July 27, 2023

University of Washington

School of Oceanography

Box 355351

Seattle, Washington 98195

Phone: (206) 221-6748

Dear Editors of BMC Bioinformatics,

On behalf of all co-authors and myself, I submit for publication consideration a manuscript entitled “Picky with peakpicking: assessing chromatographic peak quality with simple metrics in metabolomics”.

Chromatographic peakpicking continues to represent a bottleneck in automated and untargeted LC-MS workflows. The noisy signals produced by the instrument make this a difficult problem requiring extensive manual curation of the resulting dataset and tradeoffs between false positives (noise features included as real signal) and false negatives (real signal missed). Compounding this problem is a wealth of heuristic metrics of peak quality and a lack of comprehensive testing, especially in atypical environments. This manuscript attempts to fill this gap in knowledge by presenting two fully-labelled oceanographic LC-MS datasets that are then used to test a variety of quality metrics and regression algorithms. We find that complex linear models and simple machine-learning techniques typically overfit on training data and extrapolate poorly to novel datasets, but that a simple two-parameter logistic model based on our novel, intuitive metrics can reduce the proportion of false positives from >80% to <20% while still recovering more than 75% of the real features when applied to new data. We then show that the proportion of noise peaks in a metabolomic dataset can significantly affect the conclusions obtained, with a dataset consisting of high-quality features alone demonstrating significant restructuring of the metabolome in both multivariate and univariate space.

Given this manuscript’s focus on both the technical investigation and improvement of peakpicking algorithms as well as its presentation of novel oceanographic results, we believe that BMC Bioinformatics would be a good fit. Please find included in this submission the manuscript with figures embedded, the figures themselves attached at the size and resolution recommended by the guidelines for authors and named by their location in the manuscript, and supplementary figures and tables. The manuscript has been prepared as a single reproducible document available with the associated code at <https://github.com/wkumler/MS_metrics>. All metabolomic data has been deposited into Metabolomics Workbench and is accessible at project IDs PR001021, PR001317, and PR001738.

Thank you for your consideration,



Anitra E. Ingalls

Professor, University of Washington

aingalls@uw.edu