Response to Reviewer

Dear reviewer,

We greatly appreciate the efforts made by you to improve the quality of our manuscript (MS ID: egusphere-2025-1241). We have carefully reviewed and implemented all the comments provided by you and made significant revisions to the manuscript to address the concerns raised. In this response letter, your comments copied verbatim beneath are in black italic font, the author responses are in normal font, revised text is in blue, and line numbers refer to those in the Track Change manuscript.

Zhang and co-authors present a machine learning model for predicting reaction rate constants of VOC–oxidant pairs using a Siamese neural network. The model is novel in its design and application combination, especially in handling multiple atmospheric oxidants. The results demonstrate good predictive performance alongside chemical insight. The results also demonstrate varying performance on the test set depending on which oxidant is considered. The model is tested on an additional external dataset, and is used to make predictions of rate constants for compounds lacking measurements.
From my point of view, the manuscript is generally well-written and clearly structured. However, methodological and interpretative aspects would benefit from clarification to ensure reproducibility and help contextualize the findings. However, I happily recommend it for publication subject to minor revision.
Response: Thank you for your thoughtful and constructive comments. We appreciate your recognition of the significance and quality of our work. We have provided detailed responses and made revisions addressing your comments.

General comments
1.I understand that the major benefit of Vreact is the ability to predict reactivities for multiple oxidants. Could the authors further clarify the motivation for using a Siamese neural network over simpler alternative architechtures which also could provide prediction for multiple oxidants (such as a one-hot encoding of oxidant identity). Given that only four oxidants are included, it would be helpful to understand whether the architecture was chosen for scalability, improved interpretability, or flexibility. Will more oxidants be considered in the future?
Response: Thank you for the insightful comments. As you mentioned, simpler architectures can indeed provide predictions for multiple oxidants simultaneously. We chose the Siamese architecture for the following reasons:
1. Flexibility: The Siamese GNN architecture used in Vreact possesses the flexibility inherent to deep learning, which simple one-hot encoding/machine learning models lack. Because oxidants are treated as molecules rather than abstract categories, the model can leverage structural similarities between known and novel oxidants to transfer learned interaction patterns. This is particularly important in atmospheric chemistry,

where newly identified or understudied oxidants may be structurally or electronically related to those in the training set.

2. Interpretability: The pairwise design enables the extraction of interaction matrices between atoms of VOCs and oxidants, which can be visualized and interpreted (Figure 4). This level of interpretability would not be possible in architectures where the oxidant is reduced to a categorical token, and it provides mechanistic insights into reactive sites and molecular interactions.

3. Scalability: The Siamese GNN architecture of Vreact enhances its scalability. For simple one-hot encoding/machine learning models, designed descriptors/molecular fingerprints are required for the research objects. However, there are numerous reactions in the atmosphere with diverse mechanisms. Requiring a simple architecture to be applicable to non-research objects will affect its interpretability and predictive performance. Therefore, the scalability of a simple architecture is very limited. The Siamese GNN used in Vreact does not rely on predefined descriptors/molecular fingerprints but performs end-to-end modeling. This architecture grants Vreact the ability to expand to other oxidants. Currently, we have only considered four oxidants because these four are widely studied and have sufficient data, which facilitates modeling. In the future, if higher-quality and more extensive datasets become available, we will incorporate more oxidants and update the website in a timely manner.

In the main text, we further elaborate on the limitations of other methods and the advantages of Vreact:

"Despite their utility, these models generally rely on predefined descriptors and are typically limited to reactions with a single type of oxidant, which constrains the scalability of the model." (Line 76 in the revised manuscript)

"Extensions of MPNN, such as the communicative GraphRXN (Li et al., 2023) and directed MPNN Chemprop (Heid et al., 2024), have shown promise in learning reactivity across multiple reactants. They extract the interaction features of chemical reactions in depth, rather than performing simple reactant concatenating. Yet, their application has largely focused on synthesis or materials chemistry, not atmospheric multiphase oxidation." (Lines 88-89 in the revised manuscript)

"Compared to traditional and simple single-oxidant prediction models, Vreact shows significantly improved performance, achieving higher accuracy, stronger interpretability and wider scalability across multiple oxidants. Furthermore, based on the flexibility of the DL architecture, the designed interaction module captures atomic-level interaction patterns, providing mechanistic insights into VOC oxidation process *via* interpretable interaction weight matrices." (Lines 96-98 in the revised manuscript)

2.A brief discussion of quantum chemistry methods to compute these types of rate constants is not mentioned in the background, but could help position this new method in the broader context of rate constant prediction for atmospheric reactions.

Response: Thank you for the comment. We appreciate this suggestion and have now added a concise overview of quantum-chemical (QC) approaches that are widely used to estimate gas-phase rate constants for atmospheric reactions.

The new paragraph (see additions below) highlights both the strengths and limitations of quantum chemistry (QC) methods. This context helps position Vreact as a complementary, data-driven alternative that can deliver near-instant predictions for thousands of VOC–oxidant pairs while retaining mechanistic interpretability. "Traditionally, such predictions have determined either through experimental kinetic modeling methods, (Basant and Gupta, 2018; Liu et al., 2021). which track reactant and product concentrations using techniques such as chemical ionization mass spectrometry (CIMS) and apply kinetic fitting to derive Arrhenius parameters (Logan, 1982; Wells et al., 1996), or through computational methods based on high-level quantum chemical calculations that simulate reaction pathways and energy barriers. However, these methods are time-consuming and cover only a narrow subset of atmospheric VOCs. "…a narrow subset of atmospheric VOCs. QC approaches combine ab initio or density-functional theory calculations with transition-state theory (TST), canonical or variational TST to obtain temperature-dependent rate constants (Canneaux et al., 2014; Liu et al., 2021; Meana-Pañeda et al., 2024). While QC methods offer detailed mechanistic insight, their computational cost scales steeply with molecular size and conformational complexity, limiting routine application to large numbers of VOCs. However, traditional computational methods have shortcomings such as high computational complexity and low efficiency. As a more scalable alternative, QSAR model leverage molecular descriptors and statistical learning. and it has become one of the important methods for evaluating reaction rate constants." (Lines 64-70 in the revised manuscript)

3.Methods and Table S1 suggest that stratified sampling was used to balance oxidant classes across train/validation/test splits. Since the model operates on VOC–oxidant pairs, it is now unclear whether the same VOC can appear in different splits with different oxidants. If so, this could introduce information leakage. Please clarify whether VOCs were kept disjoint across splits.

Response: We thank the reviewer for raising this important concern. In our study, stratified sampling was performed on VOC–oxidant pairs, which means that the same VOC may appear in different data splits when paired with different oxidants. In other words, VOCs were not kept disjoint across splits.

We acknowledge the potential concern regarding information leakage. However, we believe that the current design is appropriate for the following reasons:

1. **Model design focus**: Vreact is trained to model interactions between VOCs and oxidants, not the VOCs alone. Each VOC-oxidant pair represents a distinct chemical reaction, and the underlying mechanisms often vary substantially across oxidants. Thus, each pair can be considered a unique input, and allowing the same VOC to appear with different oxidants across splits does not constitute classical data leakage.

2. **Chemical diversity**: Forcing the same VOC to appear only in one split (*e.g.*, training only) would eliminate the number of VOC-oxidant combinations, reducing the diversity and coverage of the training set for model learning.
3. **Empirical performance**: If leakage were present, it would likely result in inflated test performance. However, as shown in our results, the model maintains strong generalization and extrapolation performance, including on unseen VOCs and external test sets, suggesting that overfitting due to repeated VOCs is not a concern.

To clarify this in the manuscript, we have added the following statement:

"Combinations of the same VOC with different oxidants may appear across the training, validation, and internal test sets." has been added to the main text. (Lines 112-113 in the revised manuscript)

4.In Figure 3G, model performance on the external OH dataset is lower than for O_3, which is the reverse of the trend observed in the internal test set. Could this difference be a result of data quality, compound overlap, or target range?

Response: Thank you for this thoughtful comment. We agree that the inverse performance trend observed between OH and O_3 in the internal versus external (literature) test sets warrants further clarification.

We have carefully examined this discrepancy and found that it is primarily attributable to differences in **chemical space coverage**, rather than data quality or compound overlap alone. Specifically:

All external test sets are sourced from literatures and the data quality is reliable. Any duplicate pairs between the training data and the literature test sets were removed prior to evaluation.

The internal OH test set includes many VOCs with broad representation and strong overlap with the training set, resulting in high performance. In contrast, the external OH test set contains only 36 VOCs, many of which are sparsely distributed near the periphery of the training data distribution (Fig. 3E). This leads to a moderate drop in R^2 despite the model's generally strong performance for OH.

For O_3, however, the internal test set includes structurally atypical compounds that are distant from dense the training data in latent space (Fig. 3F, right region), resulting in a lower R^2. The literature O_3 test set, by contrast, is more clustered and lies closer to the training set in chemical space, allowing the model to achieve higher R^2 despite the small sample size.

This difference is analyzed in the manuscript:

"Notably, Vreact shows opposite performance trends for OH and O_3 between the internal and literature test set. To understand this, UMAP was applied to project compounds from the training, internal, and literature test sets into a shared chemical space. As shown in Fig. 3E, the internal OH test set overlaps well with the training data, leading to consistently strong performance. In contrast, the literature OH set is sparse and scattered near the dataset boundaries. Despite this, Vreact still achieves a high R^2, demonstrating good generalization. For O_3 (Fig. 3F), the internal test set lies farther from the dense training distribution, contributing to lower R^2. Meanwhile, the literature

O_3 set is better aligned with the training data, resulting in higher prediction accuracy. For NO_3 (Fig. 3G), both internal and literature sets show similar distributions, and the model achieves comparable R^2 values (~0.815). Although Vreact underperforms slightly compared to the original single-oxidant model, retraining on the literature data improves performance. This suggests that multi-oxidant training may introduce some noise but does not significantly compromise prediction accuracy.” (Lines 275-291 in the revised manuscript)

The previous Figure 3D-3F has been modified to Figure S2:

Figure S2. The chemical spatial distribution of VOCs in the OH, O_3, and NO_3 datasets used in this study and prior literatures.

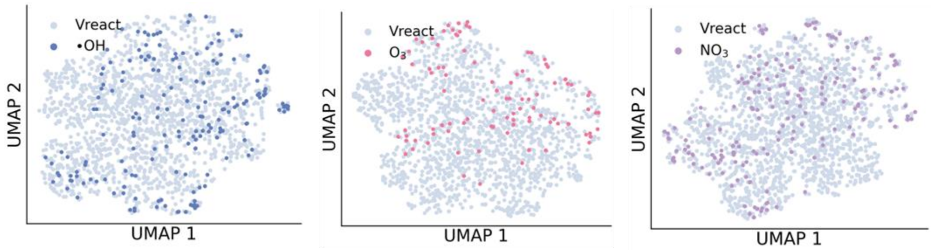
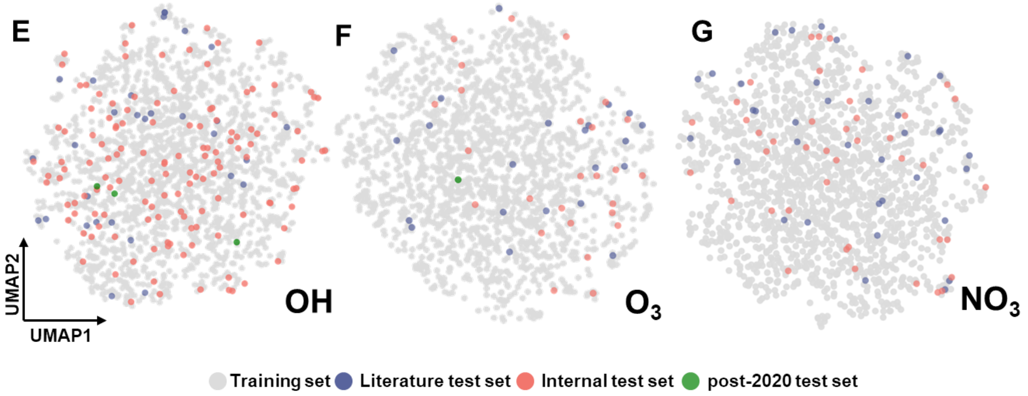


Figure 3D-3F has been modified to Figure 3E-3G:



(E-G) The chemical spatial distribution of VOCs in the OH, O_3, and NO_3 datasets used in this study and prior literatures.

5.Clustering was used to analyze molecular groups and their reactivity (Figures 2E, 3D–F), but details on how these embeddings and clusters were generated are not provided in methods. It would be good with a brief description of how the morgan fingerprint was constructed (which parameters) in the methods. Similarly, UMAP and the SOM methods could be briefly described, along with any hyperparameters, in the methods.

Response: Thank you for your rigorous suggestion. We agree that further clarification of the clustering methodology is necessary and have added relevant descriptions to the Methods section.

The Methods section now includes this information as follows:：

“2.3 Clustering analysis

Morgan fingerprints (radius 2, 1024 bits, generated using RDKit) was used as the molecular embeddings before clustering and visualization. To investigate VOC structural diversity and reactivity trends, two methods were applied: the Self-Organizing Map (SOM) (Kohonen, 2006) and the Uniform Manifold Approximation and Projection (UMAP). The SOM algorithm clustered VOCs into 100 structural groups (10×10 grid), using a sigma of 0.3 and learning rate of 0.5. The UMAP algorithm projected the high-dimensional fingerprint space into 2D for visualization, with the number of neighbors set to 50, minimum distance to 0.6, and metric as correlation." (Lines 150-155 in the revised manuscript)

Additionally, the original text describing SOM clustering in Fig. 2E was revised for clarity:

"The SOM algorithm was used to explore the relationship between VOC structural variation and $\log_{10}k_i$. Each grid in Fig. 2E represents a VOC cluster, and the color gradient indicates reactivity (average $\log_{10}k_i$ values) for the corresponding oxidants. By comparing $\log_{10}k_i$ values across clusters, oxidant-specific reactivity patterns can be assessed." (Lines 190-193 in the revised manuscript)

6.Finally, the manuscripts would benefit from an outlook contextualizing the model's performance by identifying which applications the current accuracy supports and which may require improvement. Relating how performance varies across different oxidants and how this relates to the amount of available data could further emphasize the paper's contribution to understanding data requirements for reliable model accuracy for atmospheric applications.

Response: Thank you for the valuable suggestion. We have added a dedicated outlook section in the revised manuscript to contextualize the model's performance, discuss its applicability across atmospheric chemistry tasks, and highlight the relationship between accuracy and data availability. These revisions are incorporated into the *Concluding* section (Lines 338-351) to clarify the model's application scope and remaining challenges.

"In response to growing concerns about atmospheric pollution and its impact on human health and climate, this study introduces Vreact, a deep learning model designed to predict oxidation rate constants for VOCs with multiple oxidants (OH, Cl, NO_3, O_3). Vreact demonstrates strong overall performance (MSE=0.299, R^2=0.941 on internal test data) and provides mechanistic insights by capturing atomic-level interaction patterns through a Siamese MPNN framework. Its predictive accuracy varies by oxidant, reflecting the availability and diversity of training data. The model achieves high accuracy for OH (R^2=0.929, n=1363) and Cl (R^2=0.913, n=735), supporting robust application in daytime oxidation modeling. In contrast, lower performance is observed for NO_3 (R^2=0.721, n=393) and O_3 (R^2=0.584, n=311), pointing to challenges in modeling oxidants with fewer data and more complex mechanisms. This underscores the importance of expanding high-quality experimental datasets to improve generalization, particularly for underrepresented oxidants and VOC classes.

Vreact supports high-throughput screening for emission inventories and atmospheric reactivity assessments. Its applications span VOC prioritization, emission control planning, and kinetic mechanism development, offering actionable insights for environmental policy and modeling. An interactive web interface (http://vreact.envwind.site:8001) (Fig. S3) enhances accessibility for researchers and policymakers. Further improvements in NO_3 and O_3 predictions will expand its utility in nighttime chemistry and secondary aerosol formation scenarios."

Specific comments

1.Line 29: Add citations on data-driven methods applied to atmospheric chemistry.

Response: Thank you for your valuable feedback. We have added 4 citations on data-driven methods applied to atmospheric chemistry in line 29.

"…Environmental challenges, particularly those associated with atmospheric chemistry and climate change (Chen et al., 2024; Kubečka et al., 2023; Qiu et al., 2023; Zhao et al., 2025),…"

Chen, X., Ma, W., Zheng, F., Wang, Z., Hua, C., Li, Y., Wu, J., Li, B., Jiang, J., Yan, C., Petäjä, T., Bianchi, F., Kerminen, V.-M., Worsnop, D. R., Liu, Y., Xia, M., and Kulmala, M.: Identifying Driving Factors of Atmospheric N2O5 with Machine Learning, Environ. Sci. Technol., 58, 11568–11577, https://doi.org/10.1021/acs.est.4c00651, 2024.

Kubečka, J., Knattrup, Y., Engsvang, M., Jensen, A. B., Ayoubi, D., Wu, H., Christiansen, O., and Elm, J.: Current and future machine learning approaches for modeling atmospheric cluster formation, Nat. Comput. Sci., 3, 495–503, https://doi.org/10.1038/s43588-023-00435-0, 2023.

Qiu, Y., Feng, J., Zhang, Z., Zhao, X., Li, Z., Ma, Z., Liu, R., and Zhu, J.: Regional aerosol forecasts based on deep learning and numerical weather prediction, npj Clim. Atmos. Sci., 6, 71, https://doi.org/10.1038/s41612-023-00397-0, 2023.

Zhao, Y., Zheng, B., Saunois, M., Ciais, P., Hegglin, M. I., Lu, S., Li, Y., and Bousquet, P.: Air pollution modulates trends and variability of the global methane budget, Nature, 642, 369–375, https://doi.org/10.1038/s41586-025-09004-z, 2025.

2.Line 45: "primarily" → "primary."

Response: Thanks for your comment. The "primarily" has been replaced with "primary".

3.Line 48: The phrase "with NO_3 radicals" is repeated—I suggest to remove one instance.

Response: Thanks for your valuable comment. The repeated phrase "with NO_3 radicals" has been replaced.

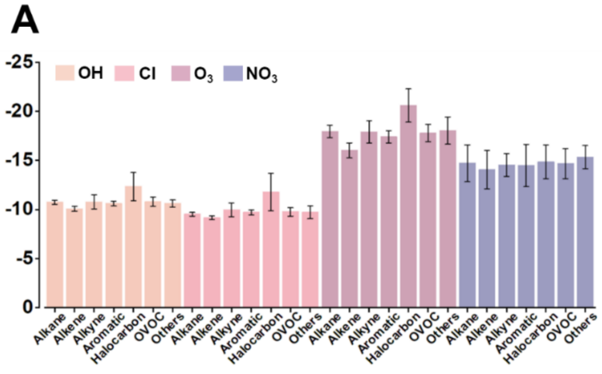
4.Line 48: "the atmosphere's self-cleaning capacity" is ambiguous; consider clarifying. rephrasing or removing.

Response: Thank you for pointing out this problem in manuscript. The "…significantly influencing the spatial and temporal variation of the atmosphere's self-cleaning capacity and the formation of organic aerosols." has been modified to "…significantly influencing the spatial and temporal variation of the formation of organic aerosols."

5.Line 90: Typo— "and and"
Response: Thanks for your constructive suggestion. The repeated "and" has been removed.
6.Line 149: "functional group" could be replaced with "molecular motif" when referring to double bonds.
Response: Thank you for the suggestion. The "including 22 functional groups" has been modified to "including 22 molecular motifs". The "(B) Number of VOCs containing each functional group" has been modified to "(B) Number of VOCs containing each molecular motif".

7.Line 191– It is mentioned in results that MSE is the metric that was used for hyperparameter optimization. This information should also be included in the Methods section for clarity.
Response: We are very sorry for our negligence. We have included the information you mentioned in the Methods section for clarity. The "After identifying the optimal hyperparameter combination (Table S3) on the validation set, and the best model was saved" has been modified to "During hyperparameter optimization, the hyperparameter combination that minimizes the Mean Squared Error (MSE) of the validation set was selected as the optimal hyperparameter combination, and the best model was saved (Table S3)".

8.Improve resolution of Figures 1–5. Figure 5A would be clearer as a conventional bar chart rather than a circular one for better being able to match bar height with y value.
Response: Thanks for you kindly comment. We have improved the resolution of Figures 1-5 and modified Figure 5A into a conventional bar chart.



Finally, we would like to thank you again for your great efforts on improving the quality of this manuscript.

Thank you very much,
Yours sincerely,
Xian Liu