

Deriving three one dimensional NMR spectra from a single experiment through machine learning

Corresponding Author: Professor Leonardo Tenori

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors demonstrate that a standard PLS approach is capable of surprising accuracy in predicting complementary 1D NMR spectra from a single 1D NOESY spectrum, which are then used to quantify metabolite concentrations for complex biofluids. The study has a number of flaws, however, that limit its significance.

- The authors fail to demonstrate the difference between their indirect approach and the direct approach of quantifying metabolite concentrations from a single experimental spectrum. They do not offer a clear rationale for the indirect approach, which could potentially, on one hand, simplify subsequent analysis using existing tools, but also holds the potential to introduce new sources of variance.
- The study may involve implicit "over-training" to samples handled in a specific manner. This concern would be mitigated by specifying a defined sample handling protocol.
- There is little analysis of the error distribution. This is important because the errors could reflect confounding clinical attributes. These could be detected using de-identified clinical data, which the authors did not use.
- MRE may not be the most sensitive or appropriate metric for assessing concordance.
- The Euclidian metric applied directly to down-sampled spectral intensities in the train/validation split loses meaning for high-dimensional data such as the data in this study because, regardless of data distribution, the difference in distances between nearest and furthest points approaches zero as dimensionality increases.
- Kennard-Stone sampling sequentially selects points that are furthest from already selected points according to a distance metric. This procedure is intended to produce a train set with good coverage of the data, reduce imbalance, and avoid highly similar samples. However, this means that the unselected points, that are used here as the validation set, do not benefit from any of these features, and may be biased towards the mean, less balanced, and contain more highly similar samples than the test set. This could lead to unrealistically high scores on the validation set, and models that generalize worse than expected.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors are introducing a machine learning approach for predicting NMR spectra with the acquisition of a single experiment. This is an important advancement that will help overcoming the challenge of time management in NMR acquisition, for example for non-stable sample (i.e biofluids for metabolomics studies).

The authors beautifully describe the approach, validate the method and proceed to apply it in clinical studies. The results are convincing but there might be some comments that I will detail after, for it to become a significant leap forward in the state of the art and to be considered before publication:

- In the section "Goodness of the results in the validation and test sets", 4 metabolites are chosen for demonstrating the robustness and reproducibility by quantifying them.

I believe the 4 are quite concentrated in serum, so I suggest performing the same comparison with more random metabolites (aliphatic, aromatic aminoacids, lipoproteins... high or low concentrated...) to be able to fully correlate the predicted and the

real spectra.

Also, SNR values in the real spectra should be indicated for the chosen metabolites to outstand the performance of the prediction even if LOD/LOQ are low.

Moreover, the quantification method should be indicated (is it by integration, line fitting with standards...?).

- In the same section, the authors show the prediction of pJRES from predicted CPMG. However, with respect to 1D NOESY, CPMG could suffer from loss of intensity in some metabolites as a side effect of the T2-filter effect that suppresses macromolecules. The authors should state how their methodology accounts for this losses to obtain a performant pJRES.

- As a general question, since metabolomics studies could rely on many other acquisition schemes (pure-shift, 2D NMR...), what would be the applicability of the machine learning algorithm presented here ?

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

It saves experimental time to directly predicting CPMG and diffusion-edited spectra from NOESY spectra, instead of acquiring them on spectrometers, especially for high-throughput applications.

I have the following comments:

1. I understand that it makes senses to predict CPMG and diffusion-edited spectra from NOESY spectra because NOESY contains signals of both small molecules and macromolecules. However, pJRS is quite different from CPMG in terms of peak splitting pattern, peak position, signal intensity, et al. Please explain why it makes sense to predict pJRS from CPMG? I believe that predicting one spectrum from another relies on the relation between the two spectra. And I disagree the statement in the Abstract, "for the prediction of 'any' NMR spectrum and constitutes a template for any future applications even beyond metabolomics

2. The authors selected alanine, glucose, lactate, and valine as representative metabolites. However, this selection is biased, as these metabolites all exhibit strong spectral intensities and minimal spectral overlap. The authors should expand their selection to include representative metabolites with weaker signal intensities and more severe spectral peak overlap. This would provide a more comprehensive analysis.

3. There is a similar work (DOI: 10.1038/s42004-024-01251-x). Please compare your work with it.

4. In lines 107, 127 and 150, I guess that it is Fig. 2, not Fig. 3. If so, Fig. 3 is not mentioned in the manuscript.

In line 150, I guess that it is pJRES, not CPMG.

5. The current results are merely validated based on a single dataset split, and it is suggested that the authors perform cross-validation to verify the stability of the method.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have substantively responded to the critiques, and the manuscript is now significantly improved. However there is one concern raised that was not directly addressed by the authors, concerning the use of the predicted spectra for downstream tasks compared to the direct use of the experimental spectrum for the same tasks. While they authors point out that the rationale for not doing this is that it is more convenient to be able to use existing computational tools, the manuscript would benefit further from at least mention of the potential limitations of prediction of the spectra in contrast to the goal of fingerprinting and determining metabolite concentrations.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors have given reasonable answers to all the concerns from reviewers, and included additional data according to the requirements.

In my opinion, the paper is ready to be published as it is, with a minor remark to consider before publication: when trying to run the code, as a beginner user in R, I found the Readme very light, I would suggest to prepare a more complete Readme file with more instructions to follow for running the code and getting the expected results.

(Remarks on code availability)

when trying to run the code, as a beginner user in R, I found the Readme very light, I would suggest to prepare a more complete Readme file with more instructions to follow for running the code and getting the expected results.

Reviewer #4

(Remarks to the Author)

I have only one concern.

Although pJRES spectra can be viewed, in a simplified sense, as CPMG spectra with suppressed scalar couplings, it is hard to predict the former from the latter. For example, for a doublet in the CPMG spectrum, why is it predicted as a singlet in the pJRES spectrum with its chemical shift in the middle of two peaks in the doublet, instead of two singlets? And thus it needs a more reasonable explanation for predicting pJRES spectra from CPMG spectra.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We sincerely appreciate the time and effort the Reviewers have dedicated to evaluating our work and we would really like to thank them for challenging us with several very constructive criticisms. Their insightful and pertinent comments have been invaluable in helping us significantly improve the manuscript. We have carefully considered all the points raised by the Reviewers and we answered each question and comment in detail.

Reviewer #1 (Remarks to the Author):

The authors demonstrate that a standard PLS approach is capable of surprising accuracy in predicting complementary 1D NMR spectra from a single 1D NOESY spectrum, which are then used to quantify metabolite concentrations for complex biofluids.

The Reviewer is right: indeed, PLS regression demonstrated to be a simple, fast, and accurate approach for the prediction of additional NMR spectra from a single 1D NOESY. The striking accuracy found encouraged us to further explore these findings to properly convince the reviewer of the relevance and robustness of our results.

- The authors fail to demonstrate the difference between their indirect approach and the direct approach of quantifying metabolite concentrations from a single experimental spectrum. They do not offer a clear rationale for the indirect approach, which could potentially, on one hand, simplify subsequent analysis using existing tools, but also holds the potential to introduce new sources of variance.

In standard NMR metabolomics, it is customary to acquire a set of spectra for each sample—typically four for serum/plasma—because each spectrum emphasizes different molecular components. As outlined in the manuscript, each type of spectrum serves a distinct analytical purpose: for instance, CPMG spectra are often most informative when the pathology primarily affects small molecule metabolism, whereas diffusion-edited spectra may be more relevant when alterations in the lipoprotein profile dominate.

While it is true that 1D NOESY spectra, in principle, contain information from all molecular species, the direct quantification of individual metabolites from NOESY spectra alone can be challenging due to peak overlap and spectral complexity. Our proposed indirect approach leverages the 1D NOESY spectrum to computationally reconstruct the other spectra (e.g., CPMG, diffusion-edited), effectively disentangling this complex information. This enables the use of existing, well-established analysis pipelines that are tailored to those specific spectra.

As the reviewer correctly notes, one potential benefit of our approach is the simplification of downstream analysis. More importantly, it leads to a substantial reduction in NMR acquisition time—approximately 75%—resulting in lower costs and higher throughput, both of which are critical in large-scale studies.

Crucially, as demonstrated in the paper and reinforced by this revised version, the predictive models introduce only minimal additional variance due to their low prediction error. These points are now

explicitly discussed in Section “A computational pipeline for prediction of NMR spectra”, with additional validation results presented in Section “Goodness of the results”.

• The study may involve implicit “over-training” to samples handled in a specific manner. This concern would be mitigated by specifying a defined sample handling protocol.

We appreciate the reviewer’s comment and acknowledge that this important aspect was not sufficiently emphasized in the original manuscript. As is standard in all metabolomic studies, the robustness and reproducibility of results critically depend on the standardization of both pre-analytical and analytical protocols. We now address this point more explicitly in the revised manuscript (Results and Discussion, pg. 13; Conclusions, pg. 19).

In this study, all samples were prepared and analyzed following the SOP detailed in Methods section. Nonetheless, our system has shown a notable tolerance to minor protocol variations. For instance, while the majority of samples were acquired using 64 scans, samples from one cohort were acquired using 32 scans, yet the resulting data remained consistent. Given that the samples were collected from different centers, across multiple studies and geographical regions, the consistency of the results further underscores the robustness of our approach.

That said, to ensure maximum sample homogeneity, we strongly recommend adhering to the SOP described in Refs. 2 and 16. This recommendation is now clearly discussed in the revised version (“Conclusions” section, pg. 19).

• There is little analysis of the error distribution. This is important because the errors could reflect confounding clinical attributes. These could be detected using de-identified clinical data, which the authors did not use.

We thank the reviewer for highlighting this important point. Our manuscript is based on spectra of serum collected in three independent projects aiming to investigate severe obesity (project ID: MTBLS242), acute myocardial infarction (project ID: MTBLS395) and breast cancer (project ID: MTBLS424). In the revised manuscript, we included a detailed analysis to evaluate the impact of different clinical conditions based on the root mean square error (RMSE) distribution across the three datasets used, as illustrated in Supplementary Figure S1. Relevant clinical and demographic differences are undoubtedly present across the three studies: our training set was intentionally designed to incorporate this variability. This was done to improve model robustness and support broader applicability across different clinical and experimental conditions.

Analyses from the breast cancer dataset exhibit slightly higher errors. We believe this increase is attributable to differences in acquisition protocols rather than clinical attributes. Notably, these spectra were acquired with only 32 scans, compared to 64 scans for the other datasets.

Smaller differences in error distribution were observed between the severe obesity and myocardial infarction studies, with the exception of diffusion-edited spectra. In this case, higher errors in the obesity group may be explained by the particularly intense lipoprotein signals in pre-surgery samples. These considerations are now reported at page 17.

• MRE may not be the most sensitive or appropriate metric for assessing concordance.

As suggested by the Reviewer, we included in the new Table 1 other metrics such as the root mean square error (RMSE), the coefficient of determination (R^2) and the ratio of performance to deviation

(RPD). In order to satisfy the requests of reviewer #4, we additionally included a 10-fold cross-validation analysis to ensure the robustness of our results. Considering the results in light of all evaluation metrics and all models, the overall performance of the proposed approach appears to be of high quality.

- The Euclidian metric applied directly to down-sampled spectral intensities in the train/validation split loses meaning for high-dimensional data such as the data in this study because, regardless of data distribution, the difference in distances between nearest and furthest points approaches zero as dimensionality increases.

We respectfully disagree with the assertion that Euclidean distance loses its meaning in high-dimensional settings *per se*. The commonly cited issue known as the “curse of dimensionality” — wherein distances between nearest and furthest points converge — indeed occurs under the assumption that data features are independent and identically distributed (i.i.d.), as shown in studies such as Aggarwal et al. (2001), Beyer et al. (1999), and Francois et al. (2007). However, this assumption is rarely valid for real-world datasets, which typically exhibit significant structure and correlations among features.

In fact, it has been shown that many real-world high-dimensional datasets lie on or near a lower-dimensional manifold. In such cases, distance metrics such as the Euclidean distance can remain meaningful and effective for tasks like clustering or nearest-neighbor search (Verleysen & François, 2005). This is one reason why simple algorithms like k-nearest neighbors (KNN) often perform surprisingly well even in high-dimensional spaces — a phenomenon sometimes described as the “blessing of dimensionality” (Houle et al., 2010).

In our case, the NMR spectral data are clearly not i.i.d.: spectral intensities are highly correlated within each peak; peaks corresponding to the same compound are also interrelated; and compounds that are part of related biochemical pathways show further correlations. As a result, the spectral data — though apparently high-dimensional — effectively reside on a structured and lower-dimensional manifold. That said, to address the Reviewer’s concern we decided to use Mahalanobis distance (retaining the number of principal components which explains at least 95% of the total variance) as the chosen metric in this revised version of the manuscript for the Kennard-Stone algorithm (see answer below). Methods were revised accordingly.

- Kennard-Stone sampling sequentially selects points that are furthest from already selected points according to a distance metric. This procedure is intended to produce a train set with good coverage of the data, reduce imbalance, and avoid highly similar samples. However, this means that the unselected points, that are used here as the validation set, do not benefit from any of these features, and may be biased towards the mean, less balanced, and contain more highly similar samples than the test set. This could lead to unrealistically high scores on the validation set, and models that generalize worse than expected.

We thank the reviewer for the insightful comment regarding the implications of the Kennard-Stone algorithm on the representativeness of the validation set. The Kennard-Stone method was intentionally chosen to ensure that the training set provides broad coverage of the spectral feature space, a critical factor in spectroscopy-based models. While it is correct that the remaining unselected samples used for validation may be more clustered or closer to the data centroid, this actually reflects typical conditions in applied settings, where many test samples are expected to lie near the average of the distribution. Moreover, since the model is trained on a well-dispersed and information-rich set, it is less

likely to suffer from underfitting and more likely to generalize effectively to both central and edge-case samples. The use of Kennard-Stone in this way is well-established in the literature (Kennard & Stone, 1969; Xu & Liang, 2001; Zheng & He, 2010; Shao & Rinnan, 2020), especially in chemometric applications where high dimensional data are often used (e.g. NIR or FTIR spectra) and calibration robustness is essential. Nonetheless, we acknowledge the concerns of the Reviewer, and to dispel all doubts, we have run again the procedure, splitting the data in training and validation sets using the KS algorithm in combination with Mahalanobis distance (i.e. distance calculated in the principal component space). In this way all effects of redundant and noisy dimensions are explicitly removed. All analyses were re-performed using the new data-split, and the results obtained are totally in line with the previous analyses. Furthermore, we tested our models in an independent test set that is not affected by the splitting strategy. Finally, a 10-fold cross validation approach is also presented, as requested by Reviewer 4 (see the new Table 1).

Reviewer #2 (Remarks to the Author):

The authors are introducing a machine learning approach for predicting NMR spectra with the acquisition of a single experiment. This is an important advancement that will help overcoming the challenge of time management in NMR acquisition, for example for non-stable sample (i.e biofluids for metabolomics studies).

The authors beautifully describe the approach, validate the method and proceed to apply it in clinical studies. The results are convincing but there might be some comments that I will detail after, for it to become a significant leap forward in the state of the art and to be considered before publication:

We sincerely thank the reviewer for this word of appreciation.

- In the section "Goodness of the results in the validation and test sets", 4 metabolites are chosen for demonstrating the robustness and reproducibility by quantifying them.

I believe the 4 are quite concentrated in serum, so I suggest performing the same comparison with more random metabolites (aliphatic, aromatic aminoacids, lipoproteins... high or low concentrated...) to be able to fully correlate the predicted and the real spectra.

Also, SNR values in the real spectra should be indicated for the chosen metabolites to outstand the performance of the prediction even if LOD/LOQ are low.

Moreover, the quantification method should be indicated (is it by integration, line fitting with standards...?).

In order to satisfy the request from the reviewer and to more strongly demonstrate the robustness of the method, we quantify a panel of 20 metabolites both in the real and predicted spectra. The results are shown in Figure 4. This is a collection of signals from both very concentrated metabolites (high SNR, like glucose, alanine, lactate, GlycA) and signals from low concentrated metabolites (like formate, phenylalanine, tyrosine). All correlations were statistically significant (all p-values < 0.001). Fifteen metabolites showed correlation coefficients (r) greater than 0.90, while four metabolites (citrate, histidine, phenylalanine, and tyrosine) showed suboptimal correlations, with r values between 0.75 and 0.90. Only formate showed a weaker correlation (r = 0.51), likely due to the low intensity of its signal. Taken together these results proved the accuracy of our predictions. These results are described at page 10. Metabolite assignment is now provided in Supplementary Table S1. Further, the method section has been expanded to better explain the quantitation approach employed.

- In the same section, the authors show the prediction of pJRES from predicted CPMG. However, with respect to 1D NOESY, CPMG could suffer from loss of intensity in some metabolites as a side effect of the T2-filter effect that suppresses macromolecules. The authors should state how their methodology accounts for this losses to obtain a performant pJRES.

We thank the reviewer for this insightful comment. We agree that the T2-filter effect inherent in CPMG sequences can lead to a loss of signal intensity for certain metabolites, which may potentially affect the accuracy of the predicted pJRES spectra. To mitigate this issue, we also trained an alternative model to predict pJRES directly from 1D NOESY spectra, which are not subject to T2-filtering and thus retain the full intensity of metabolite signals. As now reported in the new Table 1, we compared the performance of the pJRES prediction models based on predicted CPMG and on NOESY spectra. Interestingly, the model based on predicted CPMG slightly outperformed the NOESY-based model, particularly in the

independent test set. This result may be attributed to the CPMG sequence's suppression of macromolecular background signals, which likely facilitates a cleaner mapping to the pJRES domain. We would also like to emphasize that although accurate prediction of pJRES signal intensity is inherently challenging, this does not compromise the practical utility of pJRES spectra. pJRES spectra themselves are not quantitative and are not used directly for absolute quantitation but rather for disambiguating between nearby peaks and for peak assignment, where its role remains critical. These considerations are now reported in the manuscript at page 14.

- As a general question, since metabolomics studies could rely on many other acquisition schemes (pure-shift, 2D NMR...), what would be the applicability of the machine learning algorithm presented here?

Our current approach is designed to work with one-dimensional (1D) NMR data. It can read these types of spectra and produce new predicted spectra in a format that's compatible with the Bruker format. This means it can already handle different 1D NMR techniques, such as pure-shift or DIRE, with the only requirement to have a sufficiently large dataset to build a robust new training set. The machine learning method used could, in theory, be used to predict 2D spectra as well. However, we are doubtful about predicting 2D spectra from one-dimensional one, because in the training matrix (with NOESY spectra) could not be sufficient information to reconstruct the 2D cross-peaks. A limitation of using this approach to spectra prediction is the access to a sufficiently large number of spectra, acquired using standardized procedures, to build a robust training set. This consideration is now added to the manuscript's main text (Conclusions section, page 18).

Reviewer #4 (Remarks to the Author):

It saves experimental time to directly predicting CPMG and diffusion-edited spectra from NOESY spectra, instead of acquiring them on spectrometers, especially for high-throughput applications.

The Reviewer perfectly grasp the meaning of our work.

1. I understand that it makes senses to predict CPMG and diffusion-edited spectra from NOESY spectra because NOESY contains signals of both small molecules and macromolecules. However, pJRS is quite different from CPMG in terms of peak splitting pattern, peak position, signal intensity, et al. Please explain why it makes sense to predict pJRS from CPMG?

I believe that predicting one spectrum from another relies on the relation between the two spectra. And I disagree the statement in the Abstract, “for the prediction of ‘any’ NMR spectrum and constitutes a template for any future applications even beyond metabolomics”.

We thank the reviewer for this insightful comment. It is true that pJRES spectra differ substantially from CPMG, particularly in terms of peak splitting, position, and signal intensity. In fact, pJRES are also very different from NOESY spectra—perhaps even more so than from CPMG—since pJRES do not display the broad macromolecular signals characteristic of NOESY and instead present a flatter baseline, similar to CPMG.

We acknowledge that we cannot definitively state whether predicting pJRES from CPMG is the most theoretically sound approach. However, empirically, it proved to be effective: in this revised version it is reported along with the prediction of pJRES from predicted CPMG spectra the prediction directly obtained from NOESY spectra. The former approach yielded better results than the latter (see now Table 1). A possible explanation is that pJRES spectra can be viewed, in a simplified sense, as CPMG spectra with suppressed scalar couplings—i.e., with reduced multiplicities. This conceptual proximity might help explain the improved performance when using CPMG as the input domain. These considerations have been added to the section “Goodness of the results” at page 8.

The statement in the abstract and in the conclusions has been rephrased and smoothed.

2. The authors selected alanine, glucose, lactate, and valine as representative metabolites. However, this selection is biased, as these metabolites all exhibit strong spectral intensities and minimal spectral overlap. The authors should expand their selection to include representative metabolites with weaker signal intensities and more severe spectral peak overlap. This would provide a more comprehensive analysis.

In this revision, also as an answer to a similar criticism raised by Reviewer 2, we quantify an extended panel of 20 metabolites both in the real spectra and in the predicted one. The results are shown in Figure 4. Signals from both very concentrated metabolites (e.g. glucose, alanine, lactate, GlycA) and signals from low concentrated metabolites (e.g. formate, phenylalanine, tyrosine) have been analyzed. All correlations were statistically significant (all p-values < 0.001). Fifteen metabolites showed correlation coefficients (r) greater than 0.90, while four metabolites (citrate, histidine, phenylalanine, and tyrosine) showed suboptimal correlations, with r values between 0.75 and 0.90. Only formate showed a weaker correlation ($r = 0.51$), likely due to the low intensity of its signal. Taken together these results proved the accuracy of our predictions. These results are described at page 10.

3. There is a similar work (DOI: 10.1038/s42004-024-01251-x). Please compare your work with it.

We sincerely thank the reviewer for bringing this article (DOI: 10.1038/s42004-024-01251-x) to our attention. We were not aware of it prior to this revision round. We have now carefully compared this work with ours and included the discussion in the manuscript, specifically in the Results and Discussion section "A computational pipeline for prediction of NMR spectra", page 6.

4. In lines 107, 127 and 150, I guess that it is Fig. 2, not Fig. 3. If so, Fig. 3 is not mentioned in the manuscript.

In line 150, I guess that it is pJRES, not CPMG.

We thank the Reviewer for pointing out that errors. The manuscript has been carefully revised, and figure mentions checked.

5. The current results are merely validated based on a single dataset split, and it is suggested that the authors perform cross-validation to verify the stability of the method.

We thank the reviewer for this suggestion. In this revised version of the manuscript to ensure a robust evaluation of each model's performance we included the results obtained by a 10-fold cross validation scheme (see the new Table 1). This assessment reinforces the robustness of our findings.

REVIEWER COMMENTS

Reviewer #1:

The authors have substantively responded to the critiques, and the manuscript is now significantly improved. However, there is one concern raised that was not directly addressed by the authors, concerning the use of the predicted spectra for downstream tasks compared to the direct use of the experimental spectrum for the same tasks. While the authors point out that the rationale for not doing this is that it is more convenient to be able to use existing computational tools, the manuscript would benefit further from at least mention of the potential limitations of prediction of the spectra in contrast to the goal of fingerprinting and determining metabolite concentrations.

We thank the reviewer for the opportunity to further clarify the rationale behind our approach. We acknowledge that our previous response may not have fully conveyed our line of reasoning, and we appreciate the chance to elaborate on this important point.

If we correctly understand the reviewer's concern, it focuses on the question: "why predict additional spectra for downstream analysis rather than directly using the experimental NOESY spectrum to address the biological questions of interest?" At first glance, this may indeed appear to be an unnecessary detour.

In our previous reply, we noted that standard NMR-based metabolomics workflows typically require the acquisition of multiple spectra (e.g., NOESY, CPMG, diffusion-edited, and JRES). Our approach, predicting these spectra from a single NOESY acquisition, saves time and resources while enabling the use of established analysis pipelines and software that are tailored to specific spectral types. For instance, if the goal is to generate a list of quantified metabolites, NOESY spectra are suboptimal due to the presence of broad macromolecular signals; CPMG spectra, by contrast, allow for more reliable signal integration of small metabolites.

The reviewer also refers to fingerprinting approaches where non-quantified spectra are used for classification tasks via machine learning. Indeed, classifiers can be trained directly on NOESY spectra. However, depending on the biological question, we have observed that sometimes NOESY and CPMG spectra yield different classification performance. Particularly when the discriminative power relies on signals affected by macromolecular background (e.g., lipoproteins) NOESY spectra show better performance (e.g., Tenori et al. 2013, 10.1016/j.ijcard.2013.08.042), whereas when low molecular weight metabolites are more important in the discrimination CPMG spectra enable high accuracy (e.g., Hart et al. 2017, 10.1158/1078-0432.CCR-16-1153). Therefore, having access to multiple spectral types, including predicted ones, increases the flexibility and applicability of machine learning pipelines to different use cases.

That said, a natural question arises: "Is there any practical disadvantage in predicting all the needed spectra from a single NOESY acquisition?" Well, if the predictions were absolutely perfect, then definitely no: there would not be any downside. You would get three spectra essentially for free, saving time and effort, and giving the researcher the freedom to choose which spectrum to analyse depending on the biological question and the characteristics of the samples. But, as we all sadly know, there is no such thing as a free lunch. Prediction errors are, unfortunately, unavoidable, and their propagation into downstream analyses, especially those relying on subtle spectral features or low-concentration metabolites, must be carefully considered.

Our current validation is limited to specific studies, and we recognize that broader benchmarking is needed to assess how deviations in predicted spectra impact downstream applications such as fingerprinting and metabolite quantification. We have therefore revised the manuscript to explicitly discuss these limitations. In particular, we have now emphasized that the use of predicted spectra in tasks requiring high fidelity, such as absolute quantification or the detection of weak signals, may demand additional validation and awareness of the potential for error accumulation.

We appreciate the reviewer's suggestion, which has led to an important clarification in the manuscript. We would also like to reiterate that the main goal of this study is to demonstrate the feasibility of spectral prediction from a single acquisition. We are not proposing this as a replacement for multi-spectral acquisition in all future metabolomics investigations, but rather as an additional tool that may offer practical advantages in many, but not necessarily all, scenarios.

All the abovementioned considerations are now present in the manuscript (Results and Discussion and Conclusion sections).

Reviewer #2:

Remarks to the Author: *The authors have given reasonable answers to all the concerns from reviewers, and included additional data according to the requirements. In my opinion, the paper is ready to be published as it is, with a minor remark to consider before publication: when trying to run the code, as a beginner user in R, I found the Readme very light, I would suggest to prepare a more complete Readme file with more instructions to follow for running the code and getting the expected results.*

Remarks on code availability: *when trying to run the code, as a beginner user in R, I found the Readme very light, I would suggest to prepare a more complete Readme file with more instructions to follow for running the code and getting the expected results.*

We sincerely thank the Reviewer for this helpful suggestion aimed at improving the usability of our R code, especially for beginner users. In response, we have revised the Readme file to include more comprehensive and detailed instructions on how to run the code and obtain the expected results. Additionally, we have updated the Excel file input by clarifying the column headings, making it easier to understand and use. We hope these improvements will facilitate the use of our code by a broader audience.

Reviewer #4:

I have only one concern. Although pJRES spectra can be viewed, in a simplified sense, as CPMG spectra with suppressed scalar couplings, it is hard to predict the former from the latter. For example, for a doublet in the CPMG spectrum, why is it predicted as a singlet in the pJRES spectrum with its chemical shift in the middle of two peaks in the doublet, instead of two singlets? And thus it needs a more reasonable explanation for predicting pJRES spectra from CPMG spectra.

We thank the reviewer for raising this insightful and technically important point regarding the determination of pJRES spectra from CPMG spectra. We fully agree that this prediction is not trivial, particularly in terms of scalar (J) coupling behaviour, and this point needs additional explanation. As noted, pJRES spectra can be thought of, in very simplified terms, as CPMG-like spectra with suppressed scalar couplings, where multiplets are collapsed into singlets, typically located at the average chemical

shift of all the coupled peaks. In a CPMG spectrum, for example, a doublet might appear as two distinct peaks, while in the corresponding pJRES spectrum, this information is collapsed into a single peak centred between the two. This is an intrinsic behaviour of pJRES spectra, and actually it is the main reason to obtain them. The suppression of multiplicity reduces spectral crowding, thus improving spectral assignment.

The reviewer is absolutely right to highlight that predicting this transformation is not straightforward. The model is not trained with explicit rules about scalar couplings or the quantum mechanical basis of multiplet structures. Instead, it simply learns this behaviour implicitly from training on matched pairs of spectra, CPMG and pJRES, from the same samples. Regarding the specific concern: why the model predicts a singlet at the center, rather than two singlets, this behaviour is not imposed by design (the model is agnostic about the inner structure of NMR spectra) but rather emerges from the training data. Since true pJRES spectra never exhibit doublet peaks, the model learns to avoid that outcome. Indeed, because in many training examples when there is a doublet in the CPMG it found a single peak in the pJRES, the model learns to put a singlet in that position. The chemical shift of the predicted singlet reflects the center of mass of the original doublet, in line with the experimental pJRES pattern (as an example, the doublet signal of the CH₃ protons of the valine centred at 1.04 ppm is predicted as a singlet centred at 1.04 ppm as experimentally detected in pJRES, see Figure 3 and Figure 5). Across many metabolites and many examples, the PLS model consistently sees that certain multiplet patterns in the CPMG spectrum correspond to certain singlets in the pJRES spectrum, thus learning how to generalize this transformation in unseen examples.

Of course, this data-driven approach has its limitations. The model does not explicitly understand or simulate scalar coupling mechanisms and therefore may not generalize perfectly to novel coupling patterns outside of the training distribution. We now clarify this point in the revised manuscript (Section “Goodness of the results”).

We appreciate the reviewer’s comment, as it allows us to make the manuscript more precise regarding the physical basis, or, in this case, the data-learned nature of the predicted spectral features.

Reviewer #5:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

See answers to the comments of Reviewers #1, #2, and #4.