Review of **Machine learning interatomic potentials with accurate long-range interactions for molecular dynamics collision simulations of atmospherically-relevant molecules**

Neefjes et al. evaluate machine-learned interatomic potentials for molecular dynamics simulations of collisions between atmospherically relevant molecules. The study compares PaiNN, AIMNet2, and Δ-learning approaches trained on GFN1-xTB and ωB97X-3c data. The authors show that models relying purely on local atomic environments can struggle to capture long-range interactions relevant for collision dynamics, whereas AIMNet2 performs well due to its explicit long-range treatment. While the importance of long-range interactions in MLIPs is not a new observation, the manuscript provides a useful and well-executed demonstration in the context of atmospheric chemistry. I recommend publication after the following revisions.

Page 5, line 114. "While it may not match the data efficiency or accuracy of PaiNN for geometry-sensitive properties…" Please provide a citation or quantitative evidence supporting this claim.

Page 4. Similarly, the discussion of PaiNN emphasizes its equivariant representation of vectorial properties. Does AIMNet2 include a comparable treatment of directional features, or does it rely on something different? Clarifying this would help readers compare the architectures.

Page 5, line 143. "Instead, it was assumed that the GFN1-xTB PES sufficiently overlaps with the relevant regions of the ωB97X-3c PES." I understand that MD with ωB97X-3c would be expensive. However, the manuscript would be improved by adding the implications of this decision.

In Figure 2 and Table 3, please clarify what the errors are relative to in the captions.

I believe that Figure 2 isn't introduced in the text until after Table 3 is introduced. The authors should adjust the ordering of their manuscript.

Page 10, line 259. "For the H_2SO_4-H_2SO_4 system, performance is consistently low across the entire coordinate." This implies that performance is bad, when I think you mean to say that the errors are low across the entire coordinate.

Page 13, Table 4. It is my understanding that the Δ-PaiNN model here has been trained on GFN1-xTB as the baseline and target reference. I agree with the authors note on page 10 that this should yield an essentially-zero correction. Why then is the full RMSE of PaiNN 0.053 kcal/mol and the full RMSE of Δ-PaiNN 3.4 kcal/mol? I would expect these to be closer. If this difference is due to the "inherent uncertainty associated with finite umbrella sampling" (page 12), then should this have been properly mitigated or evaluated separately? Does this actually justify such a large error?

Page 17, Table 5. The text regarding this figure would benefit from identifying what magnitude of error in the collision rate coefficients is considered acceptable for atmospheric modeling applications.