

Review for "Implications of VOC Oxidation in Atmospheric Chemistry: Development of a comprehensive AI Model for Predicting Reaction Rate Constants" by Xin Zhang, Jiaqi Luo, Wenxiao Pan, Qiao Xue, Xian Liu, Jianjie Fu, Aiqian Zhang, Guibin Jiang

General comments

The study in question presents a new model for the prediction of reaction rate constants of volatile organic compounds (VOCs). The authors used the reaction rate constant dataset by McGillen et al. to train a Siamese message passing neural network (MPNN) to predict these rate constants. The outcoming model was given the name "Vreact" and it was shown to outperform existing models for reaction rate constant prediction.

The dataset used in this study comprises 2802 gas-phase reaction rate constants for 1586 VOCs and 4 oxidants ($\cdot\text{OH}$, $\cdot\text{Cl}$, $\cdot\text{NO}_3$ and O_3). The authors underline this diversity of oxidants as one of their advantages compared to previous models which only use a single oxidant per model. Because of the wide value range of reaction constants, the values were log-transformed. Vreact takes the SMILES string of the VOC and the oxidant as inputs, which is an established and modern approach in chem-informatics. Graph representations are generated from these inputs and fed to the neural network that creates the molecular feature tensors A and B. Further mathematical operations are executed to account for the effects of molecular interactions. Finally, the prediction value for the reaction rate constant is made.

Moreover, the authors evaluate how Vreact can contribute to the understanding of aerosol formation mechanisms. They showcase the oxidation of 2-methyl-4-penten-2-ol, discussing different reaction pathways and how the interaction layer of Vreact can be used for comprehension. Furthermore, the authors gathered more data from 2020 and onwards, which they called the 'post-2020 test set' to analyze the extrapolation ability of Vreact, leading to satisfactory results. Besides, more insights on the reaction rates of specific chemical classes are provided.

All in all, the article presents a modern and sustainable study. The Vreact model that is the key component of this work was built on well-established methods and principles and could overall convince with its performance. Vreact's advantages and improvements towards other models were clearly outlined in a comprehensible way. The study was conducted scientifically correct with no obvious shortcomings. Despite it being a rather data scientific topic, its atmospheric relevance became evident. The illustrations used are helpful and supporting. The supplementary material contains further details on the model architecture and is useful for a deeper understanding. Another valuable resource is the web tool version of Vreact, reinforcing reproducibility and open data.

Specific comments

The revised version does not contain any problematic issues to be discussed here.

Technical corrections

No typing errors or other technical problems were found.