Open Access MAGNETIC RESONANCE Discussions

Magn. Reson. Discuss., referee comment RC2
https://doi.org/10.5194/mr-2021-28-RC2, 2021

Comment on mr-2021-28

Anonymous Referee #2

Referee comment on "Bootstrap aggregation for model selection in the model-free formalism" by Timothy Crawley and Arthur G. Palmer III, Magn. Reson. Discuss., https://doi.org/10.5194/mr-2021-28-RC2, 2021

Crawley and Palmer propose the application of bootstrap aggregation methodology to derive uncertainty estimates for the analysis of magnetic field strength-dependent NMR relaxation data in the context of the model-free spectral density function selection protocol previously introduced by Palmer and colleagues. A weakness of the least-squares algorithm described in the widely used original protocol is its susceptibility to model-selection error. Sets of relaxation values that are statistically consistent with each other in terms of estimated experimental uncertainty may give rise to the selection of different model-free spectral density functions. In the present manuscript, the statistical method of bootstrap aggregation is used to enable a joint refinement in the parameter uncertainty estimation and the dynamical model selection process.

As applied to ^{15}N relaxation data for the bZip domain of GCN4, bootstrap aggregation is reported to reduce residue-to-residue variations in optimal model-free parameters, particularly in the partially disordered basic region. Regarding the more general practical utility of the proposed sampling protocol, while experimental relaxation data collected at four magnetic field strengths yields 6859 suitably filtered combinations of bootstrap samples, as noted by the first reviewer, the robustness of the statistical analysis may appreciably decline when this value drops to 343 for three magnetic field strengths, and presumably will decrease significantly further when it drops to only 27 for data from two magnetic field strengths.

A key step in the proposed joint refinement process calculates each of the averaged dynamical parameters by summing over the estimates obtained from each of the five spectral density representations being utilized, as weighted by how often each of these five models have been selected (Eq. 10). A potential concern over this approach arises from the fact that while the same set of symbols (τ_m, S_f^2,τ_f,S_s^2,τ_s) are utilized in each of the five dynamical models used, the functional significance of each symbol is defined within the context of the specific equation being used. Each of these five model equations that are used to represent the spectral density function is capable of accurately fitting only a small subset of the physically plausible spectral density curves. Systematic bias can

potentially arise not only with respect to a given dynamics parameter being utilized in distinct model representations but also as a result of the inadequacy with which each of the five model spectral density equations are capable of representing the physical dynamics of the system. While such biasing effects are surely diminished for Model 4 and 5 which incorporate four and five adjustable parameters, respectively, more promising might be the utilization of alternative model equations for the spectral density function that can more robustly represent the range of motion occurring in protein molecules which utilize a smaller set of adjustable parameters for optimization against experimental relaxation data.