.4	Classic neural network	57 sesquiterpene lactones	11 organic molecules	[39]
1/2002	0.25	—	Classic neural network	120 organic molecules	259 organic molecules	[43]
8/2002	—	1.2	Classic neural network	SpecInfo	Taxol	[37]
8/2002	—	2.4	Classic neural network	SpecInfo	Taxol	[37]
8/2002	—	3.8	Pretsch increments	ChemDraw Pro	Taxol	[37]
8/2002	—	3.7	Pretsch increments	SpecTool	Taxol	[37]
8/2002	—	1.0	HOSE	Specinfo	Taxol	[37]
8/2002	—	2.7	HOSE	PredictIt	Taxol	[37]
8/2002	—	1.7	HOSE	CNMR 6.0	Taxol	[37]
8/2002	—	3.4	Cosmas 4.5	n/a	Taxol	[37]
8/2002	—	4.0	Gaussian 98	n/a	Taxol	[37]
8/2002	—	1.9	Classic neural network	SpecInfo	1547 shifts/100 molecules	[37]
8/2002	—	2.5	Classic neural network	SpecInfo	1547 shifts/100 molecules	[37]
8/2002	—	1.4	SpecInfo	SpecInfo	1547 shifts/100 molecules	[37]
5/2007	—	1.59	ACD/Labs predictor	ACD/Labs	NMRShiftDB	[52]
5/2007	—	2.22	CSEARCH	CSEARCH	NMRShiftDB	[52]
1/2008	0.28	—	CHARGE	n/a	Wiley ^1H NMR database	[24]

TABLE 1 (Continued)

Date	^{1}H MAE (ppm)	^{13}C MAE (ppm)	Method	Training	Test dataset	Literature
1/2008	0.30	—	Pretsch increments	n/a	Wiley ^{1}H NMR database	[24]
1/2008	0.21	—	Pretsch 3D increments	n/a	Wiley ^{1}H NMR database	[24]
1/2008	0.18	—	Combined	n/a	Wiley ^{1}H NMR database	[24]
9/2008	0.154	—	HOSE	NMRShiftDB	[a]	[35]
9/2008	0.18	—	J48	NMRShiftDB	[a]	[35]
9/2008	0.182	—	RF	NMRShiftDB	[a]	[35]
9/2008	0.125	—	SVM	NMRShiftDB	[a]	[35]
5/2009	—	1.85	Custom encoding partial least squares regression	190,000 structures	16,000 structures	[35]
1/2010	—	1.58	HOSE	ACD/Labs	205 molecules/2531 shifts	[36]
1/2010	0.1081	—	Improved HOSE	in house	282	[41]
1/2010	—	1.91	NN	ACD/Labs	205 molecules/2531 shifts	[36]
1/2010	—	2.15	Increment	n/a	205 molecules/2531 shifts	[36]
1/2010	—	3.29	DFT-GIAO	n/a	205 molecules/2531 shifts	[36]
4/2019	0.25	2.82	Stereo HOSE	nmrshiftdb2	[a]	[6]
4/2019	0.29	3.52	HOSE	nmrshiftdb2	[a]	[6]
5/2019	—	1.63	Modgraph (HOSE + NN)	Modgraph	13 molecules	[53]
5/2019	—	2.2	ML	MestreLab	13 molecules	[53]
5/2019	—	1.7	ML	MestreLab	13 molecules	[53]
5/2019	—	1.3	Ensemble of previous three	As per method	13 molecules	[53]
8/2019	0.28	1.43	CNN	Subset of nmrshiftdb2	[a]	[8]
11/2019	0.23	2.45	ML	DFT calculations	410 structures	[7]
4/2020	0.224	1.355	MPNN	Subset of nmrshiftdb2	[a]	[54]
2/2021	0.11[c]	0.70[c]	Δ-machine learning	57,456 DFT calculations	3780 DFT calculations	[55]

Abbreviations: CNN, convolutional neural network; DFT-GIAO, density functional theory-gauge-independent atomic orbitals; HOSE, hierarchically ordered spherical description of environment; MAE, mean absolute error; ML, machine learning; MPNN, message passing neural network.

[a]Employs cross-validation with identical dataset.

[b]Standard deviation instead of mean absolute error, error excludes atoms that are considered impossible to predict.

[c]Root mean squared error instead of mean absolute error.

overview of the development of the precision of predictions over time. The publications were chosen on the one hand to give a number for the prediction error as mean average error/mean deviation and on the other hand to represent a significant method, give a number for a method for the first time, or several comparable numbers. For several reasons, this table needs to be read with care (mentioned in Section 2), but we believe it still constitutes a valuable overview. In the table, "method" gives an indication of the method used, and "training" indicates either a training set used if the method is data-driven or a specific software implementing the method. n/a means that is not a data-driven method and nothing more specific than in the method column is reported.

"Test dataset" indicates the test dataset used. In all cases, the literature indicated in the "literature" column should be consulted for details. Some general issues worth noting are as follows:

- Some methods are represented several times, which does not mean they are identical. This is clear for neural networks (classic or CNN), where the architectures can be vastly different, but also, for example, HOSE code implementations tend to be vary the original publication. For increments, there are many different sets of rules and values.
- Data may vary over time. For example, NMRShiftDB in 2007 had different data from 2008.
- Not all publications are the original publication for a method, because they do not necessarily contain quantitative information (e.g., the original HOSE code paper[33] does not quantitatively evaluate the method).
- Not all publications contain ^{13}C and ^{1}H predictions.

In order to get an overview of those numbers, we have put them into a line chart in Figure 2. The caveats given in the previous section apply here as well, and in particular, the ordering by time is partly arbitrary here. Firstly, as said before, the publications are not necessarily the first use of the method, and secondly, numbers from the same publication are next to each other. Still, it seems plausible to argue that after some initial progress from early increment systems, improvements were not fundamental for a long time. In particular, HOSE codes and machine learning methods were performing roughly similar, as well as different ML methods not clearly outperforming each other. Only recently some promising results have been reported in Jonas and Kuhn,[8] Kwon et al,[54] and Unzueta et al.[55] In particular, the results from Unzueta et al[55] are better than anything reported so far. We would cautiously conclude that new machine learning methods, like CNNs, MPNNs, or δ machine learning have the potential for a breakthrough effect, even though a final verdict is hard due to mentioned factors.

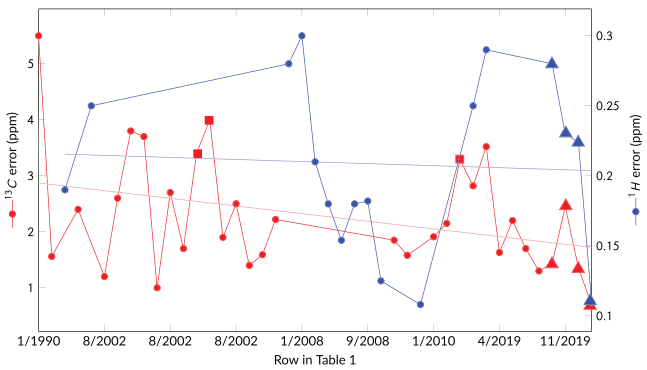

FIGURE 2 Selected prediction results for ^{1}H and ^{13}C NMR predictions from 1990 until now, ordered by publication time. The light lines give the least squares linear regression. Squares indicate ab initio calculations, triangles deep learning methods. See the text for an explanation and caveats to consider

3.3 | Coupling constants and other nuclei

In Section 3.2, we reported results for ^{13}C and ^{1}H prediction. Work has also been done on other nuclei, but is much more limited. Reasons for this are the generally lower importance of the methods and therefore less interest in them, and the lack of data, which is in particular problematic for data-driven methods. Furthermore, reliable error estimations are rare for other nuclei. Many work reports on highly specialized subclasses of compounds, which makes comparison difficult. We therefore report a few results here, without giving a history as for ^{13}C and ^{1}H:

- Unzueta et al[55] do ^{17}O and ^{15}N predictions with an RMS of 1.69 and 2.47, respectively. The method used is the same as for $^{13}C/^{1}H$ shifts, using DFT for generating the data in this data-driven approach.
- DFT methods have been used for various nuclei, resulting in, for example, $^{15}N/^{17}O$ predictions with MAEs of 3/36.4,[56] in an MAE of 5.1 for ^{15}N,[57] or that "95% of linear scaled N-15 chemical shifts are within a ±9.56 range".[58]
- Commercial software by ACD/Labs predicts ^{15}N, ^{19}F, and ^{31}P shifts[59] and MestreLab software predicts ^{11}B, ^{15}N,^{17}O, ^{19}F, ^{29}Si, and ^{31}P shifts.[60]
- For ^{19}F shifts, Saunders et al[61] achieved a maximum deviation of 6.5 ppm for fluorinated (hetero)aromatic compounds. Vulpetti et al[62] report a root mean squared error (RMSE) of 1.08 for a CF3 test set and 2.37 for a CF test set, but does not offer a prediction for all cases.
- Emerenciano et al[63] predict ^{125}Te and ^{13}C shifts for diorgano tellurides. A MAE of 0.67 is reported for ^{13}C, which is a very good result, but only applies to a specialized class of compounds. This program matches atom types and distances in 3D models in order to find similar instances.

As we have seen in Section 2, for a spectrum, simulation coupling constants need to be predicted as well. This is done for HH couplings using associative neural networks in Binev et al.[64] Here, the networks are used to find most similar atoms, and their coupling constants are

used as prediction. The reported error is 0.6–0.8 Hz. Shibata and Kaneko[65] use RDKit descriptors and decision trees via LightGBM and achieves errors from 0.880 to 0.99 for ^{1}H, ^{2}H, and ^{3}H HH, CH, and NH couplings. Ito et al[66] use a combination of DFT and ML and report an error of 1.21 Hz for HH couplings.

If peaks and coupling constants are predicted, they can be combined into a realistic spectrum simulations. This is in particular relevant for ^{1}H spectra. An early work on this is Tusar et al,[67] which uses increments for the prediction. A current spectrum simulation systems is available, for example in SPINUS,[68] based on neural network prediction.

4 | CONCLUSION

Methods for chemical shift prediction can be classified as either ab initio or data-driven methods. Ab initio techniques like DFT are able to give very good results and remain the gold standard. Data-driven methods, on the other hand, have for most of the time failed to achieve comparable results. Still, their comparatively fast execution still makes them attractive. They can be divided into three broad, groups, namely, increment-based, HOSE code, and machine learning approaches. All of them have been used for ^{13}C and ^{1}H prediction. Usage for other nuclei has been demonstrated as well, but is limited by availability of data.

Even though numbers have to be interpreted with caution for the reasons given, it seems that overall HOSE codes and machine learning now perform comparably. This was concluded in Elyashberg et al[36] and is supported by our findings. For example, the best value reported for a ^{1}H prediction is 0.1081 ppm for a HOSE code-based method. That beats machine learning (0.125), but is better than results reported for HOSE codes with bigger databases (0.25 achieved using nmrshiftdb2). The best result for increments is 0.21 (0.19 is a standard deviation instead of MAE), suggesting that increments may not do as well as machine learning or HOSE codes. As explained above, inconsistencies in datasets and training data make it difficult to make more definitive statements.

Looking at the development over the last three decades given in Table 1 and Figure 2, it is not possible to spot a clear trend, in particular when leaving out the last four items, which are deep learning based. Even though various improvements were presented, a consistent and clear tendency of the error going down is not visible.

Deep learning technologies, which have become prominent recently, claim to achieve comparable results to DFT. If this is true, it should be reflected in the error numbers. Although an exact proof of this is impossible, our numbers, as given in Table 1 and Figure 2, show that, at least for ^{13}C predictions, there was a significant improvement recently, which goes well beyond what was possible so far. This is demonstrated by the red line in Figure 2 showing a trend downwards on the right, and being well below historical values. On the other hand, this is not visible for ^{1}H predictions, even though the same technologies have been applied. There could be an argument for leaving out the prediction of 0.1081, which might partly be due to the dataset used, but even then other historic values are very close to the latest results. The challenges of robust, consistent ^{1}H measurements, and the sensitivity of ^{1}H spectra to temperature, solvent, and other experimental factors make further improvement in performance difficult given the quality of data.

ACKNOWLEDGEMENTS

N.E.S. thanks for funding received by the Deutsche Forschungsgemeinschaft (DFG) during the IDNMR project (grant SCHL580/3-2) and also for support through the NFDI chemistry research data initiative, NFDI4Chem. NFDI4Chem is supported by DFG under project number 441958208.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/mrc.5234.

ORCID

Stefan Kuhn https://orcid.org/0000-0002-5990-4157
Nils Schlörer https://orcid.org/0000-0002-0990-9582

REFERENCES

[1] M. Elyashberg, D. Argyropoulos, *Magn. Reson. Chem.* **2021**, *59*(7), 669. https://doi.org/10.1002/mrc.5115
[2] C. Cobas, *Magn. Reson. Chem.* **2020**, *58*(6), 512. https://doi.org/10.1002/mrc.4989
[3] F. V. Toukach, V. P. Ananikov, *Chem. Soc. Rev.* **2013**, *42*, 8376. https://doi.org/10.1039/C3CS60073D
[4] M. W. Lodewyk, M. R. Siebert, D. J. Tantillo, *Chem. Rev.* **2012**, *112*(3), 1839.
[5] A. Cavalli, X. Salvatella, C. M. Dobson, M. Vendruscolo, *Proc. Natl. Acad. Sci.* **2007**, *104*(23), 9615. https://www.pnas.org/content/104/23/9615
[6] S. Kuhn, S. R. Johnson, *ACS Omega* **2019**, *4*(4), 7323.
[7] W. Gerrard, L. A. Bratholm, M. J. Packer, A. J. Mulholland, D. R. Glowacki, C. P. Butts, *Chem. Sci.* **2020**, *11*, 508. https://doi.org/10.1039/C9SC03854J
[8] E. Jonas, S. Kuhn, *J. Cheminformatics* **2019**, *11*(1), 50.
[9] V. Sessions, M. Valtorta, The effects of data quality on machine learning algorithms, Iciq, **2006**.
[10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo,

R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S. A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, *Sci Data* **2016**, *3*, 160018.

[11] S. Kuhn, N. E. Schlörer, H. Kolshorn, R. Stoll, *J. Cheminformatics* **2012**, *4*(1), P52. https://doi.org/10.1186/1758-2946-4-S1-P52

[12] National Institute of Advanced Industrial Science and Technology, Sdbsweb, https://sdbs.db.aist.go.jp, **2021**.

[13] H. Dashti, J. R. Wedell, W. M. Westler, M. Tonelli, D. Aceti, G. K. Amarasinghe, J. L. Markley, H. R. Eghbalnia, *Anal. Chem.* **2018**, *90*(18), 10,646. PMID: 30125102.

[14] P. H. Willoughby, M. J. Jansma, T. R. Hoye, *Nat. Protoc.* **2014**, *9*(3), 643. http://www.ncbi.nlm.nih.gov/pubmed/24556787

[15] Y. Yesiltepe, J. R. Nuñez, S. M. Colby, D. G. Thomas, M. I. Borkum, P. N. Reardon, N. M. Washton, T. O. Metz, J. G. Teeguarden, N. Govind, R. S. Renslow, *J. Cheminformatics* **2018**, *10*, 1. https://doi.org/10.1186/s13321-018-0305-8

[16] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16, Revision A.03, Gaussian, Inc., Wallingford CT, **2016**.

[17] S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, F. Neese, *Angew. Chem. Int. Ed.* **2017**, *56*(46), 14,763. https://doi.org/10.1002/anie.201708266

[18] E. E. Kwan, R. Y. Liu, *J. Chem. Theory Comput.* **2015**, *11*(11), 5083. https://doi.org/10.1021/acs.jctc.5b00856

[19] R. Bürgin Schaller, E. Pretsch, *Analytica Chimica Acta* **1994**, *290*(3), 295. https://www.sciencedirect.com/science/article/pii/0003267094801169

[20] K. Jackson, R. Paton, *in Applied theoretical organic chemistry*, (Eds: D. J. Tantillo), World Scientific **2018**, 165.

[21] F. Rastrelli, A. Bagno, *Chem.–Eur. J.* **2009**, *15*(32), 7990.

[22] K. Ermanis, K. E. B. Parkes, T. Agback, J. M. Goodman, *Org. Biomol. Chem.* **2019**, *17*, 5886. https://doi.org/10.1039/c9ob00840c

[23] R. J. Abraham, M. Mobli, *Spectrosc. Eur.* **2004**, *16*(4), 16.

[24] A. N. Davies, *Spectrosc. Eur.* **2008**, *20*(1), 20.

[25] J. N. Shoolery, *Varian Tech. Inf. Bull.* **1959**, *3*(3).

[26] E. G. Paul, D. M. Grant, *J. Am. Chem. Soc.* **1963**, *85*(11), 1701. https://doi.org/10.1021/ja00894a045

[27] E. Pretsch, T. Clerc, J. Seibl, W. Simon, *Tables of Spectral Data for Structure Determination of Organic Compounds*, Springer-Verlag, Berlin Heidelberg **1983**.

[28] A. Fürst, E. Pretsch, *Analytica Chimica Acta* **1990**, *229*, 17. https://www.sciencedirect.com/science/article/pii/S0003267000851053

[29] J. Støohle, G. Widmalm, *in NME in Glycoscience and Glycotechnology*, The Royal Society of Chemistry **2017**, 335–352. https://doi.org/10.1039/9781782623946-00335

[30] J. Dubois, M. Carabedian, I. Dagane, *Anal. Chim. Acta* **1984**, *158*, 217.

[31] A. Panaye, J. P. Doucet, B. T. Fan, *J. Chem. Inf. Comput. Sci.* **1993**, *33*(2), 258. https://doi.org/10.1021/ci00012a012

[32] B. A. Jezl, D. L. Dalrymple, *Anal. Chem.* **1975**, *47*(2), 203. https://doi.org/10.1021/ac60352a047

[33] W. Bremser, *Anal. Chim. Acta* **1978**, *103*(4), 355.

[34] W. Bremser, M. Klier, E. Meyer, *Org. Magn. Reson.* **1975**, *7*(2), 97. https://doi.org/10.1002/mrc.1270070211

[35] S. Kuhn, B. A. Egert, S. Neumann, C. Steinbeck, *BMC Bioinformatics* **2008**, *9*(1), 400. https://doi.org/10.1186/1471-2105-9-400

[36] M. Elyashberg, K. Blinov, Y. Smurnyy, T. Churanova, A. Williams, *Magn. Reson. Chem.* **2010**, *48*(3), 219. https://doi.org/10.1002/mrc.2571

[37] J. Meiler, W. Maier, M. Will, R. Meusinger, *J. Magn. Reson.* **2002**, *157*(2), 242. https://www.sciencedirect.com/science/article/pii/S1090780702925991

[38] N. A. B. Gray, J. G. Nourse, C. W. Crandell, D. H. Smith, C. Djerassi, *Org. Magn. Reson.* **1981**, *15*(4), 375. https://doi.org/10.1002/mrc.1270150412

[39] F. A. M. M. Magri, J. A. S. L. MilitÃ£o, M. J. P. Ferreira, A. A. J. C. Brant, V. P. Emerenciano, *Spectroscopy* **2001**, *15*, 99. 2.

[40] J. Xu, *Molecules* **1997**, *2*(8), 114. https://www.mdpi.com/1420-3049/2/8/114

[41] S. G. Spanton, D. Whittern, *Magn. Reson. Chem.* **2009**, *47*(12), 1055. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/mrc.2512

[42] A. M. Castillo, A. Bernal, R. Dieden, L. Patiny, J. Wist, *J. Cheminformatics* **2016**, *8*(1), 26. https://doi.org/10.1186/s13321-016-0134-6

[43] J. Aires-de Sousa, M. C. Hemmer, J. Gasteiger, *Anal. Chem.* **2002**, *74*(1), 80.

[44] P. C. Jurs, J. W. Ball, L. S. Anker, T. L. Friedman, *J. Chem. Inf. Comput. Sci.* **1992**, *32*(4), 272. https://doi.org/10.1021/ci00008a002

[45] G. W. Small, P. C. Jurs, *Anal. Chem.* **1983**, *55*(7), 1121. https://doi.org/10.1021/ac00258a033

[46] K. A. Blinov, Y. D. Smurnyy, T. S. Churanova, M. E. Elyashberg, A. J. Williams, *Chemometrics and Intelligent Laboratory Systems* **2009**, *97*(1), 91. https://www.sciencedirect.com/science/article/pii/S0169743909000112, Selected papers presented at the 6th Winter Symposium on Chemometrics Kazan, Russia 16-22 February 2008.

[47] S. Kang, Y. Kwon, D. Lee, Y.-S. Choi, *J. Chem. Inf. Model.* **2020**, *60*(8), 3765. PMID: 32692561.

[48] H. Han, S. Choi, *J. Phys. Chem. Lett.* **2021**, *12*(14), 3662. PMID: 33826849.

[49] Y. Guan, S. V. Shree Sowndarya, L. C. Gallegos, P. C. St. John, R. S. Paton, *Chem. Sci.* **2021**, *12*, 12,012. https://doi.org/10.1039/D1SC03343C

[50] Z. Yang, M. Chakraborty, A. D. White, *Chem. Sci.* **2021**, *12*, 10,802. https://doi.org/10.1039/D1SC01895G
[51] L. Chen, W. Robien, *Fresenius J. Anal. Chem.* **1992**, *344*(4), 214. https://doi.org/10.1007/BF00322714
[52] R. Sasaki, Nmrshiftdb, acd/labs, and modgraph, https://acdlabs.typepad.com/my_weblog/2007/05/nmrshiftdb_acdl.html, **2021**.
[53] C. Cobas, Ensemble nmr prediction, https://resources.mestrelab.com/ensemble-nmr-prediction/, **2021**.
[54] Y. Kwon, D. Lee, Y.-S. Choi, M. Kang, S. Kang, *J. Chem. Inf. Model.* **2020**, *60*(4), 2024. PMID: 32250618.
[55] P. A. Unzueta, C. S. Greenwell, G. J. O. Beran, *J. Chem. Theory Comput.* **2021**, *17*(2), 826.
[56] L. Olsson, D. Cremer, *J. Phys. Chem.* **1996**, *100*(42), 16,881. https://doi.org/10.1021/jp961316p
[57] D. O. Samultsev, V. A. Semenov, L. B. Krivdin, *Magn. Reson. Chem.* **2014**, *52*(5), 222. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/mrc.4055
[58] D. Xin, C. A. Sader, U. Fischer, K. Wagner, P.-J. Jones, M. Xing, K. R. Fandrick, N. C. Gonnella, *Org. Biomol. Chem.* **2017**, *15*, 928. https://doi.org/10.1039/C6OB02450E
[59] Nmr predictors from acd/labs, https://www.acdlabs.com/products/adh/nmr/nmr_pred/index.php, **2021**.
[60] Nmr predict - mestrelab, https://mestrelab.com/software/mnova/nmr-predict/, **2021**.
[61] C. Saunders, M. B. Khaled, J. D. Weaver, D. J. Tantillo, *J. Org. Chem.* **2018**, *83*(6), 3220. PMID: 29470063.
[62] A. Vulpetti, G. Landrum, S. Rüdisser, P. Erbel, C. Dalvit, *J. Fluor. Chem.* **2010**, *131*(5), 570. https://www.sciencedirect.com/science/article/pii/S0022113910000059
[63] V. P. Emerenciano, D. G. Diego, M. J. P. Ferreira, M. T. Scotti, J. V. Comasseto, G. V. Rodrigues, *J. Braz. Chem. Soc.* **2007**, *18*(6), 1183. http://www.scielo.br/pdf/jbchs/v18n6/a12v18n6.pdf, CONDENSED MATTER PHYSICS, SUPERCONDUCTIVITY AND SUPERFLUIDITY.
[64] Y. Binev, M. M. B. Marques, J. Aires-de Sousa, *J. Chem. Inf. Model.* **2007**, *47*(6), 2089. PMID: 17956083.
[65] K. Shibata, H. Kaneko, *Anal. Sci. Adv.*, *n/a*(n/a). https://doi.org/10.1002/ansa.202000180
[66] K. Ito, X. Xu, J. Kikuchi, *Anal. Chem.* **2021**, *93*(18), 6901. PMID: 33929838.
[67] M. Tusar, L. Tusar, S. Bohanec, J. Zupan, *J. Chem. Inf. Comput. Sci.* **1992**, *32*(4), 299. https://doi.org/10.1021/ci00008a007
[68] Spinus - structure-based predictions in nuclear magnetic resonance spectroscopy, http://neural.dq.fct.unl.pt/spinus/, **2021**.

How to cite this article: E. Jonas, S. Kuhn, N. Schlörer, *Magn Reson Chem* **2022**, *60*(11), 1021. https://doi.org/10.1002/mrc.5234