Interactive comment on "Analysis of functional groups in atmospheric aerosols by infrared spectroscopy: method development for probabilistic modeling of organic carbon and organic matter concentrations" by Charlotte Bürki et al.

Anonymous Referee #3

Received and published: 6 January 2020

Bürki et al. present a probabilistic modeling framework for estimating organic carbon concentrations and OM/OC ratios from infrared spectroscopy, and they apply it to infrared spectra of PM2.5 samples from 17 monitoring sites of the IMPROVE network. The presented approach is based on previous developments regarding functional group analysis from infrared spectroscopy and statistical calibration strategies for organic aerosol quantification. Here, the authors apply Bayesian calibration to provide plausible estimates for parameters in the probabilistic model, and they obtain OC

concentrations and OM/OC ratios consistent with other estimation approaches. The presented work is well within the scope of AMT. However, in its present form the manuscript is difficult to follow especially for atmospheric scientists who do not use linear algebra in their day-to-day work. Thus, the manuscript requires major revisions to improve clarity and to focus on the novelty of the approach before publication in AMT.

While I appreciate the rigorous explanations in the appendices and in the supplementary material, some parts of the main text either require additional information or should be moved to the supplementary material in order to focus on the main objective of the manuscript. In particular, the introduction of the probabilistic framework in section 1.2 might be revised with the general reader in mind. Also, the description of the cluster analysis might be shortened, and Figures 4 and 6 might be moved to supplement section S3.

When comparing the FTIR estimates of OM/OC with the reconstructed fine mass (RCFM) regression solved both by ordinary least squares (OLS) regression and error-in-variables (EIV) regression, the extensive comparison of OLS and EIV seems distracting. It may be beneficial to briefly introduce both OLS and EIV in section 3.3 but then restrict the comparison in section 4.4 and Figures 8 and 10 to FTIR and only OLS, or only EIV.

Functionalization by aldehyde, peroxide, aromatic, phenolic, organonitrate, and organosulfate groups is not included in the presented set of calibrations. While it is prudent to prioritize functional groups that are expected to be highly abundant, one has to be very careful when interpreting the results, as stated for example when discussing low mass recovery fractions that "we cannot rule out the need to examine additional FGs" (l.349). For example, organosulfates may become more important in situations when biogenic VOCs are processed in anthropogenically influenced air masses. Thus, changes in OM/OC ratios between 2011 and 2013 observed in the RCFM regression but not observed in the FTIR estimates could also indicate an increasing influence of FGs not taken into account in the presented set of calibrations. I recommend a brief

discussion.

Minor comments: Figure 2: What is the reasoning for fitting Weibull distributions to the prior "fractional carbon" coefficients but a normal distribution to the mass recovery fraction? Figure 7 is introduced after Figure 4 and before Figures 5 and 6. Please re-order figure numbers. Conclusions: The final notion that additional constraints from additional measurements such as NMR or photometry can be added is really helpful and important.

Technical comments: Please introduce all abbreviations when first used, e.g. "RCFM" in line 141, L-BFGS-B in l.562, etc. l.5: "For instance, a subset of model..." - please revise the sentence. l.14: Carboxylic acid should be "COOH". l.82 "... proposed an extension to this approach..." l.138/Figure 1: Please indicate the four regions SW/NW/SE/NE by adding lines on the map in Fig. 1. l.165: "L2 norm" might need explanation. l.201: Revise "...regression of to y". l.355: "but it more likely reflects" instead of "but is more likely reflects" l.394: Add parentheses to references. l.451: Remove the extra "from" from "...that estimate OM/OC from from mass balance...". Figure 2: "representing" instead of "representating"