

Response to Editor Comments

Dear Editor,

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.

Thank you very much for your revision. Before the paper can be accepted in ACP, I would like to reiterate on an important point made by reviewer #2, which I believe was not sufficiently addressed and is crucial to show the significance of this work. The original comment read:

"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 architectures 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."

I would prefer if the authors give quantitative results for their newly added statement of "achieving higher accuracy, stronger interpretability and wider scalability".

1. Accuracy: It would really strengthen the paper if the authors could show and quantify, using unbiased calculations (e.g. with same training time), that the Siamese neural network architecture is superior to simpler alternative model architectures. Can the authors show what happens if you train the same type of neural network (just without the Siamese architecture), for each oxidant? Is the result worse? I think it's not enough here to compare with other published models that may have different numbers of neural network hyperparameters, training time, etc.

2. Interpretability: Can the authors show what precisely can be learned about the detailed chemical interactions of two molecules with the Siamese neural network architecture? If not, it should be indicated that this is only a hypothetical feature that was not yet explored.

Response: We thank the reviewer for the careful observation and valuable feedback.

1. Accuracy:

To quantitatively demonstrate the benefits of the Siamese neural network architecture under identical training conditions, we conducted an ablation study in the revised manuscript. In the **Vreact-Ablation** model, we removed the oxidant input and interaction module from the original Vreact architecture, retaining only the VOCs input. This results in a simpler single-input MPNN architecture, while keeping all other settings—including training/validation/test splits, data source (McGillen et al., 2020), hyperparameters, training time, and evaluation metrics—

identical to those of Vreact. We trained both models on individual oxidant datasets (OH, Cl, NO₃, O₃) as well as on the combined dataset. As shown in **Fig. 3D**, Vreact consistently outperformed Vreact-Ablation across all oxidants.

- The improvements of R^2 for OH, Cl, NO₃, and O₃ are 0.049, 0.113, 0.184, and 0.021 respectively.
- When evaluated on the full dataset (all oxidants), Vreact-Ablation achieved an R^2 of only 0.035, indicating that the simple MPNN without oxidant information fails to generalize beyond single-oxidant learning.
- Moreover, their runtime per iteration is essentially consistent.

We have added the following content:

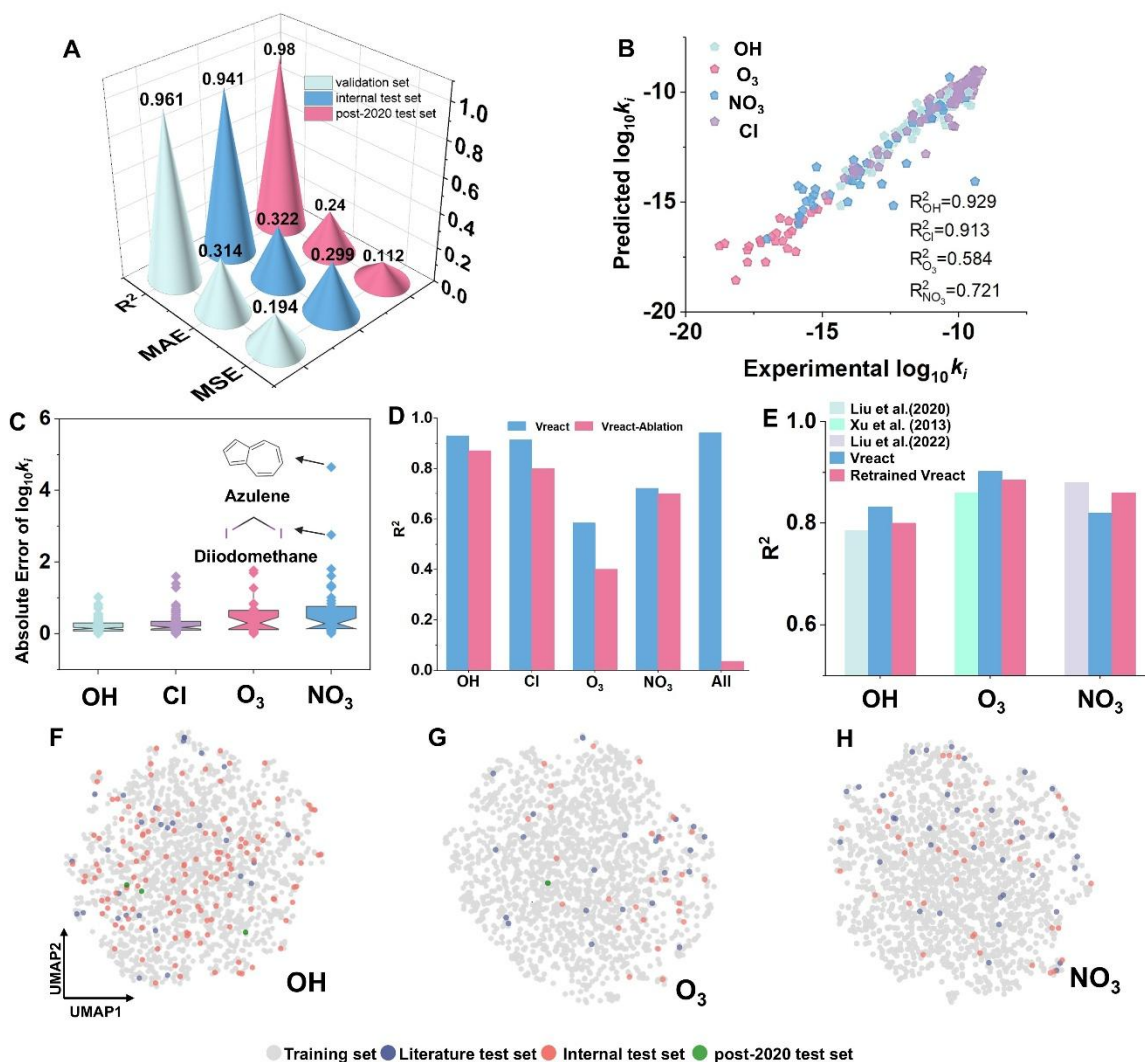


Figure 3. Evaluation and comparison of the predictive performance of the Vreact model. (A) MSE, MAE, R^2 of Vreact (trained on the McGillen et al. dataset) on the validation set, internal test set, and external post-2020 test set. (B) R^2 values for $\log_{10}k_i$ predictions of four oxidants' reactions in the internal test set. (C) Distribution of AE between predicted and experimental $\log_{10}k_i$ values for the four oxidants in the internal test set. (D) R^2 of the Vreact and Vreact-Ablation on the OH, Cl, NO₃, O₃, and combined test sets. (E) R^2 comparison among previously published single-oxidant models, the original Vreact (evaluated on cleaned

literature test sets), and Retrained Vreact (trained and tested using the same original splits as the literature) highlighting adaptability. (F-H) The chemical spatial distribution of VOCs in the OH, O₃, and NO₃ datasets used in this study and prior literature.

“3.3 Model Ablation Study

To evaluate the contribution of the Siamese neural network architecture in Vreact, we performed an ablation study. In the ablation model (Vreact-Ablation), the oxidant input and interaction module were removed, leaving only the VOC input. Both Vreact and Vreact-Ablation were trained, validated, and tested on the OH, Cl, NO₃, O₃, and combined datasets. All experimental settings were kept consistent, including data sources (McGillen et al., 2020), hyperparameters and evaluation metrics. As shown in Fig. 3D, Vreact consistently outperformed Vreact-Ablation across all four oxidants, with R² improvements of 0.049 (OH), 0.113 (Cl), 0.184 (NO₃), and 0.021 (O₃). When evaluated on the combined dataset, Vreact-Ablation achieved an R² of only 0.035, indicating that it fails to generalize across multiple oxidants. Additionally, both models showed comparable runtime per iteration. These results demonstrate that, under the same training conditions, the Siamese MPNN architecture significantly enhances predictive performance and generalization. By explicitly modeling VOC-oxidant interactions, the architecture enables the network to capture shared patterns across reaction types, thereby improving its practical applicability in multi-reactivity prediction.”

2. Interpretability:

Regarding interpretability, Vreact’s architecture was specifically designed to capture atomic-level interactions between VOCs and oxidants. The model incorporates an interaction module that learns cross-molecular attention weights, enabling insights into which atoms in each molecule contribute most to reactivity.

In Section 3.4, we present visualization and analysis of these interaction weights, illustrating how the model differentiates chemical behaviors of VOCs toward different oxidants. While still preliminary, this direction demonstrates the architectural potential for interpretable chemical learning. We have clarified that this aspect is explicitly explored, not merely hypothetical.

l.64: "QC approaches combine ab initio or density-functional theory calculations ..." - DFT is generally considered an ab initio method (though not a first-principles-method).

Response: Thank you for your comment. “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” has been modified to “QC approaches use density-functional theory calculations such as transition-state theory (TST) or variational TST to obtain temperature-dependent rate constants”.

l.90: "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."

- Can the authors explain what they mean with the first sentence?
- Second sentence: this study does not look at multiphase chemistry, either.

Response: Thank you for the comments. The following is an explanation of your comments one by one.

- First sentence: We are very sorry that this sentence has caused you confusion. We would like to show through two examples that GNN and especially MPNN have been deeply applied when modeling chemical reactions. These two methods can transform molecules into molecular graphs, then input them into MPNN for feature depth extraction, and finally splice the features for reaction prediction. They do not perform simple concatenation for prediction by converting reactants into molecular fingerprints/descriptors. The MPNN provides more chemical information for chemical reaction modeling. So, the “They extract the interaction features of chemical reactions in depth, rather than performing simple reactant concatenating” has been modified to “Compared with the simple concatenation using molecular fingerprints/descriptors, they all use MPNN to deeply extract task-relevant representations of chemical reactions, provide abundant chemical information for subsequent reaction modeling, and achieve good prediction results”.
- Second sentence: We apologize for the inappropriate expression of the second sentence. Our study indeed does not look at multiphase chemistry. The “atmospheric multiphase oxidation” has been modified to “atmospheric oxidation reaction”.

l.98: "Vreact shows significantly improved performance"

- Please provide the level of significance for this statement

Response: Thank you for the valuable suggestion. In the ablation experiment, Vreact outperforms Vreact-Ablation in predictive performance across all four oxidants, with R^2 differences between the two models on OH, Cl, NO_3 , and O_3 being 0.049, 0.113, 0.184, and 0.021, respectively. Additionally, when both models are tested using the combined dataset, Vreact-Ablation achieves an R^2 of only 0.035, indicating that it lacks predictive capability. Vreact shows significantly improved performance compared to simpler alternative architectures. Then, compared with the simple single-oxidant prediction models reported in three literature, Vreact exhibits predictive performance that is either superior to or comparable with those models. Vreact did not demonstrate significantly improved prediction performance on the NO_3 literature test set.

We have removed the imprecise statement: “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”.

Add new statements:

“The dual-input design of Vreact enhances scalability and generalization across multiple oxidants. Ablation experiments show that Vreact significantly outperforms a structurally simpler single-input MPNN trained under identical conditions. The interaction module within Vreact provides atomic-level attention maps that offer

mechanistic insights into VOC-oxidant reactivity patterns, improving interpretability.”
(Lines 97-100)

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

Thank you all,
Yours sincerely,
Xian Liu