

The manuscript by Topping et al. reports on a computationally efficient method for predicting activity coefficients of organic constituents in multicomponent systems (organics + water). As noted by the authors, efforts to reduce the computational expense of such models are greatly needed, as the benchmark models (e.g., UNIFAC) are too expensive to be incorporated into large scale chemical transport models and the consequences of ignoring or simply assuming fixed deviations from ideality may be significant. In addition to being more computationally efficient, the method was developed to ultimately allow for consideration of systems with inorganics and organics, which are of greater atmospheric relevance. The paper is generally well written, though some attention to detail is needed, particularly in Section 2 (describing the framework) and with regard to the Tables and Figures. It is recommended that once the revisions are made and comments addressed, the manuscript be accepted for publication in Geoscientific Model Development.

Technical comments:

p. 1766, lines 24-28: The analysis of the binary vs. ternary vs. UNIFAC statistics described here is not clear. The authors state that the "...average predictions from ternary are smaller than those from binary..." Does that mean that overall the ternary model is more accurate (which is what is shown in Table 4, referenced in the same sentence, but has already been stated)? Or does it mean that the ternary model predicts smaller deviations from unity (which would be interesting but not shown anywhere...)? The authors then say, "For example, for "4", "5" and "8" the difference between both variants is large...", which suggests the comparison of binary vs. ternary looking for compounds with either the largest difference between the binary and ternary percentage deviations (though the word "both" is confusing) or the compounds for which both the binary and ternary predictions are poorest; the listed compounds (when compared with other compounds in Table 4) do not really serve to clarify. Finally, the authors then reference the reported percentages for compound "7" which is not listed (but supports the suggestion of identifying the largest differences in the percentage deviations; though not clear then how that supports the first sentence).

p. 1769, line 11: Is compound "12" dominant with and without the consideration of activity coefficients? It is recommended that a clarification be made here (e.g., "12" is the dominant compound in the ideal case") and if another compound became dominant in the non-ideal case it may be interesting to note or discuss. This may be particularly interesting given that the compounds chosen were the most abundant particle-phase compounds in a modeling study that (presumably) neglected activity. Note: "dominate" should be "dominant".

Were there any significant differences in the mass or number size distributions between the ideal and non-ideal cases? While it would be difficult to generalize from such differences, it may add weight to the significance of including activity coefficient corrections.

Comments on Section 2:

Overall it is recommended that the authors review each of the symbols and indices used in Equations 2-15 to ensure that they are unique, defined, and complete. Some examples of symbols/indices needing revision follow.

Eq 2: “ j ” is not defined; assumed that p. 1758, line 7: component “ i ” should be component “ j ”. “ N ” is not defined. Suggest replacing “ $i = 1$ ” in the summation with “ i ” to be consistent with summations including components “ i ” and “ j ” (or alternatively, replace all “ $j \neq i$ ” with “ $j (\neq i) = 1$ ”).

p. 1758, line 5: remove italics from “ln”

Eq. 3-5: It is not clear why m (which is not defined) is used to represent the number of components “ j ” in Eq. 4 and 5, but “ N ” is used in Eq. 3. It is suggested that “ m ” also be used in Eq. 3, even if $N = m$.

p. 1759, line 1: r_j and q_j are not defined.

p. 1759, following line 11: It is suggested that the reference and standard states be defined here.

Eq. 9: The summation is not needed.

p. 1760, line 14: In the partial derivative, “B” should be “ x_B ”.

p. 1761, line 11: The notation β_{x_B} is confusing; it is suggested that “ x'_B ” be removed as a subscript. It seems that retaining the notation $\beta(F(x'_B))$ would be better, as has been done with the α variable. (Same comment for p. 1762, line 11.)

Other minor editorial comments:

Abstract, line 15: Replace “upto” with “up to”

p. 1764, line 12 and p. 1765, lines 21 and 26: Replace Eq. (B1) with Eq. 13.

p. 1764, line 23: “compound” should be “compounds”; line 24: suggest that “as defined within the MCM” be placed in parenthesis

p. 1765, line 10: Replace “A example” with “An example”; line 28: “computational” should be “computationally”

p. 1765, line 12: Replace “Figs” with “Figures”.

p. 1767, line 5: “PD-FITE” should be “PD-FiTE”

Tables and Figures:

Table 2: Polynomials for only 12 compounds are listed.

Please check all Table numbers and references in citations. Table 4 is referenced in the text as Table 5; Table 5 is referenced as Table 6, etc. More description is needed in the Table 4 header (e.g., describing that values are averages over multiple runs/RH values, etc.).

Fig. 1: The color scale is not defined/described (Figs 1-3). " x_B " in figure should be " x'_B "

Fig. 2: " x'_B " in figure should be " x'_3 " and " $\ln f_A$ " should be " $\ln f_1$ " (in Fig. 3 too)

Fig. 3: " x'_B " in figure should be " x'_4 "

Fig. 4: It is recommended that the compounds in Fig. 4 be specified; may have implications for future applications and understanding robustness of binary vs. ternary approach for specific compounds.