Supplement of Hydrol. Earth Syst. Sci., 30, 6843–6865, 2026
https://doi.org/10.5194/hess-30-6843-2026-supplement
© Author(s) 2026. CC BY 4.0 License.



Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

A physically integrated GNN surrogate for microbially-mediated kinetic reactive transport: PHYGNET

Jinbo Wang et al.

Correspondence to: Mingzhu Liu (liumz@cugb.edu.cn)

The copyright of individual parts of the supplement might differ from the article licence.

Sect. S1 Temporal Evolution of Spatial Prediction Residuals in Autoregressive RTM Simulations

Since the ML surrogate model is trained on single-step state transitions (Δt = 1 h) but performs continuous, autoregressive sequential predictions for the 240-hour reactive transport model (RTM), it is crucial to evaluate whether the prediction errors accumulate and diverge over time. To quantify this temporal evolution, we mapped the spatial residuals at specific intervals: t = 10, 50, 100, 150, 200, and 240 h.

As illustrated in Fig. S1 to Fig. S7, the autoregressive errors do not diverge; rather, they exhibit bounded saturation. For the concentration variables (e.g. HC, NO_3^-, O_2), throughout the 240-hour continuous sequence, the maximum spatial residuals consistently fluctuate within the order of 1e-5 to 1e-6 mol/L. This indicates that the iterative feedback loop does not induce exponential mass drift.

For the highly non-linear thermodynamic variables, pH and pe, the errors display a saturation behavior rather than continuous amplification. For example, the residual error for pH experiences slight local increases at intermediate steps (reaching a maximum of approximately 0.802 at t = 200 h) but stabilizes at a lower bound (maximum 0.551) by the end of the simulation at t = 240 h.

Furthermore, the spatial distribution of these residuals remains highly localized throughout the time series. Both concentration and thermodynamic errors are tightly confined to the fracture intersections. These regions inherently feature the steepest concentration gradients and the most complex multi-species mixing dynamics. The ability of the surrogate to restrict errors to these highly localized, physically complex zones—without allowing them to propagate and corrupt the entire domain—further confirms the stability and robustness of the autoregressive predictions over extended simulation scales.

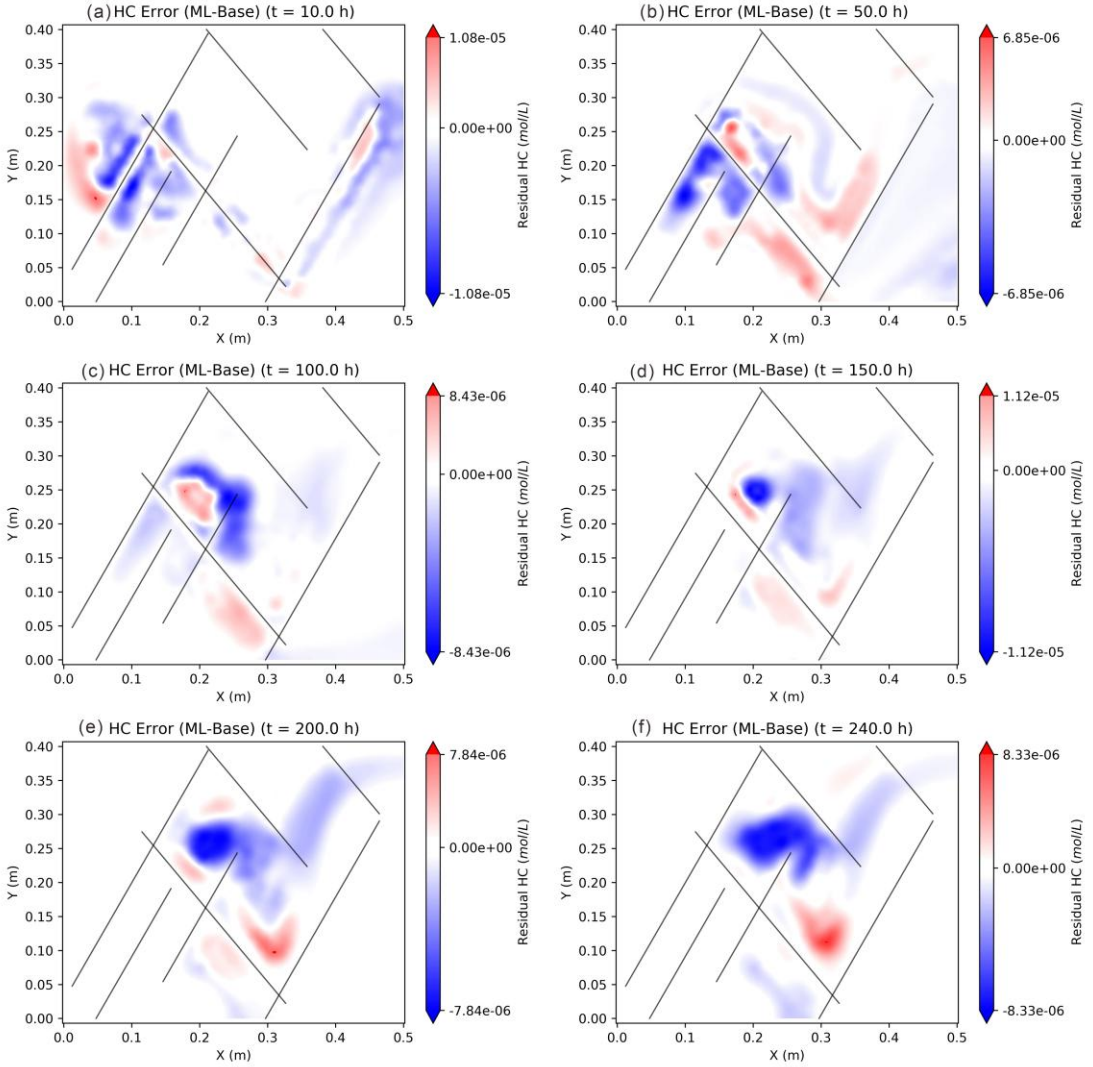


Figure S1. Temporal evolution of spatial residuals for HC concentration.

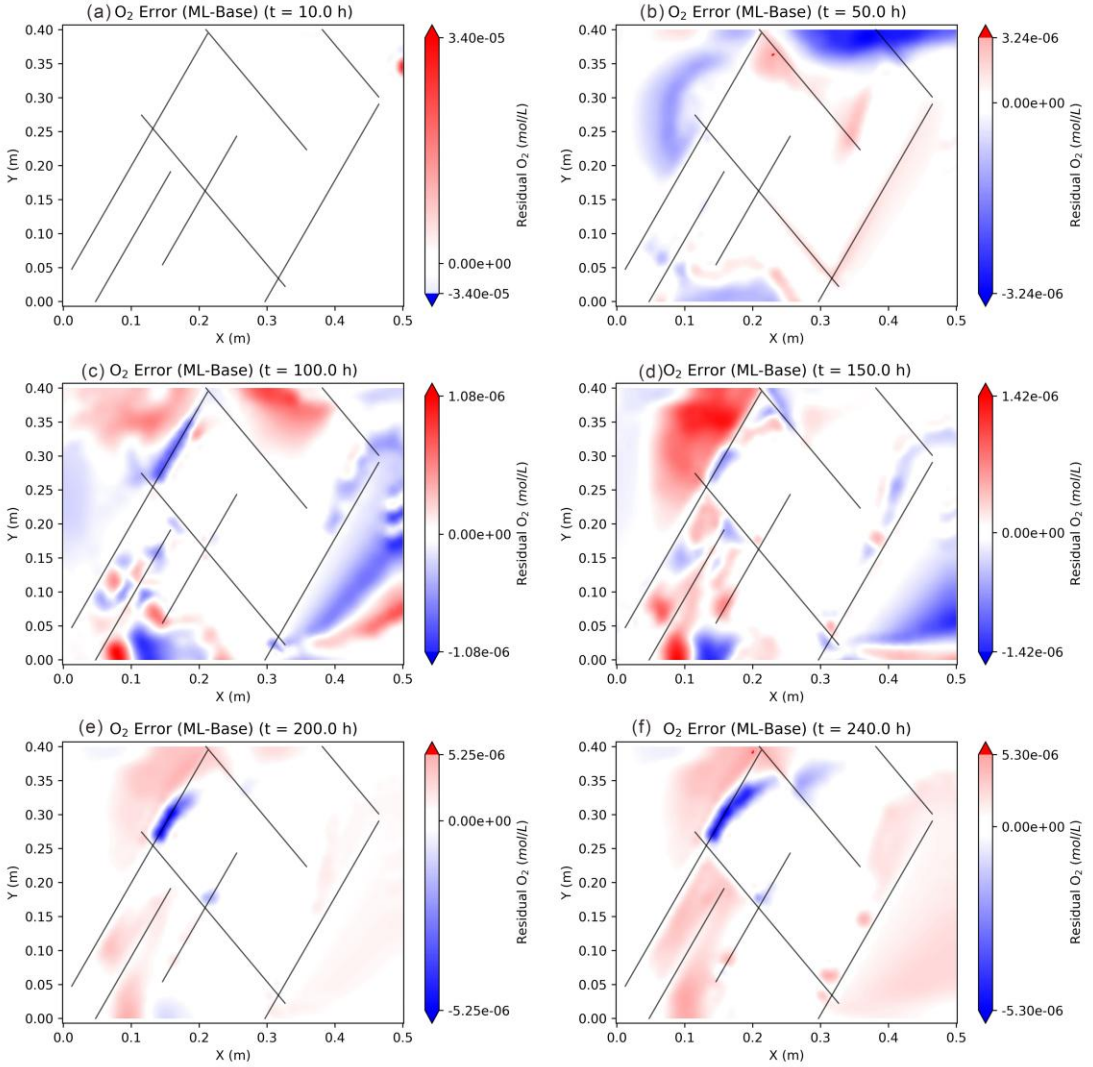


Figure S2. Temporal evolution of spatial residuals for O_2 concentration.

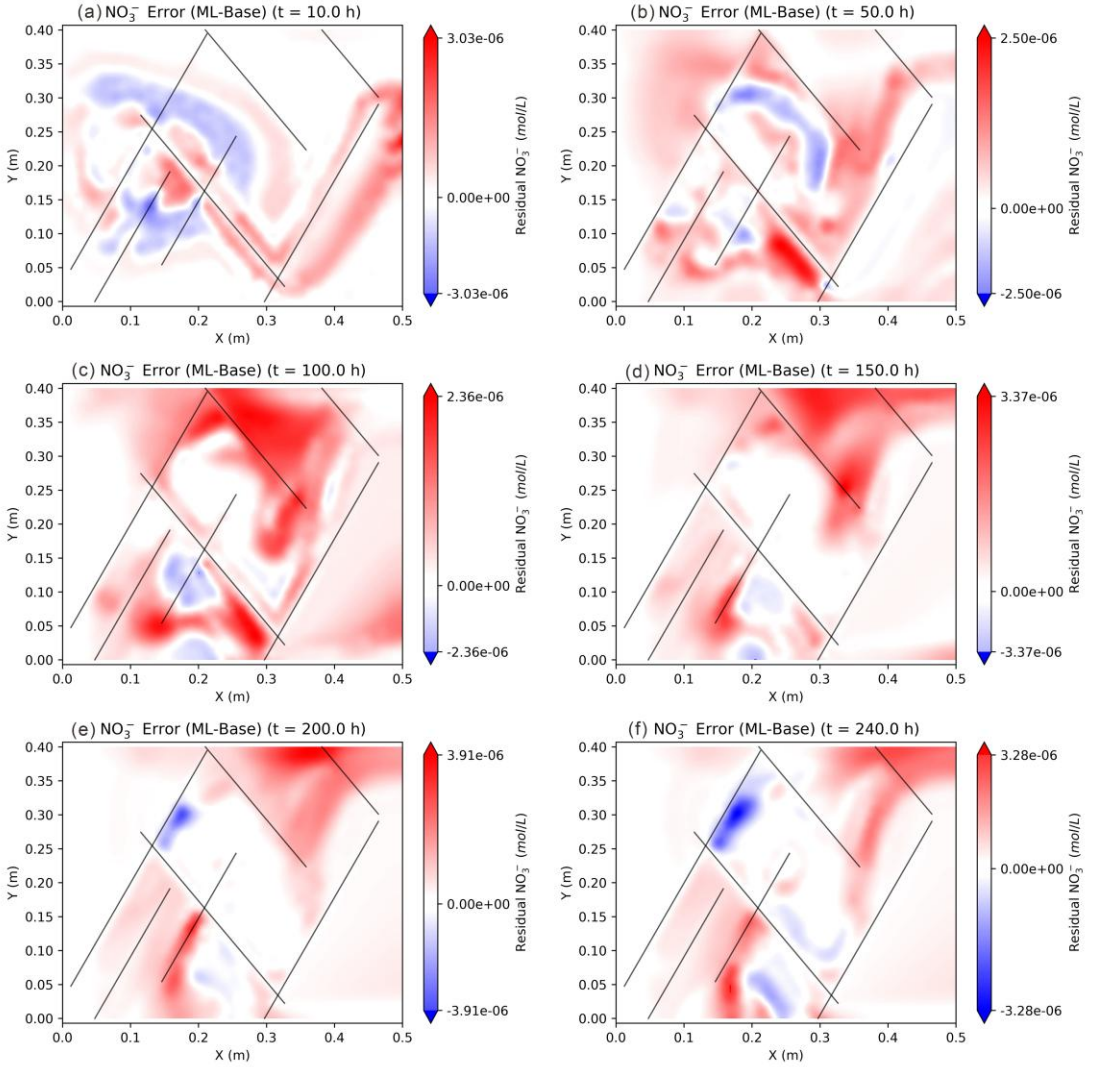


Figure S3. Temporal evolution of spatial residuals for NO_3^- concentration.

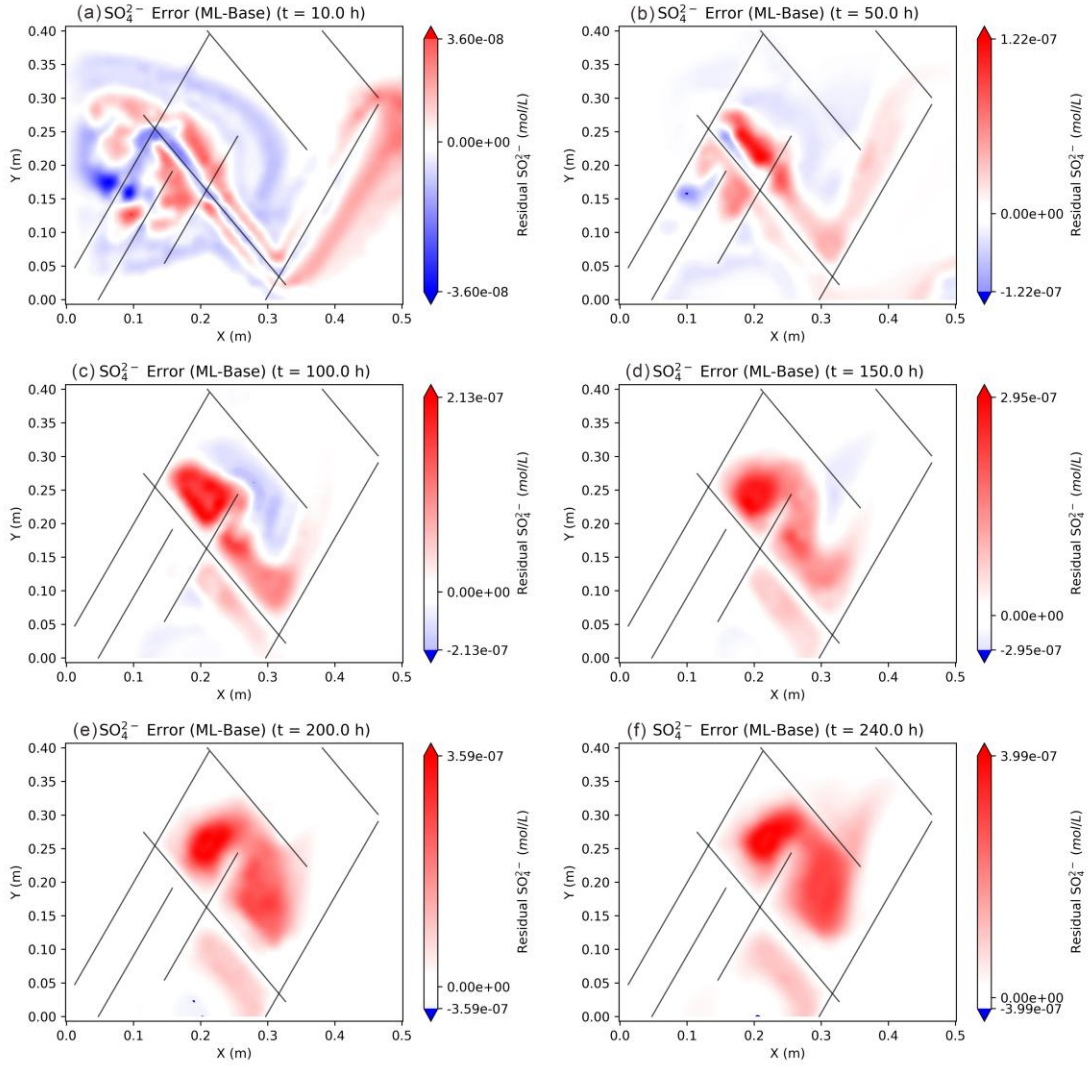


Figure S4. Temporal evolution of spatial residuals for SO_4^{2-} concentration.

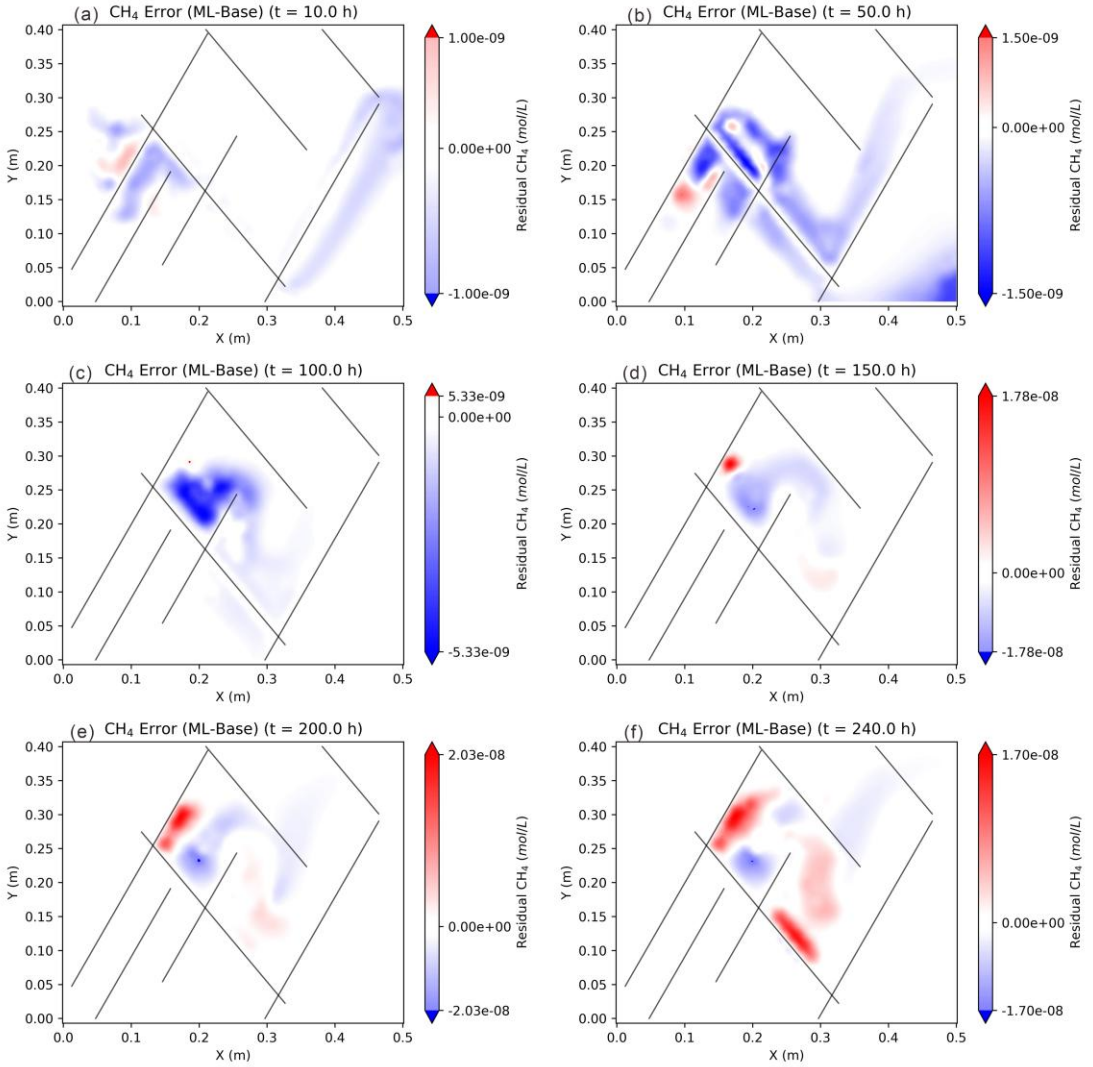


Figure S5. Temporal evolution of spatial residuals for CH_4 concentration.

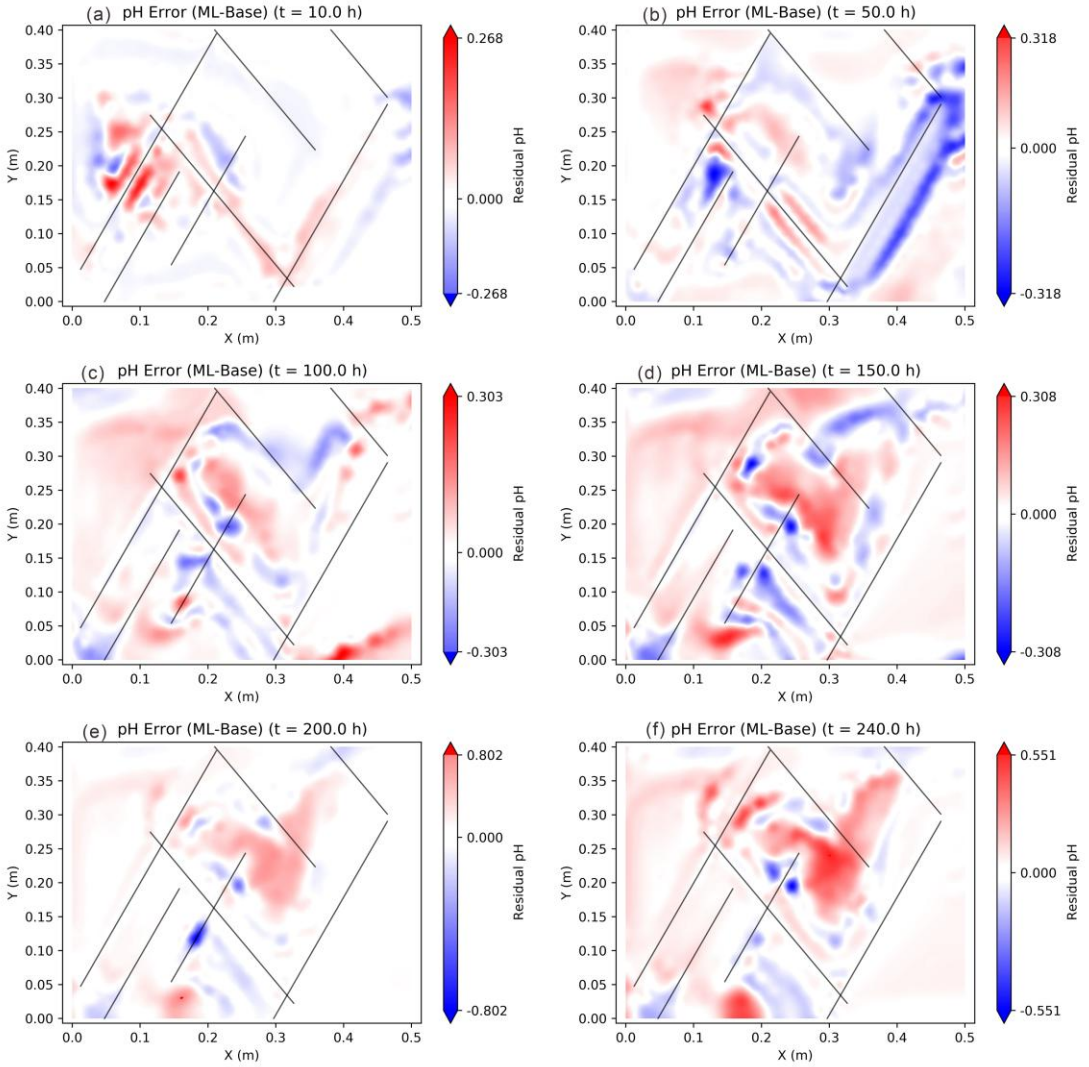


Figure S6. Temporal evolution of spatial residuals for pH concentration.

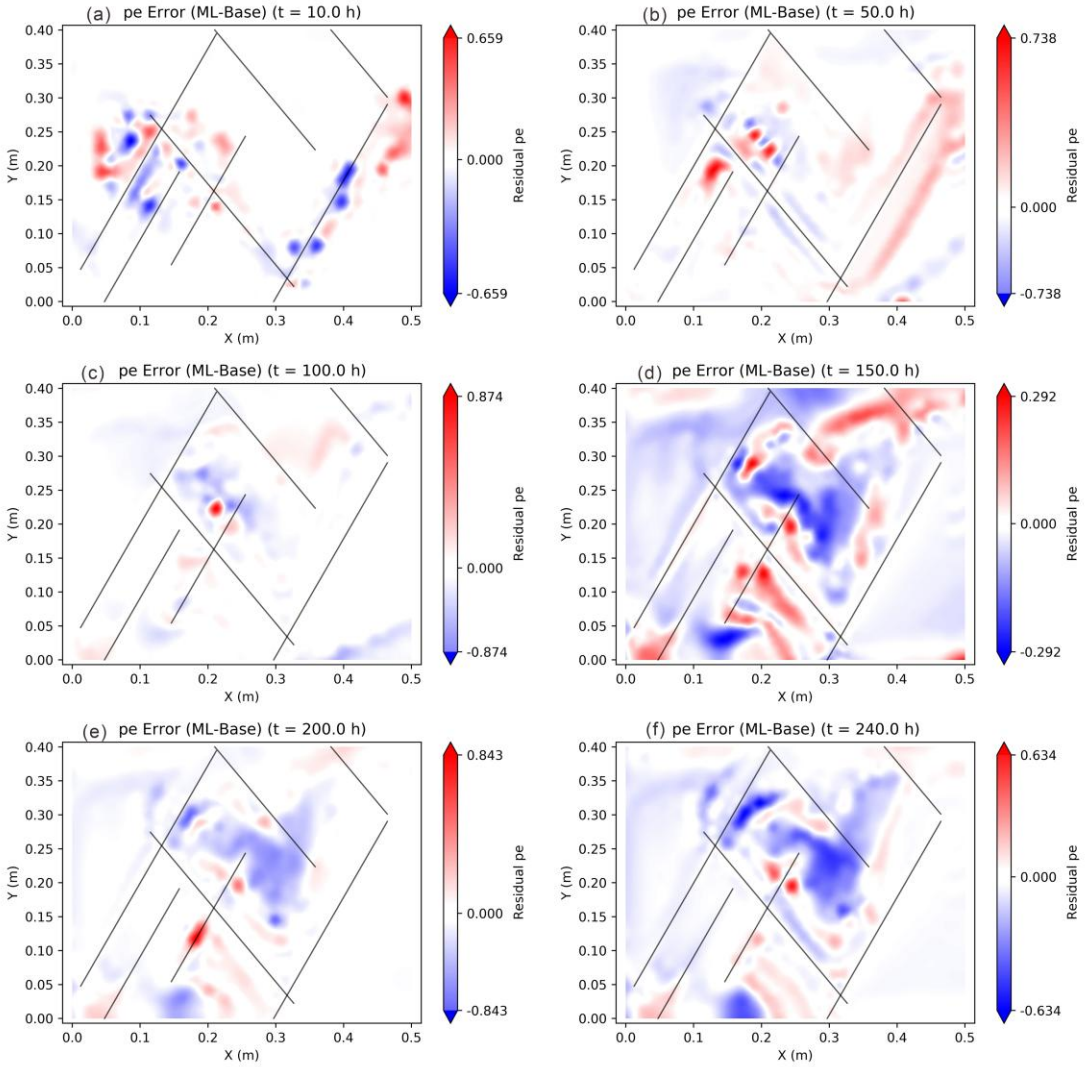


Figure S7. Temporal evolution of spatial residuals for pe concentration.

Sect. S2 Sensitivity Analysis of Recalibration Frequencies and Future Strategies

To quantitatively evaluate the long-term robustness of the hybrid CPqPy workflow and determine the optimal frequency for geochemical intervention, sensitivity tests were conducted across five different recalibration intervals: 5, 10, 15, 20 steps and without calibration.

As can be clearly observed from Fig. S8h, regardless of the calibration frequency used, the total RMSE exhibits a "cliff-like" drop after each physical calibration intervention, rapidly falling to an extremely low level close to zero. This phenomenon intuitively demonstrates that periodically introducing full physical calculations to update the system state is extremely effective; it forces the elimination of minute deviations generated during the neural network's forward propagation, thereby maintaining the model's high fidelity over the long term.

Moreover, A comparison of the vertical scales in Figs. S8a–e reveals that the spatial RMSE for the major chemical species generally remain at extremely low levels (10^{-5} to 10^{-2}). In contrast, the RMSE for pH (Fig. S8f) and pe (Fig. S8g) is at the highest order of magnitude (reaching the 10^{0} order of magnitude), and is the primary source of fluctuations in the total RMSE shown in Fig. S8h. As analyzed in Text S1, this imbalance in error distribution is determined by the network's underlying architecture. This disparity in the strength of physical constraints leads to pH and pe which are more data-driven being more prone to accumulating offsets during multi-step autoregressive simulations. It is also worth noticed that, although variables lacking explicit physical constraints can lead to the accumulation of errors, these errors are strictly limited to a controllable range. Observing the evolution of the "PHYGNET Only" curve (purple) in the Fig. S8 reveals that even under extreme test conditions with absolutely no calibration intervention, its total RMSE could still be stabilized at a specific order of magnitude (around 4.0) during the middle to late stages of the simulation, without experiencing severe numerical explosion or infinite divergence.

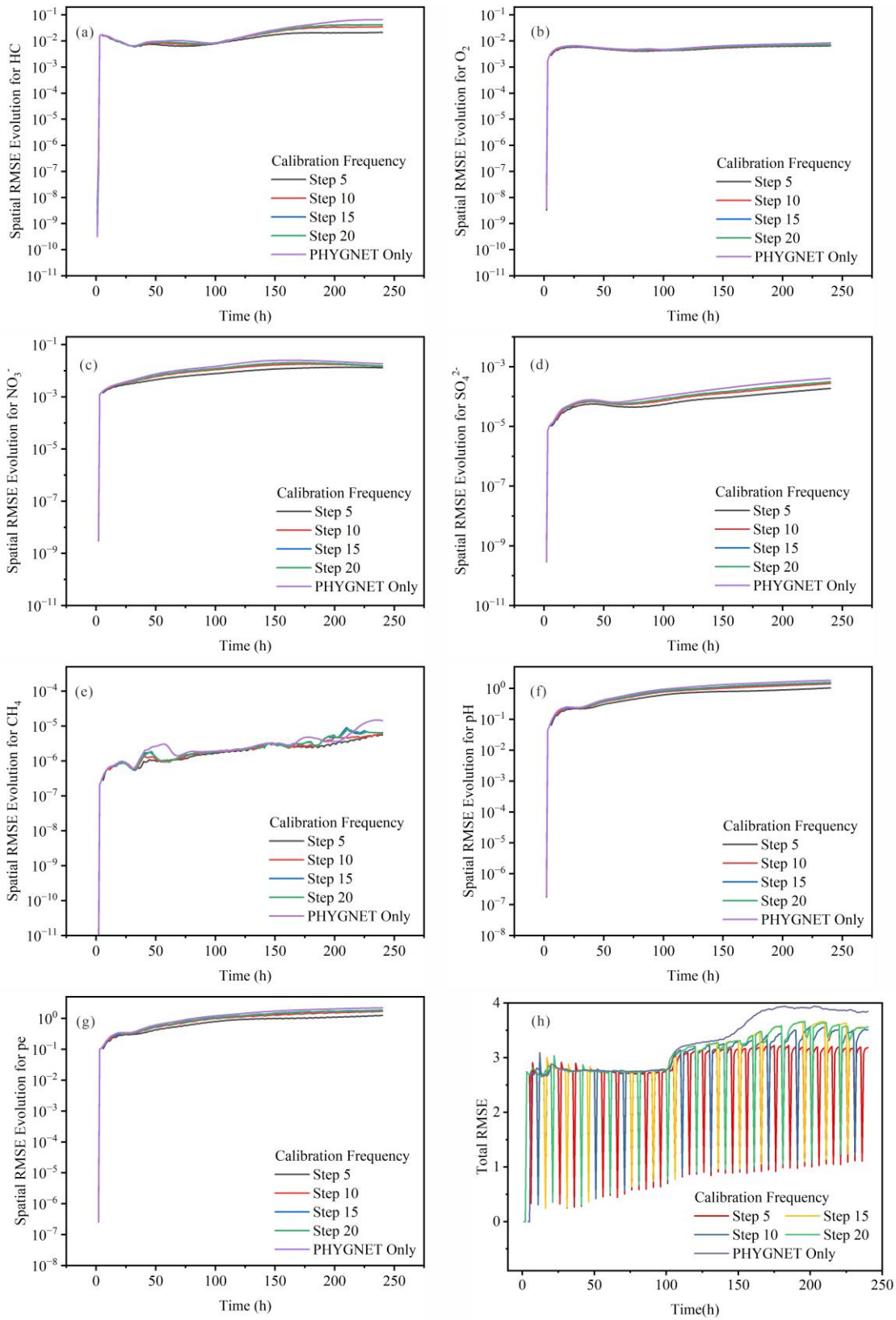


Figure S8. Spatial RMSE evolution under different calibration frequencies. (a)-(e) are the evolution of RMSE for each species, (f)-(h) are the evolution of RMSE for environmental conditions while (h) shows the total RMSE evolution under different calibration strategies.

Furthermore, while the fixed 10-step interval provides a robust baseline for current simulations, it presents inherent limitations when applied to highly heterogeneous fractured aquifers characterized by steep reaction fronts. In such complex domains, fixed intervals can lead to redundant computations in smooth transport zones while delaying necessary error corrections in steep-gradient areas. Therefore, as a crucial future improvement, this framework will benefit significantly from an adaptive recalibration strategy. Instead of relying on predefined steps, future iterations of the model could dynamically trigger the full-physics solver based on real-time spatial residual variability or abrupt chemical gradients across the graph network, thereby maximizing both accuracy and computational resource allocation.