**MAGNETIC
RESONANCE**
Open Access
Discussions

Interactive comment on "Improving the Accuracy of Model-based Quantitative NMR" *by* Yevgen Matviychuk et al.

Anonymous Referee #2

Received and published: 25 March 2020

The authors propose an improved algorithm for quantitative NMR (qNMR) through model adjustments. The introduction section provides a good review of a number of challenges that are commonly faced in model specification. Although, a full discussion related to the specification of number of components in the model – a factor that can contribute significantly to model misspecification – is not offered. Overall, the discussion in the Introduction section is well-presented.

There is an example involving the modeling of a mixture in the manuscript which leads me to think that deconvolving mixtures is a supported feature. If this is correct, then perhaps the authors implicitly assume that the number of components are a known quantity? If so, then this assumption should be stated explicitly since the number of components in a linear mixture model in the presence of noise can have influence the

calculated parameters of the model in a significant way.

On the other hand, if the authors assume that the number of components is an unknown parameter, then two additional points for discussion should be included in order to help readers develop a fuller picture: a) what procedures should be used for deriving the number of components?, and b) what would be the impact of the number components on the "model adjustment" process proposed by the authors?

It is worth noting that NMR signals are non-stationary, and there is supporting data for the view that the appropriate noise model for NMR spectra is not Gaussian. As a result, the outcome of the model fitting stage, depending on the objective function used for optimization, may lead to a residual noise background that is challenging to characterize. More specifically, the underlying distribution of this residual noise may change (according to the standard metric) in every iteration of an algorithmic optimization and at every point of the spectrum. This phenomenon, combined with lack of a priori knowledge about the number of components in the model, can hinder the adjustment procedure for weaker peaks.

As stated in the manuscript, the model adjustment process attempts to "... find a model spectrum, which after the subtraction from the experimental data would lead to exclusively noise in the residual." This is an intuitively appealing objective, but only a heuristic (not rigorous) form of this statement is given. The authors state in a later paragraph that the specification of this objective through equation 5 is ill-posed, but their statement does not explicitly address the issue of how "exclusively noise" can be determined. Recall that two noise vectors of identical power (or different power) can be far apart in the standard metric but rather close in an appropriately selected metric. For example, the cosine distance for two noise vectors generated from the same random process can be close to 1 ($1 - r$, where r is the cosine between two noise vectors, is considered very far). Therefore, the precise specification of how the residual becomes exclusively noise is necessary.

The authors address the challenge of addressing "noise equality" by focusing on the component of the residual signal that is related to the model "misfitting" in the decomposition presented in equation 5. Their proposal is to remove the noise through the use of symlet wavelets and a set of filters and consider the residual as "misfitting". Since symlets are dependent on the order of their filter, and there are a multitude of filtering approaches, the proposed approach has the side effect of introducing additional parameters. The authors present a figure showing the impact of symlet order on RMSE in peak ratios which seem to indicate the superiority of the symlet over reference deconvolution. However, the RMES in peak ratios seem to suggest very small improvements as a result of applying the proposed algorithm. Is this correct? There are no quantitative guidelines for the selection of these parameters and the implicit assumption that the reader is expected to arrive at based on the examples presented is that the impact of parameters is negligible. Is this the claim that is being made?

As a tool for user-supervised spectral fitting, the proposed approach appears to be useful, assuming that the software in source form is provided to the community. Do the authors envision the use of this software as a user-supervised tool? Or, do they consider the software as providing an automated procedure? There is plenty of computational power in a typical desktop to perform near real-time changes and visual feedback based on user-adjustable parameters, if a friendly GUI for the task is provided. Perhaps this GUI could be something developed by the community if one had not been developed already. It is unclear if this material is available and what licensing rights are available for future open developments.

Some additional discussions will add significantly to the science presented in the manuscript: a) Additional clarifications regarding the comments/questions mentioned above. b) Clear and concise statements regarding the assumptions made in devising this algorithm. For example, is it assumed that the number of components is known a priori? c) The capabilities and limitations of the approach. For example, how many peaks can be handled and how much overlap is tolerated by the algorithm? d) modes

of parameters specification in the software, and guidance on the method of comparing outcomes will add. If it is being used as a user-guided tool, what utilities are available for guiding the user in selecting the "best" solution? If it is left to the user, then please state it. e) the use of software as a user-supervised tool. Do the authors view this as an automated tool, or is it viewed as a user-operated tool to gain insight? f) Availability of the software and license rights.