This study combined borehole soil sampling, groundwater level data, and groundwater flow and transport modeling to examine the impact of subsurface heterogeneity on dispersion at the basin-scale. Overall, the dataset and modeling framework are promising, and the study has the potential to make a valuable contribution once the issues below are addressed.

Major comments:

1. The novelty and knowledge gap can be better articulated. In particular, the current introduction may not set the stage for this study well. Consequently, the three research objectives lack literature review support and strong motivations.

- For instance, the reviewer would be curious what motivated the author to perform an uncertainty analysis. Was it because existing literature shows a wide range of different observations in the field? Was it because a new model was used and the model involved some uncertain parameters whose uncertainty is unknown? Was it because there has been an inconsistency between existing observations and models? Or something else?

- Another example is why did they want to quantify the relative contributions of sedimentary architectural attributes, hydraulic statistics, and source size? There seems to be no well-organized discussions on the existing knowledge about how these three factors control the dispersion processes under field-representative aquifers and large-scale flow fields.

- A similar question arises for the uncertain proportion part.

2. The writing and logical flow can be improved through substantial revisions. Please see the specific comments for some examples.

3. The manuscript involves quite a number of jargons and unclear sentences, which may require clarification or rephrasing. Below are some examples:

- “From pre-asymptotic dynamics to uncertainty propagation”
- “geometric connectivity”
- “general influence”
- “hierarchical organization”
- “Although the general influence of heterogeneity has been extensively investigated, the scale-dependent effects of hierarchical organization, particularly under basin-scale flow conditions, remain inadequately quantified.”
- “lithofacies”
- “macroform scales”
- “macro-dispersion”
- “finer-scale heterogeneity”
- “Non-ergodic”
- “basin-scale structure”
- “macroscopic spreading and dispersion across scales”

- “A total of 57 boreholes … were collected”
- “To bias-correct parameter estimates affected by incomplete exposure of sections, …”

Specific comments:

1. What is the relationship between “finer-scale”, “micro-scale”, “meso-scale”, “macro-scale”, “macroform scales”, “large-scale”, “laboratory-scale”, “facies-scale” “field-scale”, “region-scale”, “kilometer-scale”, “basin-scale”? Without clear definitions or quantitative descriptions (like “kilometer” did), it is hard to understand the specific scales.

2. Line 30: Has the "laboratory- and sandbox-scale study” been done in this work? If not, this argument can be misleading.

3. Line 30: The reason why the current study can support studying groundwater management at data-limited regions is not trivial? Please briefly explain the reasoning herein or in the main text.

4. Line 60: It is unclear how the arguments “Recent work …” and “High-resolution …” logically connects with each, as well as how they connect with the prior discussions. That said, it is hard to catch the key message behind these arguments.

5. Line 65: There was no discussion on solute transport at other scales previously. It is hard to see the connection between the prior paragraph to this statement.

“Most validations have been limited to laboratory or site-scale studies.” cannot be justified by the current introduction. Also, it’s unclear what point the authors would like to make.

6. Line 70: How can "kilometer-scale" be considered as "high-resolution"? What resolution can be considered as "high"?

7. Line 75: “late-time tails …” This can be good motivation for this work, while the argument "Even with these advances, ..." makes the gap ambitious.

8. Line 80: As suggested in my first major comments, the current introduction didn’t provide a comprehensive literature review and clear research trajectory to motivate the objectives.

9. “Part II the” is missing a verb.

10. Section 2 - Method: Each subsection of the method section includes both methodology(i.e., general approach) and specific model setup (i.e., parameters and data), which sometimes can affect readability. I would suggest separating them into a methodology section and a parameterization (or data collection and model setup) section. I believe this reorganization will substantially improve the clarity and quality of the paper.

11. Line 90: “The Nen River defines …” Why is this matter?

“The regional terrain is gently…” Is there any support from literature or the collected data? if so, I would either cite the references or clarify the data support.

12. Figure 1 caption: More details should be provided about the selections of boreholes and cross-sections, as well as the rational or purposes of such decisions.

Additionally, what were the cross-sections used for? Or what kind of information was collected from the cross-sections?

13. Figure 2: The figure is hard to interpret without providing more details about the meanings of the acronyms and the plotted regions.

In the right figures, how can one interpret the figure? Are they referring to the horizontal or vertical view of the aquifer, or something else?

How are the right figures connected to Figure 1b.

14. Line 130: Can you elaborate what does it mean by "parameterized by lithofacies volume proportions and mean lengths." or provide a brief introduction about the parameterization?

15. Line 130: What is the "bias-correct parameter"? What does it mean by “To bias-correct parameter estimates affected by incomplete exposure of sections,”?

16. Line 135: How were P_S and P_D defined?

17. Table 1: It is hard to understand the physical meanings of the parameters and how they were derived. I would suggest providing a schematic or conceptual plot and label them in the plot at least for some of them (e.g., Lx, Ly, Lz), while providing the formula to compute P.

18. Line 155: I guess you were referring to the characteristic diameters of the grains. Please clarify.

What is USBR?

19. Table 2: The unit should be ln(m/d). Is simga dimensionless? If not, the unit should be provided.

20. Line 175: The authors are recommended to justify that 50 realizations are statistically grounded by adding relevant references.

21. Figure 3 provides a more clear explanation of part of the methodology. Based on that, I would suggest using some plots (e.g., a and c) to create a conceptual figure to describe the entire methodology and adding a short subsection to introduce the conceptualization at the very beginning of Section 2 Method. This will make the rest easier to follow.

22. Figure 4. What is the blue rectangle representing? Are the boundary conditions applied only to the lines or the faces?

23. Line 225: I guess planar source was indicated in Figure 4, while point source was not. Additionally, the authors should explain the rationale of simulating the two different sources, regarding the real-world processes they are representing.

24. Line 250: I would suggest reporting the error or correlation between model and observation. Additionally, in addition to the fitness between the averaged simulated water levels and measured ones, reporting the errors that reflect the fitness of simulated water levels for each realization or their statistical values (e.g., mean and standard deviation of the error of each realization) would better verify the robustness of the model results.

25. Line 280: Add references to the Borden site results.

Why does higher mean velocity accelerate stability? Was it due to stronger mechanical dispersion, or something else?

26. Line 305: The sensitivity and uncertainty methods were not introduced in section 2. Please provide the details about the specific methods and the analyzed factors and parameters.

27. Figure 7 caption: Please clarify what each model was representing

28. Line 350: What does scale-dependent tread mean? How can realization-to-realization variability reveal the "scale-dependent tread"? What is the underlying rational and theoretical basis?

29. Line 355: It seems Scale II was used for these analyses. Can you comment on whether the impact of varying K values on flow and solute transport will remain similar between Scale I and Scale II models?

30. Line 415: What is the reason for this difference? What is the implication for characterizations and simulations of contamination transport at different sites?

I will probably talk about the key finding of these studies and then make the comparisons with the Borden site in the second last paragraph of this section given that there were several comparisons (similarity vs difference).

31. Line numbers should be appended to each line for the reviewer to pinpoint their comments.