Response to Reviewer 2:

We would like to thank the reviewer for the positive feedback and thoughtful suggestions. Below are our responses to the comments raised:

General comments

The article addresses the dissolution trapping of CO_2 during geological carbon sequestration in deep saline aquifers, a critical process that serves as a key indicator for evaluating the safety of long-term CO_2 storage. By trapping CO_2 in the aqueous phase, dissolution trapping helps prevent CO_2 leakage. However, the dissolution process can result in gravity-driven instabilities, such as Gravity-Driven Convection (GDC), and is influenced by formation properties like permeability and porosity, which are subject to uncertainties due to heterogeneities. The article aims to quantitatively analyze the explicit relationships between GDC-driven dissolution rates and permeability heterogeneities reported in the literature through numerical experiments.

The article is well-written and easy to understand. It is evident that the authors have carefully guided readers step by step through their work, supported by an in-depth literature review. Additionally, they have conducted a substantial number of numerical simulations across various setups to provide robust evidence. Based on my assessment of the manuscript, I believe this work is highly deserving of publication, subject to revisions.

R: We are gratified that the reviewer has acknowledged the manuscript as well-written and clear. We sincerely appreciate the recognition of the meticulous explanations and the robustness of the numerical simulations conducted in this study. Additionally, we are thankful for your suggestion that the work merits publication, pending revisions. We are eager to address the comments to further enhance the clarity and impact of the manuscript. The following revisions will be incorporated into the subsequent version.

Major Comments:

These are primarily comments from me and do not involve major revisions to the article.

(1). *The 2D numerical setup is well-chosen, in my opinion, and all choices are clearly explained. Simulation data on a 100x100 grid over an extended period are obtained from 554 realizations of the permeability distributions. Additionally, the time-step size is governed by CFL conditions, which can result in a significant number of time steps. I assume this process is computationally intensive, and I was wondering how challenging it would be to extend this work to 3D, as doing so could provide valuable insights. Perhaps providing an approximate*

value of the CPU time per simulation could serve as a good starting point and indicator.

R: Thank you for raising the important point about extending our study to a 3D framework and acknowledging the computational challenges we have encountered. We fully agree that a 3D setup would offer a more realistic depiction of permeability heterogeneities in geological formations. However, the current computational resources pose significant limitations on the feasibility of conducting ensemble-based studies in 3D. As you correctly pointed out, the 2D simulations were already computationally intensive due to the large number of realizations (554) and the constraints imposed by the CFL time-step conditions. Extending these simulations to 3D would introduce considerable additional computational complexity.

To elaborate, the current 2D simulation, of which the grid discretization is 100×100, involves a $(2\cdot10^4)\times(2\cdot10^4)$ sparse matrix (here, 2 means we have two independent variables), taking approximately **4–7 hours** to complete. In contrast, a 3D simulation, of which the grid discretization is 100×100×100, would correspond to a $(2\cdot10^6)\times(2\cdot10^6)$ sparse matrix. Based on our estimates, the simulation time for a single 3D case would be at least 4×100 hours (around **17 days**). **Moreover, increasing the matrix size would also significantly elevate the memory requirements. Actually, we tried to give a 3-D example for you on my own desktop (not a cluster), but the simulation was extremely slow and it broke down with "Out of memory." warning, during estimating the condition number with LU using** `condest`**. The CPU and the RAM of the computer used are "12th Gen Intel(R) Core(TM) i9-12900K 3.20 GHz" and "64.0GB", respectively.** Given that we need to perform a total of 554 simulations, the computational burden and memory demands associated with a 3D setup would be formidable and currently beyond our capacity.

Second, it is worth noting that the stabilized mass flux in the 3D scenario is approximately 25% higher than that observed in a comparable 2D simulation (Pau, 2010). Although this difference is statistically significant, it is relatively minor when compared to the several orders of magnitude variability in permeability commonly seen in geological media (e.g., $\kappa = 10^{-16} - 10^{-12}$ m^2 and $\sigma^2_{\ln\kappa}$ = 3-10, as reported by Wang et al., 2022). This indicates that the impact of additional spatial dimensions on the stabilized mass flux is secondary to the influence of permeability heterogeneity. In the revised manuscript, we have included a concise discussion on the differences between 2D and 3D simulations of GDC.

The following sentences will be added to the conclusion Section 4.1 . "We note that in a more realistic 3D scenario, the dissolution rate may be approximately 25% higher than that observed in 2D cases. However, this difference is relatively minor when compared to the significant variability in permeability commonly encountered in geologic media Wang et al. (2022)." The revisions will be incorporated into the subsequent version.

(2). *The authors clearly state the software used for the numerical experiments and provide open access to the code and data. The user guide is extremely valuable, as it includes all the numerical details of the solver. However, I believe that for full reproducibility of this specific work, it would benefit from an additional short note, which I may have overlooked.*

R: Thank you for the suggestion regarding the reproducibility of the work. We agree that for full reproducibility, it would be beneficial to provide an additional note to clarify specific implementation details. We are sorry for not giving a very clear description to reach our conclusions. In the following, we give a summary that provides any potential information for full clarity, ensuring that all steps in the numerical experiment are fully accessible to users. Given that the reader has access to this document, we will not make further modifications to the original manuscript. We will neither put this on the general user guide for the MRST simulator because flow simulations are only used in Steps 2 and 3 described below.

Step 1. Generating heterogeneous fields based on the parameter given in Table 4 in the article. Random fields were generated using the sequential Gaussian simulation method implemented in the SGSIM code of GSLIB. *Note: The fields are generated in a random manner and could differ from those used in the present work. However, this does not impact the statistical conclusions drawn in this study.*

Step 2. Measuring the equivalent vertical (κ_z^e) and horizontal (κ_x^e) permeability for each realization of the random fields. To estimate κ_i^e ($i = x, y$), we neglect gravity and saturate the porous medium with only water. We then set the domain sides perpendicular to the *i*th direction as impermeable, and we impose a pressure decrement $|\Delta_i p|$ along the *i*th direction. κ_i^e is estimated by the total volumetric flow Q_i passing through the system in the ith direction as $\kappa_i^e = \mu Q_i L_i/(A_i|\Delta_i p|)$, where L_i is the domain size along the *i*th direction and A_i the corresponding cross-sectional area. *Note: To maintain the simplicity and clarity of the code library, we have omitted this specific calculation in the uploaded code, https://zenodo.org/records/5833962. However, the calculation can be readily implemented by making minor modifications to the code provided in our uploaded file. Alternatively, it can also be achieved using existing functionalities in the MRST https://www.sintef.no/projectweb/mrst/.*

Step 3. Conducting GDC (Gravity-Driven Convection) simulations with the provided code (located in *benchmarks and examples/example_unstable_finger*). It is essential to capture two key pieces of information during the simulation process: the mass flux through the top boundary, denoted as $F(t)$, and the detailed flow field. *Note: The example code is designed to be broadly applicable. Users are required to tailor the simulation parameters—discretization, domain size, permeability, and porosity—according to the specifications outlined in Table 3. The methodology for calculating $F(t)$ is as follows:*

$F(t) = \frac{1}{L}\int_0^L \left[\rho X^C q_z - \phi D\rho \frac{\partial X^C}{\partial z}\right]_{z=0} dx$ (c.f. Equation (15) in the manuscript)

The flow field is used to calculate the finger velocity $v(t)$*, which is given as follows:*

$v(t) = \max_{0<z<B}\left\{\frac{1}{L}\int_0^L \frac{1}{\phi}|q_z|dx\right\}$ (c.f. Equation (16) in the manuscript)

Step 4. Estimating the Asymptotic Values of the Dissolution Rate (F_∞) and Vertical Finger Velocity (v_∞). In our simulations, the asymptotic values of the global dissolution rate (F_∞) and the vertical finger velocity (v_∞) are determined by computing the temporal averages of $F(t)$ and $v(t)$ over the interval $[\frac{t_b}{3}, t_b]$. Here, t_b denotes the time at which the earliest aqueous CO_2 finger reaches the bottom, marked by the moment when the maximum CO_2 mass fraction at the bottom exceeds 25% of the CO_2 concentration at the top boundary.

Step 5. Obtaining the predictors for the asymptotic dissolution rate by performing regression analysis for the simulation results (summarized in https://zenodo.org/records/14061632) based on

$F_\infty^* = \gamma_1 \left(\frac{\kappa_z^e}{\kappa_g}\right)^{\alpha_1} \left(\frac{\kappa_x^e}{\kappa_z^e}\right)^{\beta_1}$ (c.f. Equation (17) in the manuscript)

and

$F_\infty^* = \gamma_1 \left(\frac{v_\infty}{v_c}\right)^{\alpha_2} \left(\frac{\kappa_x^e}{\kappa_z^e}\right)^{\beta_2}$ (c.f. Equation (18) in the manuscript)

Finally, we will get the results listed in Table 5 of the manuscript.

(3) *Multiple realizations of permeability distributions are provided, and Figure 4 effectively illustrates the differences in numerical simulation results between them. I was wondering whether the upscaling of highly heterogeneous permeability fields could yield similar results for the predictors or if a relationship could be established between the upscaled and original permeability fields.*

R: This is a very good point. The question of upscaling permeability fields and its impact on our findings is an important one. The upscaling of heterogeneous permeability fields should statistically yield similar results as the original permeability fields, although the shapes of concentration profiles are different in original and upscaled fields.

Proof by by Elenius and Gasda (2013): The dissolution coefficient γ obtained in heterogeneous fields by Elenius and Gasda (2013) is very similar to that in homogeneous field using equivalent permeability. This means that gravity-driven convection is governed by a rule similar to Darcy's law, with the density difference acting as the driving force, this relationship can be illustrated by the following equation: (cf. Equation (1) in the manuscript):

$$F_\infty = \gamma X_0^C \rho_0 \frac{\Delta\rho g \kappa}{\bar{\mu}} = \gamma X_0^C \rho_0 \frac{\kappa}{\bar{\mu}} \Delta\rho g \quad \text{(R1)}$$

This means that the dissolution rate is statistically the same for heterogeneous fields and corresponding upscaled homogeneous fields of equivalent permeability.

Proof from current work: We first perform 10 GDC simulations with different random noise in a homogeneous field (or upscaled field) with permeability $\kappa = 10^{-12}$ m^2, and then perform GDC simulations in different isotropically heterogeneous fields with geometric mean permeability $\kappa_g = 10^{-12}$ m^2, we obtain that the relation of the dissolution rate follows $\frac{F_{hetero}}{F_{homo}^{mean}} = \frac{\kappa_z^e}{\kappa}$, where κ_z^e is the equivalent permeability, F_{homo}^{mean} and F_{hetero} are the statistic dissolution rate in homogeneous field (or upscaled field) and the dissolution rate in the heterogeneous field. This means the dissolution rate in the heterogeneous field can be obtained using an upscaled homogeneous field of equivalent permeability.

For anisotropic fields, we did not perform simulations in upscaled anisotropically homogeneous fields, but we can expect that upscaling should also work for anisotropic fields, because our predictors $F_\infty^* = \gamma_1 \left(\frac{\kappa_z^e}{\kappa_g}\right)^{\alpha_1} \left(\frac{\kappa_x^e}{\kappa_z^e}\right)^{\beta_1}$ (c.f. Equation (17) in the manuscript) and $F_\infty^* = \gamma_1 \left(\frac{v_\infty}{v_c}\right)^{\alpha_2} \left(\frac{\kappa_x^e}{\kappa_z^e}\right)^{\beta_2}$ (c.f. Equation (18) in the manuscript), which are based on equivalent permeabilities, perform well for all anisotropy, as shown in Figure R1 (c.f. Figure 7 in the manuscript):

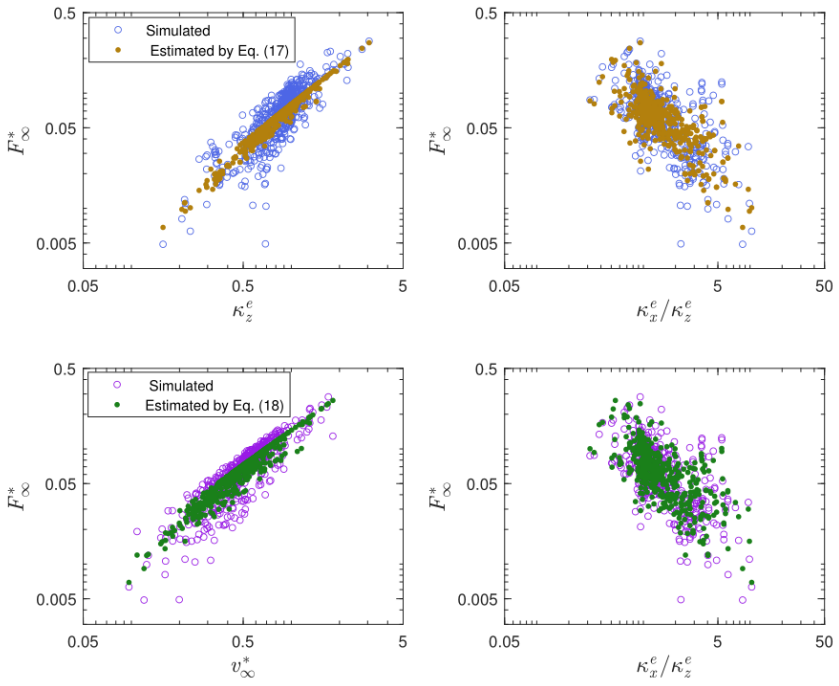

Figure R1. Performance of predictors in fields of different equivalent vertical and horizontal

permeabilities κ_z^e and κ_x^e.

Inspired by your comments, we add the following sentences in the manuscript:
"The results indicate that employing an upscaled permeability field with equivalent permeability does not compromise the depiction of dissolution efficiency in GDC simulations, although permeability upscaling does alter the shapes of the dissolution profiles." The revisions will be incorporated into the subsequent version.

Minor Comments:

(1). *Table 3: Could you comment on the potential effects of introducing a relationship between porosity and permeability realizations? Specifically, how might such a relationship influence the dissolution rates.*

R: This is indeed an insightful point. Introducing a relationship between porosity and permeability to account for heterogeneous porosity distribution may subtly influence the morphology of instability fingers, because the pore size may affect the interstitial velocity. However, it does not statistically impact the dissolution rates. This is because gravity-driven convection is governed by a rule similar to Darcy's law, as illustrated by the equation (R1), which clearly shows that the dissolution rate depends solely on the equivalent permeability and is independent of porosity.

We further explan the use of constant porosity in our simulations for two additional reasons:
（i）Limited Porosity Variation: The range of porosity variation (0.1–0.38) is relatively narrow compared to the wide range of intrinsic permeability ($\kappa = 10^{-16} - 10^{-12}$ m^2 and $\sigma^2_{\ln \kappa}$ = 3-10), as documented in Table 3 of Wang et al. (2022) and Table 4 of Elenius and Johannsen (2012).（ii）Ambiguity in Permeability-Porosity Relationships: On one hand, clay particles are significantly smaller than sand particles, resulting in a higher total pore space in clay soils. However, these pores are typically small and poorly connected, leading to low permeability. In contrast, sand particles are larger and more irregularly shaped, creating larger and better-connected pores that facilitate higher permeability. Thus, permeability is influenced not only by pore volume but also by pore shape, meaning that high porosity does not necessarily imply high permeability. On the other hand, for a given aquifer, increasing porosity through acid water erosion often leads to an increase in permeability, as described by the well-known Kozeny-Carman model (Saaltink et al., 2013). However, even if we employ the Kozeny-Carman model to represent the permeability-porosity relationship, the model parameters are typically site-specific, limiting their generalizability.

(2). *Line 231: 553 realizations and line 354: 554 realizations.*

R: We apologize for the typo. We corrected this in the manuscript. The total number of simulations should be 554.

(3). *In Figure 4, the range of permeability values appears to be relatively narrow (from -14 to -10 on a logarithmic scale). Could you comment on the potential impact of using a wider range of permeability values?*

R: This is a very good point. We will respond your comments in two aspects.

(1) When the geometric mean permeability and correlation length are held constant while the variance (σ_Y^2) increases to produce a broader spectrum of permeability values, the influence of permeability heterogeneity becomes more pronounced, and the connectivity within the medium improves. Consequently, preferential flow tends to occur within the interconnected high-permeability zones. In this context, the uncertainty associated with the development of instability fingers is primarily governed by the permeability heterogeneity, while the role of white noise, which initially triggers the instability, becomes relatively minor. Essentially, the flow becomes focused in the high-permeability regions regardless of the specific initial perturbation.

Note: In a medium with minimal heterogeneity, instability fingers can emerge due to minor white noise present in the initial conditions. This can result in variations in finger shapes across different simulations using distinct white noise inputs. However, the statistical dissolution rate remains consistent, as demonstrated by Pau et al. (2010) and this work.

Irrespective of the behavior of CO_2-rich fingers, the overall vertical mass flux of CO_2 can be reliably predicted based on the equivalent vertical and horizontal permeabilities. These equivalent permeabilities can be calculated using the method detailed in Section 4.5 of the manuscript.

Therefore, we expect that employing a large variance (σ_Y^2) values amplifies preferential channeling within interconnected high-permeability zones and may consequently affect dissolution rates. Nonetheless, this observation does not undermine the conclusion that CO_2 dissolution rates can be reliably estimated using equivalent vertical and horizontal permeabilities. To put it succinctly, the fundamental relationship between dissolution rate and equivalent permeabilities remains consistent, regardless of the permeability variability.

(2) When the variance (σ_Y^2) is negligible while the average permeability changes, the size of the instability fingers is inversely proportional to the permeability, as described by the relation $l_c = 70 \cdot \frac{\mu\phi D_m}{\Delta\rho g \kappa_g}$. This relationship indicates that the finger size is very large in media with

very low permeability. This insight is particularly valuable when designing simulation domains or laboratory experiments, since an inadequately chosen domain size may fail to accurately represent gravity-driven convection phenomena.

For instance, consider a scenario where the characteristic length of the fingers (l_c) is 1 meter. In such cases, employing a simulation domain or experimental reservoir smaller than 1 meter may fail to accurately capture the development of instability fingers. Given a specific size of the experimental reservoir, meticulous selection of sand permeability becomes essential to ensure that the observed finger distributions are both representative and meaningful.

Therefore, we claim that changing the mean permeability in a field with negligible heterogeneity will change the finger size, and the size of the simulation domain should be changed accordingly to efficiently match the density instability fingers.

We add the following comments in the revised manuscript:
"From Figure 7, it is also evident that the performance of our predictors is not influenced by permeability. This suggests that the findings of this study can be extended to fields with greater permeability heterogeneity." The revisions will be incorporated into the subsequent version.

Reference:

Elenius, M. T. and Johannsen, K.: On the time scales of nonlinear instability in miscible displacement porous media flow, Computational Geosciences, https://doi.org/10.1007/s10596-012-9294-2, 2012.

Elenius, M. T. and Gasda, S. E.: Convective mixing in formations with horizontal barriers, Advances in Water Resources, https://doi.org/10.1016/j.advwatres.2013.10.010, 2013.

Pau, G. S., Bell, J. B., Pruess, K., Almgren, A. S., Lijewski, M. J., and Zhang, K.: High-resolution simulation and characterization of density-driven flow in CO2 storage in saline aquifers, Advances in Water Resources, 33, 443 – 455, https://doi.org/https://doi.org/10.1016/j.advwatres.2010.01.009, 2010

Saaltink, M. W., Vilarrasa, V., De Gaspari, F., Silva, O., Carrera, J., and Rötting, T. S.: A method for incorporating equilibrium chemical reactions into multiphase flow models for CO2 storage, Advances in Water Resources, 62, 431 – 441, https://doi.org/10.1016/j.advwatres.2013.09.013,2013.

Wang, Y., Fernàndez-Garcia, D., and Saaltink, M. W.: Carbon Dioxide (CO2) Dissolution Efficiency During Geological Carbon Sequestration (GCS) in Randomly Stratified Formations, Water Resources Research, 58, e2022WR032 325, https://doi.org/10.1029/2022WR032325, 2022.