GENERAL COMMENTS

This manuscript presents an extension of the work by Rizzo and de Barros (2017). The developed surrogate model is tested on an example, where the goal is to determine the permeability character of three faults in a domain for which the conductivity field, the position of the sources and the position of the faults is known.

I was not one of the reviewers of the original submission, and, therefore, I will add some comments to the remarks of the colleagues who reviewed the original submission.

The manuscript is well organized and written with a good language.

However, in my opinion the innovative content of the work is quite limited and the test used to validate the surrogate model is excessively simple.

I think that the work suffers from some scientific weaknesses, that are listed in the specific comments below.

Overall, I think that the manuscript cannot be considered for publication in its present form, but it requires a major revision.

SPECIFIC COMMENTS

1) Lines 2 & 3. I found this sentence rather obscure and, if I understood it properly, I do not fully agree with the concept it conveys. More generally, at the end of the abstract, I had the idea that the work proposes a surrogate model, based on graph theory, rather than a physically-based model, in order to exclude some hydrostratigraphic or hydrogeological scenarios. However, I am not sure that I got the right point.
2) Lines 120 to 124. This remark is intriguing, but I do not understand the physical motivation. Moreover, if I understood correctly, equation (4) implies that the local flow direction is always the same as the main direction of flow. In the case of a highly heterogeneous medium, the geometry of low permeability structures could yield a local flow with a direction opposite to the average flow. Therefore, I am afraid that this formula could not be optimal for some conditions.
3) Section 2.3 Observation time. I think that this discussion is not well developed. In particular, the surrogate model proposed here gives a steady-state picture of flow, which is then used to estimate transient solute transport, without solving the ADE. So, the problem of the selection of the observation time should be analysed very carefully.

4) Section 2.4 Metrics. The proper selection of metrics for this comparison is a well known problem, for which, to my knowledge, there is no definite solution. the authors provide an interesting discussion, but I would appreciate if they can support their remarks with some numerical tests.
5) Lines 224 to 229. The authors assume that the conductivity field and the position of the faults (this is not explicitly stated in the text, but I deduced it from the rest of the description here) are known. Then the problem is restricted to the determination of the permeability o the faults. This problem could be solved quite simply with few runs of a numerical, physically-based model. Indeed, the problem is often that scarce data do not allow to map preferential flow paths (permeable faults, fractures, permeable sediments, etc. for different types of aquifers) and their connectivity.
6) In several steps of the work, thresholds are introduced. As it is common in such approaches, the selection of thresholds is often a subjective operation. A thorough sensitivity analysis with respect to the applied thresholds, extending what is found in Appendix B, would be necessary.
7) Throughout the whole work, MODFLOW is mentioned as the model applied to generate synthetic data. However, to my knowledge, MODFLOW solves the flow equations, whereas MT3D or other models are used in cascade to model transport. Is this the case? I think this should be specified in a better way.

TECHNICAL COMMENTS

1) Line 2. What is "it"? the subject of the previous sentence is "groundwater contaminant transport problems".
2) Lines 14 to 16. Sentence "The study of… heterogeneous environments" is quite tautological. Sentence "Contamination… mitigation strategies" is quite generic and it ends with a couple of citations that do not seem to be the most relevant to support this statement.
3) Line 40. "CO_2" should be corrected.
4) Line 48. Expression "multi-heterogeneous-layer" should be rephrased.
5) Figure 1. The three geological units mentioned at line 67 are not represented. It would be useful to provide the length of the domain along the x and y directions, otherwise using the scale length remains quite imprecise. These values are given at line 79; they could be given earlier.
6) Line 70. Sentence "Flow… equation" is rather imprecise. ADE (advection-diffusion equation) is used to model transport and it is based on Darcy's law and on Fick's law for diffusion and dispersion. Moreover, Darcy's law is used to model flow, together

with the continuity equation (i.e., the mathematical formulation of the mass conservation principle).

7) Line 71. The authors mentioned MODFLOW, which is a finite difference code, but here they underline the use of finite elements. OK, this is not a big problem for the specific work presented here, but I'd prefer to avoid confusion. Moreover, these sentences are not well related with the following sentences. And at lines 79 & 80, cell size is mentioned: this seems to be related to a finite difference approximation, rather than to a finite-element simulation.
8) Line 72. Increase or decrease K with respect to what?
9) Lines 76 & 77, 80 & 81. Measurement units are missing for the coordinates of the reference point (lines 76 & 77) and of the border planes (lines 80 & 81).
10) Table 1. The position of wells is given with a precision of 10^{-8} m, i.e., 10 nm. This is not physically significant!
11) Line 81. The average hydraulic gradient is 1/70, which is slightly greater than the typical value of hydraulic gradients, whose magnitude is of the order of 1/1000.
12) Line 85. I do not like the "e" format used here to denote values: 3.5×10^{-5} ms^{-1} is much better, in my opinion. Measurement units should be attached to each value of a list, as required by rules of the SI system.
13) Lines 104 & 105. I do not understand the link between the two sentences "The conductivity field… discrete fields" and" here… 3D space".
14) Lines 196 & 205. Using *i* for both the point index and the coordinates might be confusing.