Response to reviewers' comments on "Evaluating Long-Term Effectiveness of Managed Aquifer Recharge for Groundwater Recovery and Nitrate Mitigation in an Overexploited Aquifer System" by Y. Zhu, Z. Guo, S. Wan, K. Chen, Y. Wang, Z. Zeng, H. Shen, J. Ye, and C. Zheng

Reviewer's comments in black; Response to reviewer's comments in blue; Revisions in the revised manuscript in red.

We would like to thank the editor and the reviewer for their constructive comments, which have helped us improve the presentation of this work. We have revised our manuscript according to the reviewer's comments and have provided below a point-by-point response to the reviewer's comments.

Reviewer #2

This manuscript presents a regional-scale assessment of MAR impacts on groundwater level recovery and nitrate dynamics in the Xiong'an New Area using a three-dimensional coupled flow and reactive transport model. The topic is relevant for managing groundwater depletion and nitrate pollution in intensively cultivated regions. The integration of MAR hydraulics with nitrate biogeochemistry is an important contribution, and the authors provide clear insights into how dilution, denitrification, and geological heterogeneity jointly shape water-quality outcomes. The discussion is generally well-structured and highlights both the strengths and limitations of MAR. However, several aspects require clarification and improvement before the manuscript can be considered for publication. The model structure, boundary conditions, and parameterization of reactive processes need more transparency, particularly regarding organic carbon availability, redox controls, and the justification for simplifying nitrogen pathways. Some interpretations appear to overstate the certainty of denitrification estimates given the strong assumptions applied. While the authors have mentioned the influence of spatial resolution, recharge configuration, and heterogeneity, this part would benefit from more rigorous sensitivity analyses. The discussion could also more explicitly connect findings to broader MAR design principles and regional management implications. Furthermore, I suggest the authors discuss how their exclusion of other nitrogen transformation reactions might affect the conclusions regarding denitrification's minor role.

Overall, the study has clear potential, but refinements in methodological justification, uncertainty communication, and the contextualization of results will substantially strengthen the manuscript. Therefore, I recommend minor revisions.

Response:

Thank you for your encouraging feedback on the value of this work. We greatly appreciate your insights

and recommendations, which have notably improved the clarity and depth of our research. In response to your comments, we have made the necessary revisions. Below, we provide a point-by-point response to each comment and indicate the corresponding changes in the revised manuscript.

1. Line 16: The abstract states that "regional-scale hydrological and geochemical impacts remain insufficiently understood," but several recent MAR modeling papers at regional scales exist. It is unclear what is truly new: the model structure? the scale of the simulation? integrating heterogeneity? long-term simulation horizon? nitrate processes?

Response:

Despite recent advances in managed aquifer recharge (MAR) research using regional-scale models, studies that explicitly incorporate multi-component reactive transport processes under complex geological field conditions remain scarce. Accordingly, we have revised the abstract to more clearly articulate the specific contributions of this work.

Revisions have been made in line l6:

"Recent riverine MAR operations have shown significant local groundwater recovery; however, the long-term regional fate and spatial evolution of nitrate remain poorly quantified. In particular, it remains challenging to assess how geological heterogeneity interacts with biogeochemical processes to control remediation efficacy."

2. Line 13-24: Currently the abstract provides interpretation but no numbers, making the conclusions feel generic. Consider adding a sentence on the role of heterogeneity in shaping nitrate reduction patterns.

Response:

Thank you for this helpful comment. We have added a sentence on the role of heterogeneity in shaping nitrate reduction patterns, making the conclusions more specific and informative.

Revisions have been made in line 15:

"The results indicate that MAR leads to a basin-wide mean groundwater level rise of 1.11 m, with a maximum increase of 7.5 m near the river. Nitrate reduction is dominated by physical dilution (~91%) rather than denitrification (~9%). Furthermore, geological heterogeneity governs the spatial variability of water quality evolution by channeling flow through preferential pathways, which creates localized reduction hotspots, despite having a minimal impact on the total nitrate mass removal."

3. Line 33: "become as"

Response:

We have corrected the error by removing the redundant word "as" in the revised manuscript.

Revisions have been made in line 33:

"Amid these challenges, groundwater quality degradation, specially salinization and nitrate pollution, has become a pressing concern in many arid regions."

4. Line 113: The rate expression multiplies terms for electron donor (ED), terminal electron acceptor (TEA) and inhibition. However, there is no mention of an explicit microbial biomass pool (or active biomass concentration). Denitrification in aquifers is often biomass-mediated and can be substrate-limited and biomass-limited; excluding biomass (or at least an active biomass term) risks mischaracterizing dynamics, especially under long-term MAR where biomass may grow or be flushed

Response:

We sincerely appreciate the reviewer's insightful comment. We fully agree that denitrification is fundamentally a biomass-mediated process and that, in reality, reaction rates can be constrained by the active biomass concentration, particularly during the transient phases of microbial growth or decay under long-term MAR conditions. We employed an effective Monod-type formulation in this study, assuming a quasi-steady effective active biomass at the regional scale rather than explicitly simulating a dynamic microbial biomass pool. This decision is based on the following considerations:

(1) This is a regional-scale simulation (>500 km²). Introducing a dynamic biomass variable would require defining multiple unconstrained parameters (e.g., initial biomass distribution, yield coefficients, decay rates) that are unknown at this scale. Recent studies (Schäfer Rodrigues Silva et al., 2020) indicate that adding such unconstrained parameters often increases model uncertainty without significantly improving predictive capability for regional mass balance.

(2) Theoretical advances in reactive transport modeling suggest that at the macro-scale, complex micro-scale biological dynamics often collapse into effective kinetics driven by mixing and substrate availability. For instance, Le Traon (Le Traon et al., 2021) demonstrated that upscaled reaction rates are often controlled by physical transport limitations rather than intrinsic microbial physiology.

We have added a clarifying paragraph after Equation (5) to justify the simplified kinetic approach. While acknowledging this as a current limitation, we have further discussed the assumption in the revised manuscript. We agree that incorporating dynamic microbial growth is an important direction for future research, particularly for localized simulations near infiltration points, and will prioritize this in subsequent studies. We thank the reviewer again for highlighting this key mechanism, which has helped refine the boundary conditions and limitations of our present model.

Revisions have been made in line114:

"It is important to note that an explicit dynamic microbial biomass pool was not simulated due to the lack of spatial data for microbial parameters at the regional scale. Instead, a quasi-steady-state biomass concentration is assumed to be present in the aquifer sediments. Therefore, the effect of biomass abundance is implicitly incorporated into the effective rate constant (μ_{max}). This simplification avoids the high parametric uncertainty associated with unconstrained biological parameters (Schäfer Rodrigues Silva et al., 2020) and aligns with theoretical frameworks for effective kinetics in heterogeneous media (Le Traon et al., 2021)."

Revisions have been made in line 374:

"Furthermore, the current model assumes a constant microbial capacity (μ_{max}), neglecting potential biomass growth or washout during long-term recharge. It may overlook local biogeochemical dynamics,

such as biofilm development near infiltration zones, which warrants detailed investigation in future fine-scale studies."

5. Line 117: The Introduction criticizes prior studies for omitting microbial OM degradation and C–N coupling, but the method section does not convincingly demonstrate that these processes are represented.
    Response:
    Unlike many previous regional assessments that simplify denitrification as a first-order decay process, our model employs dual-Monod kinetics Equation (5). This formulation mathematically links the nitrate reduction rate to the concentration of the electron donor (Acetate). This kinetic dependence constitutes the actual "coupling" of Carbon and Nitrogen dynamics in our framework. Furthermore, we utilize Acetate (CH_3COO^-) as a representative proxy for bioavailable Dissolved Organic Carbon (DOC). The stoichiometric reaction in Table 1 explicitly simulates the consumption (degradation) of this organic matter concurrent with nitrate reduction.
    Revisions have been made in line 52:
    "Previous studies often simplify denitrification as a first-order decay process, thereby neglecting the kinetic coupling between nitrate reduction rates and the availability of electron donors."

6. Line 123: The manuscript specifies that the model domain is discretized using 100 m horizontal grid spacing but does not provide justification for this choice. Given the scale of the study area (tens of kilometers) and the strong emphasis on preferential flow paths, MAR-induced infiltration fronts, and heterogeneity, did the authors perform mesh and time-step sensitivity analyses (e.g., Courant number assessment or grid refinement tests) to ensure numerical stability and accuracy of both flow and reactive transport simulations?
    Response:
    We thank the reviewer for this valuable comment regarding grid resolution and numerical accuracy. To address this concern, we performed a grid resolution sensitivity analysis using three horizontal grid spacings (50 m, 75 m, and 100 m), while keeping all hydraulic parameters, boundary conditions, and recharge configurations identical.
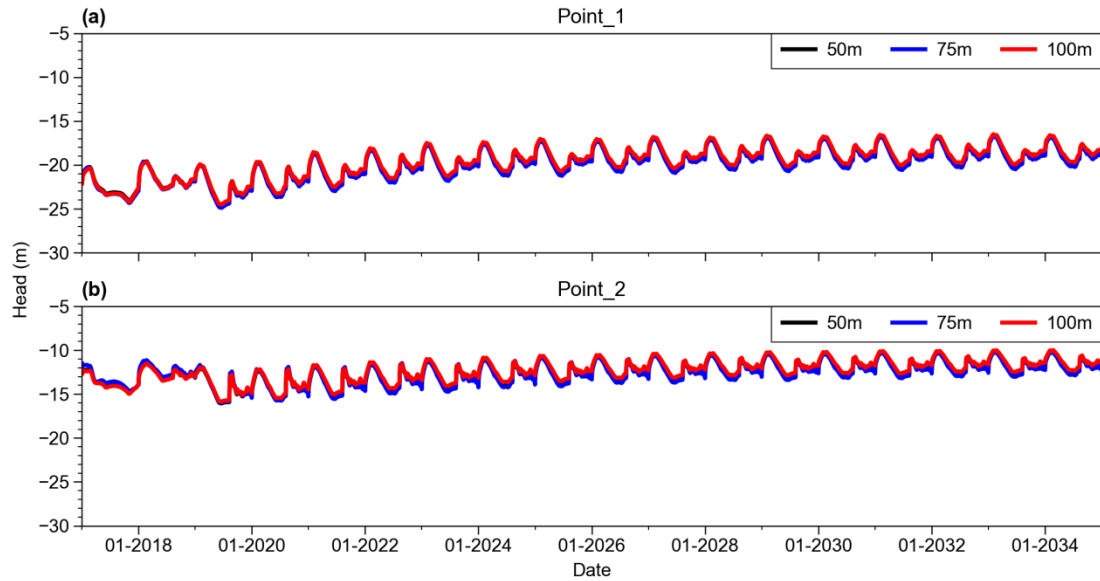
Figure R1. Grid-resolution sensitivity of simulated groundwater heads at representative points.: (a) Point 1 and (b) Point 2 using three horizontal grid resolutions (50 m, 75 m, and 100 m).

The sensitivity analysis focused on the temporal evolution of the simulated groundwater head at representative observation points used in the manuscript. Figure R1 presents a comparison of groundwater head time series at two representative locations (Point 1 and Point 2) under the three grid resolutions. As shown in Fig. R1(a–b), the simulated groundwater head responses exhibit similar temporal patterns and comparable magnitudes across the tested grid spacings, indicating that the regional-scale hydraulic response is robust with respect to grid resolution within this range. Based on these results, the 100 m horizontal grid spacing was adopted for the full set of simulations, as it provides a reasonable balance between computational efficiency and spatial resolution for a regional-scale groundwater flow study covering a large domain.

Regarding temporal discretization, PFLOTRAN employs an adaptive time-stepping scheme for groundwater flow simulations, which automatically adjusts time steps to maintain numerical stability. No numerical instability or nonphysical oscillations were observed in the groundwater head simulations across the different grid resolutions.

Revisions have been made in line 123:

"The model domain (33 km × 26.9 km × 135 m) was discretized using a horizontal grid spacing of 100 m and three vertical layers with thicknesses of 15 m, 40 m, and 80 m, representing the vadose zone, the phreatic aquifer, and the underlying unit, respectively. A grid-resolution sensitivity test was conducted to ensure the numerical robustness of the groundwater flow simulation, based on which the 100 m horizontal grid spacing was adopted for the full set of simulations."

7. Line 129: Only four lithofacies are represented based on 14 boreholes. This is a very sparse conditioning dataset for a 3D domain of this size. The uncertainty introduced by such limited conditioning should be acknowledged.

"The geological heterogeneity was characterized based on a sparse borehole dataset, inevitably introducing structural uncertainty in the delineation of localized contaminant migration, although its impact is partially mitigated through stochastic ensemble simulations."

8. Line 203: How sensitive are these values to recharge rates or hydraulic conductivity? The results repeatedly attribute attenuation to "distance and connectivity," but this remains qualitative. A clearer linkage between MAR response zones and the permeability map is needed.

Response:

We appreciate the reviewer's constructive comment. We agree that our original discussion on "distance and connectivity" was predominantly qualitative. To address this, we have performed a quantitative zonal analysis that explicitly links the MAR response (Δh) to the mapped hydrostratigraphic units.
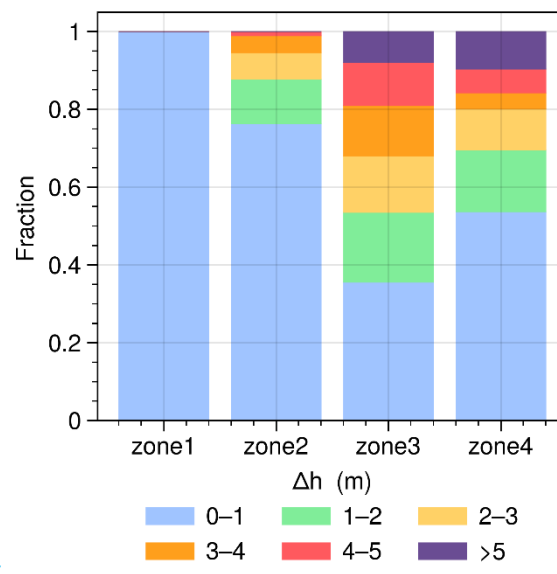

:

Figure R2. Distribution of Groundwater Head Recovery (Δh) Across Zones by 2035

As illustrated in Figure R2, the statistical results reveal a stark contrast controlled by hydraulic conductivity:

(1) Zone 1 behaves as a hydraulic barrier, with 100% of its cells showing minimal recovery ($\Delta h < 1$ m), confirming that low permeability effectively blocks the propagation of the recharge signal. Similarly, the intermediate Zone 2 is dominated by weak responses (76% of cells < 1m), with only localized recovery

(mean rise ~2.23 m) in limited areas.

(2) In contrast, the permeable zones facilitate significant recovery. Zone 3 exhibits the most extensive response coverage, with over 64% of cells exceeding Δh >1m and one-third exceeding 3 m. Zone 4 (Highest Permeability), while spatially more focused (46% coverage), facilitates the most intense local recovery. It achieves the highest mean head rise (~3.18 m) and peak values (~7.5 m) among the responding cells.

This zone-by-zone quantification explicitly demonstrates that strong MAR benefits (Δh > 3 m) are confined to high-permeability corridors (Zones 3–4), while low-permeability zones remain largely unaffected.

Revisions have been made in line 208:

"A quantitative zonal analysis further confirms this control: Zone 1 acts as a hydraulic barrier with 100% of cells showing minimal recovery (Δh < 1 m). In contrast, the high permeability corridors facilitate significant recovery, with Zone 3 showing extensive coverage (64% cells > 1 m) and Zone 4 achieving the highest mean rise (~3.18 m) and peak values (~7.5 m). This demonstrates that strong MAR benefits are spatially confined to conductive pathways (Zones 3-4) rather than being uniformly distributed."

9. Line 215: The concept of comparing points with different distances and permeabilities is good, but the interpretation mixes permeability effects and distance effects. For example, line 220 states that "their amplitudes diverge with permeability: Point1 rises by ~4–7 m, while Point2 increases by ~3–5 m across the projection horizon". But figure 5b clearly shows a larger amplitude than 5a, which means MAR posed a higher effect on Point2. Also, no lag time between Point1&2 is surprising.

Response:

We thank the reviewer for this careful observation. We apologize for the confusion caused by our terminology.

(1) We would like to clarify that our analysis was based on monthly average water levels. The values mentioned in the text (rising by ~4–7 m vs. ~3–5 m) refer to the cumulative magnitude of water level rise (the net change over the simulation period), rather than the "seasonal fluctuation amplitude."

(2) The absence of a significant lag time between Point 1 and Point 2 is attributed to their immediate proximity to the recharge source. Both observation wells are located within < 500 m of the river channel. Given the rapid propagation of pressure waves in these near-field zones, the hydraulic response appears synchronous at the model's temporal resolution.

We have incorporated the updated figure into the main text and revised the corresponding descriptions accordingly.

Revisions have been made in line 220:

"Although the two sites respond nearly synchronously, their cumulative water-level rise: Point1 rises by ~4–7 m, while Point2 increases by ~3–5 m across the projection horizon (Fig. 5a–b)."
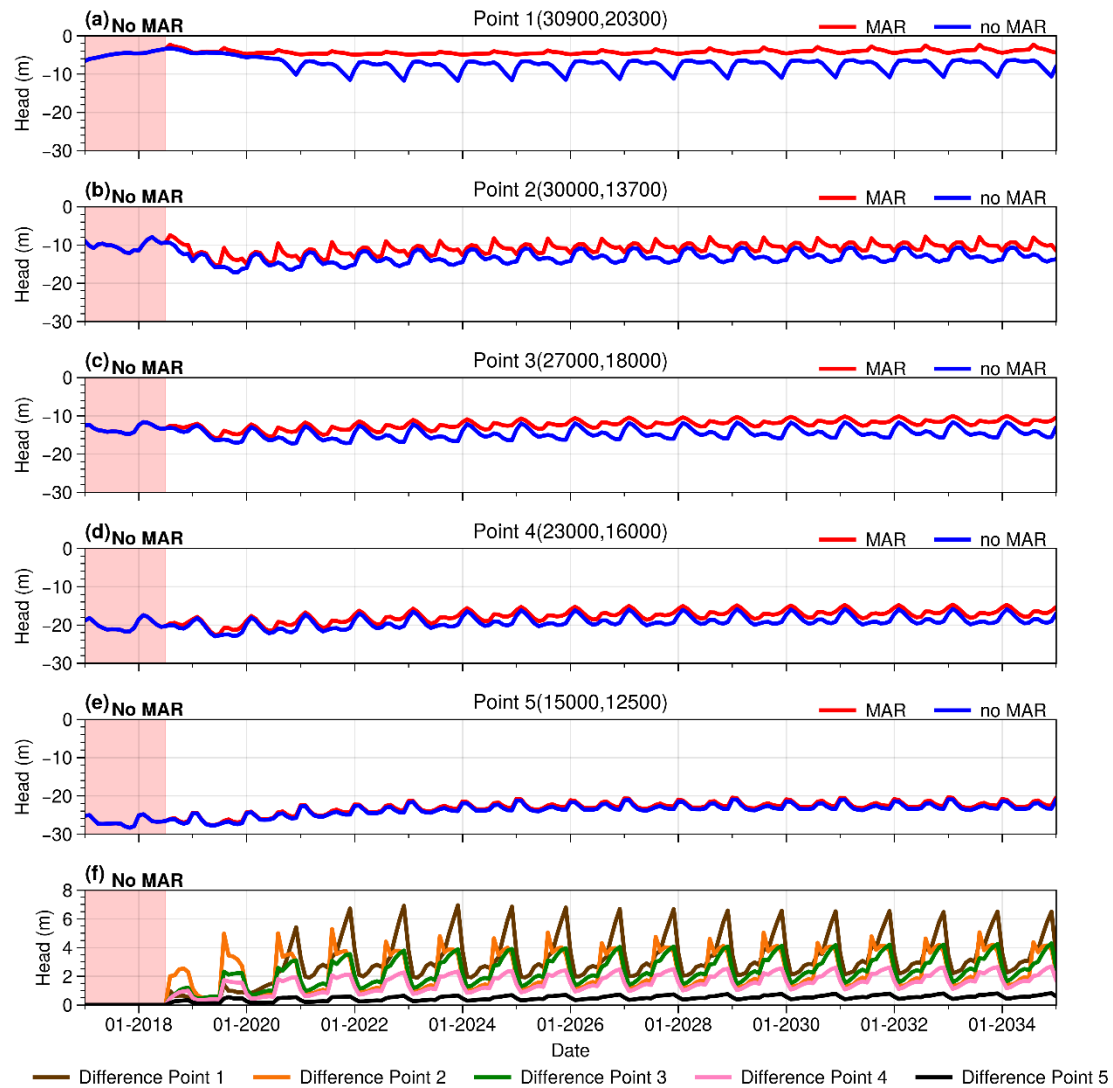
Figure 5. Groundwater level changes and differences at observation points; (a)–(e) Time series of water level changes at observation points (Point 1 to Point 5) with and without MAR; (f) Differences in water levels (MAR vs Without MAR) for all observation points

10. Line 255-256: Be cautious with statements such as "strongest decreases" when the absolute reductions remain quite small (10^{-5} mol/L is extremely minor relative to typical groundwater nitrate levels). You may want to discuss ecological significance, not just the mathematical magnitude.

Response:

We sincerely thank the reviewer for this valid caution regarding the magnitude description. We agree that the term "strongest" could be misleading, given the small absolute reduction. Instead of adding a lengthy discussion, we have refined the description directly in Section 3.3 to provide immediate context and avoid overstatement.

Revisions have been made in line 255:

"The response classes correspond to orders of magnitude: -8 to -7 (very small decrease), -7 to -6 (small),

-6 to -5 (moderate to large), and > -5 (the most distinct decreases, at least on the order of 10^{-5}), while the absolute magnitude (10^{-5}) appears minor, it represents the primary zones of cumulative mass removal in the system."

11. Line 278: Obviously, denitrification is effectively negligible in this system. How reliable this result is? Plus, it is surprising that the spatial pattern is nearly uniform, given strong variability in residence time, electron donor availability, and organic carbon distribution in real systems.

Response:

We appreciate this insightful comment. Although residence time usually drives reaction patterns, the hydrogeochemical constraints (high oxygen, low carbon) in our study area suppress denitrification basin-wide. This strong chemical limitation effectively masks the influence of physical heterogeneity, resulting in a uniform and negligible reaction rate.

(1) The result is reliable because it is constrained by site-specific monitoring data used as model inputs (Table 2). The shallow groundwater system in Xiong'an is characterized by oxidizing conditions and carbon limitation. According to the reaction kinetics Equation (5), this high oxygen level exerts a strong inhibition effect. Consequently, the system is thermodynamically unfavorable for denitrification. The model correctly predicts that nitrate removal is dominated by physical dilution (~91%) rather than biological reduction (~9%) under these aerobic conditions.

(2) The reviewer correctly notes that variability in residence time usually creates heterogeneous reaction patterns. However, this applies primarily when the reaction is active. In our system, the reaction is globally suppressed by the widespread presence of oxygen and the scarcity of electron donors. Since the oxygen inhibition is applied uniformly strong across the domain, the reaction rate is clamped to near-zero levels everywhere. This chemical limitation acts as the dominant control, masking the secondary effects of physical heterogeneity (residence time).

Revisions have been made in line 285:

"The modeled negligible and uniform denitrification is fundamentally driven by the hydrogeochemical regime of the study site. Field data (Table 2) indicates that the aquifer is generally oxidizing and carbon-limited. Under such conditions, the reaction kinetics are globally suppressed by oxygen inhibition and electron donor starvation. Consequently, the influence of physical heterogeneity (i.e., variability in residence time) becomes secondary to this overwhelming chemical limitation. Even in zones with long residence times, the reaction rates remain low due to the lack of favorable redox conditions, resulting in a spatially uniform distribution of minimal denitrification."

12. Line 353: You highlight the differing effects of heterogeneity on dilution vs. denitrification but do not clearly link these insights to the broader literature on preferential flow and residence-time control on reactive transport. Adding comparison to previous heterogeneity–redox studies would position your results more convincingly.

Response:

We sincerely thank the reviewer for this constructive suggestion. In the revised Discussion, we have explicitly interpreted the differing effects of heterogeneity through the lens of groundwater residence times and contact time. We clarified that high-permeability zones act as preferential flow paths, which drastically increase flow velocity and reduce the time groundwater spends in the aquifer. As established in classical reactive transport literature (Zheng and Gorelick, 2003), when solutes are transported rapidly through the system, the effective contact time between nitrate and reactive microbial communities is insufficient for substantial biodegradation to occur.

Revisions have been made in line 369:

"This aligns with previous studies (Zheng and Gorelick, 2003; Green et al., 2010; Chen et al., 2024), which indicate that while high-preferential flow enhances physical dilution via dispersion, they simultaneously restricts denitrification by shortening groundwater residence times. Consequently, rapid transport results in insufficient effective contact time for substantial biodegradation to occur."

13. Line 362: "he"

    Response:

    We have corrected "he" to "The" in the revised manuscript.

14. Line 373: While you note that the recharge scheme assumed constant conditions, you do not discuss how sensitive the nitrate response is to recharge timing, distribution, or water chemistry. Without acknowledging potential model sensitivity, the management implications may appear overstated.

    Response:

    We thank the reviewer for this crucial observation. We fully agree that assuming constant recharge conditions simplifies the complexity of real-world operations, where seasonal timing, spatial distribution, and hydrochemistry vary significantly. Neglecting these factors could indeed lead to overstated management implications.

    Revisions have been made in line 376:

    "The simulated nitrate reduction represents a baseline scenario under constant recharge. In practice, operational intermittency (wetting-drying cycles) could enhance oxygen intrusion into the vadose zone, strengthening denitrification inhibition. Furthermore, fluctuations in source water chemistry, particularly reductions in Dissolved Organic Carbon (DOC) load, would limit electron donor availability. Consequently, actual nitrate mitigation efficiency may be lower than simulated if hydraulic saturation and sufficient carbon supply are not maintained."

    Revisions have been made in line 394:

    "Overall, MAR proves to be a robust tool for hydraulic recovery; however, its effectiveness for nitrate mitigation is primarily driven by physical dilution. Therefore, its implementation as a remediation strategy requires careful consideration of source water chemistry and continuous injection regimes to maximize benefits."

Chen, K., Guo, Z., Zhan, Y., Roden, E. E., and Zheng, C.: Heterogeneity in permeability and particulate organic carbon content controls the redox condition of riverbed sediments at different timescales, Geophysical Research Letters, 51, e2023GL107761, https://doi.org/10.1029/2023GL107761, 2024.

Green, C. T., Böhlke, J. K., Bekins, B. A., and Phillips, S. P.: Mixing effects on apparent reaction rates and isotope fractionation during denitrification in a heterogeneous aquifer, Water resources research, 46, https://doi.org/10.1029/2009WR008903, 2010.

Le Traon, C., Aquino, T., Bouchez, C., Maher, K., and Le Borgne, T.: Effective kinetics driven by dynamic concentration gradients under coupled transport and reaction, Geochimica et Cosmochimica Acta, 306, 189–209, https://doi.org/10.1016/j.gca.2021.04.033, 2021.

Schäfer Rodrigues Silva, A., Guthke, A., Höge, M., Cirpka, O. A., and Nowak, W.: Strategies for simplifying reactive transport models: A Bayesian model comparison, Water Resources Research, 56, e2020WR028100, https://doi.org/10.1029/2020WR028100, 2020.

Zheng, C. and Gorelick, S. M.: Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale, Groundwater, 41, 142–155, https://doi.org/10.1111/j.1745-6584.2003.tb02578.x, 2003.