Review of the paper titled: "Modelling 3D permeability distribution in alluvial fans using facies architecture and geophysical acquisitions" by Zhu et al.

Does the paper address relevant scientific questions within the scope of HESS? Yes it does. Modelling physical heterogeneity in aquifers is still very important topic with myriad of theoretical and practical significances.

Does the paper present novel concepts, ideas, tools, or data? To some extent. The zonation approach to address the non stationarity of the K distribution is not new and it has been used in other previous studies. However, the integration of geological and geophysical data to derive K estimates is interesting. However, it is not completely novel since the same group of authors already presented the same approach in a recent previous publication (Zhu et al,. 2016b) based on the same dataset.

Are substantial conclusions reached? To some extent. Results indicate that the proposed method can be used to model *K* distribution in alluvial fans. However, this conclusion is mostly based on a qualitative assessment rather than a rigorous direct or indirect validation method. Moreover, it seems to me that a very similar conclusion was reached in their previous work (Zhu et al., 2016).

Are the scientific methods and assumptions valid and clearly outlined? To some extent. The methodology section could be expanded to better explain the criteria based on which the four facies were chosen, and how the lithological model in Figure 7A was developed.

Are the results sufficient to support the interpretations and conclusions? Yes.

Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? To some extent. A description of how lithofacies were defined and simulated could be beneficial.

Do the authors give proper credit to related work and clearly indicate their own new/original contribution? Generally yes. However, they should explain the novelty of this work with respect to their previous one.

Does the title clearly reflect the contents of the paper? Yes.

Does the abstract provide a concise and complete summary? Yes.

Is the overall presentation well structured and clear? To some extent. See comments.

Is the language fluent and precise? Some descriptions sound colloquial.

Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? Yes.

Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? Yes. See comments.

Are the number and quality of references appropriate? In part. See comments.

Is the amount and quality of supplementary material appropriate? None provided.

General comments

Although the methods described here are not new, the paper presents an interesting case study that could be worth of publication. However, I think that major revisions are necessary to address the following issues:

1) What is the novelty of the results presented in this work with respect to the recent publication by the same group of authors (i.e., Zhu et al., 2016b)? This not clear to me, and I think it should appear in the introduction of this paper, together with the motivation for expansion or improvement (if any).

2) I do not understand how the facies C, FS, MS and G have been defined. Are the names of these units referring to the prevalent grain size value? The reason for this question is that the *K* values estimated for these units in the three zones show inconsistencies. In table 2, for instance, how do the authors explain that the facies called "Medium-coarse sand" in zone 3 is less conductive (0.81 vs. 1.07) than the facies called "Fine sand" in zone 1? Is the "fine sand" unit in zone 1 the same as the "fine sand" unit in zone 3? If yes, then why the mean *K* value is about 8 times larger? What I mean is that a deposit consisting largely of fine sand is a "fine sand" regardless of its location with respect to the apex of the alluvial fan. The same applies to all the other units. It is the proportions of these facies that changes with distance from the apex, and these changes are responsible for the non stationarity of the K distribution. The average *K* value of the lithofacies (e.g., "fine sand") should be consistent (plus minus uncertainty) between zones. Please clarify this point because it seems to me a major flaw of the proposed methodology. I also wonder if a different classification with more units would be also more appropriate in this case. In particular, fan deposits especially in the proximal part (your zone 1) are also characterized by debris flow deposits (matrix supported gravel). This type of deposits has been ignored.

3) I think it would vastly improve the impact of this analysis if the authors can include some quantitative assessment of the accuracy of the reconstructed K distribution. For instance, it seems that there is enough data to apply a split sample validation test. A comparison with results in which K is assumed stationary would also be beneficial.

Specific comments

1) Line 34. The reference Zappa et al. (2006) refers to a different depositional environment. Please use a more specific reference.

2) Line 35. I suggest to add "hydraulic".

3) Line 65. Insert a period before "The Chaobai".

4) Line 79. Deposited instead of "laid down"

5) Line 156. Please specify what you consider as representative grain size diameter. Is it the d_{10}?

6) Lines 161 – 162. Only ranges are provided for the parameters. What specific values have been used? Why? Please explain.

7) Equations 2 and 3. How do the systematic errors and uncertainty in the parameters associated with these empirical equations affect the uncertainty in the K estimations? Some comment on this would be beneficial.

8) Line 169. I suggest to show the histograms of the *K* values within each facies to justify the lognormality assumption.

9) Line 172. What information? Grain size analyses?

10) Line 180. Are those data different then the data used in this work? The fact that you say that you found consistency suggests that they are different, but then I wonder why these 694 boreholes were not considered. Please clarify.

11) Line 217. I understood that the volumetric proportions p_k were derived from the borehole data rather than estimated through inversion. Please clarify.

12) Line 244. Variance of what?

13) Line 246. "… is highly uncertain"

14) Line 248 - 249. This sentence does not find correspondence in your analysis. On the other hand it confirms my doubt that the facies classification at the basis of the proposed methodology is not correct (see general comment 2). You rightly write "The deposits consist of wide ranges of sediment categories and grain sizes" to justify the fact that *K* has higher variance. But this is not the case in this work because the presented analysis is based on the assumption of only four (even three here) units G, FS, MS. So it seems to me that in order to include that variability you are talking about, your units do not actually represent a specific lithology as the names imply but a wider range of lithologies. For instance, your unit called "Fine sand" may actually include deposits that would be classified as fine sands as well as fine gravels. Am I wrong?

15) Line 250. Heterogeneity of what?

16) Line 254. Sorting and grain size are not the same. A poorly sorted sediment can still have a very high *K*.

17) Line 263. I suggest to provide the variance also for zone 1. After all, if I understood correctly, there are 102 samples. This is not such a small number.

18) Lines 266 – 269. Please revise the sentence. The meaning is not clear.

19) Lines 274 - 276. Same comment as #14.

20) Lines 280 – 281. Is this conclusion not obvious? I understood that the SGS realizations of *K* are mapped on the basis of the facies model (Figure 7a).

21) Line 281 – 282. This should not be caused by assigning larger average *K* of three units. This should be the consequence of the fact that coarser units are more frequent in this zone and therefore the average *K* is larger. The *K* distribution is the product of the lithological heterogeneity; it is not the opposite as it is implied here and in general in the paper.

22) Lines 293 – 296. It depends. Are you considering arithmetic or geometric average?