Review of "Interpretation of Multi-scale Permeability Data through an Information Theory Perspective" by Aronne Dell'Oca et al. (2019).

Summary and Recommendation

In this manuscript (MS), the authors use a series of different methods taken from information theory to estimate and compare the information content of different permeability measurements of two geological settings.

Overall, the MS is structure and written in a clear manner. In addition, the general idea behind the study is interesting and of relevance for the potential readers of HESS. However, I have to admit that I was a bit surprised because I could not find a scientific discussion in the entire MS. While there a few interpretations of the results in section 4 there is not a single reference after page 7 neither a discussion. This left me with a lot of open questions like:

> How does your results link to the work of Tidwell and other who used for instance geo-statistics on the same data set.
>
> Which findings are new and could have not been drawn if you have used of more classical statistical approaches instead of Information theory?
>
> What are the merits of using Information Theory if most of your conclusions can be drawn by looking at the pdfs in figure 3 a.

Given the nicely written introduction and the overall interesting topic of the MS I am however very positive that the authors are able to re-work this MS in a way that it can be published in HESS. For this, I believe, however, that a substantial amount of work needs to be put in this MS before it can be published.

Major comments:

No scientific discussion and comparison with other research.

Technical comments:

Section 3.2 Implementation Aspect Line 247-269: Here, the authors chose a couple of crucial parameters, which are in my opinion not all well justified. For instance, they use a kernel density estimator to estimate their pdfs from their datasets. However, they give not much details how they chose their related parameters, neither how changing them influences the results nor why they do this besides stating that: "*This step enables us to smooth and regularize the available finite datasets*".

How do you know that you do not smoothed out information that is of relevance?

Furthermore, why do you chose 100 bins. Is this choice based on, for instance, the measurement uncertainties (physics; e.g. *Loritz et al. 2018*) or on a statistical analysis (statistics; e.g. *Gong et al. 2013*)? How do your results change if you only pick 50 bins? Remember that the bin width is pretty much your a-priori assumption of similarity so you need to be careful here.

Line 83: Well, again you need to choose your bin width, which is a strong a-priori assumption.

Line 106: Information is always about something. Please be more specific here.

Line 155: The formula is correct but rather uncommon in this form.

Line 162: The nature of the Shannon entropy does not change if you use *nats,* however, I would argue that the interpretation is much more straightforward if you use the binary logarithm calculate it. This is the case because it is then directly linked to the average number of binary questions one needs to ask to infer in which state X is as well as it is then directly linked to the maximum compressibility of your dataset. A perfect lossless compression is thereby a perfect upscaling.

Line 244 – 246: Why? Could you explain that in your specific context.

References:

Gong, W., Yang, D., Gupta, H. V. and Nearing, G.: Estimating information entropy for hydrological data: One-dimensional case, Water Resour. Res., 50(6), 5003–5018, doi:10.1002/2014WR015874, 2014.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U. and Zehe, E.: On the dynamic nature of hydrological similarity, Hydrol. Earth Syst. Sci., 22(7), 3663–3684, doi:10.5194/hess-22-3663-2018, 2018.