review:WR034441

May 2023

1 Third reviewer (Shlomo P. Neuman)

The authors have conducted an interesting statistical analysis of raw variograms (I assume that this is what the authors mean by variogram clouds) of published data from a variety of geographic locations and geologic settings. Though the data are said to include saturated hydraulic conductivities from aquifers and shallow soils, as well as permeabilities and transmissivities, I assume (but the authors need to confirm or otherwise clarify) that the data consist of logarithms of these variables. A strength AND a weakness of the analysis is the lack of clear distinction between varied hydraulic properties (conductivity versus transmissivity) from varied settings (sediments, rocks, porous versus fractured media) measured by varied means (I presume single and multihole pressure tests, perhaps other) on varied scales of measurement. The strength of the approach is that it reveals commonalities among the data (scaling and amenability to analysis by a variety of variogram models); its weakness is lack of clarity about the way hydrogeologic settings, methods and scales of measurements, affect variogram behavior.

We appreciate the reviewers comments and overall supportive feedback on our study. The reviewer is right that we used the logarithm of the conductivity, transmissivity and permeability values. To make this more clear, we explicitly mention this now in the revised version of the manuscript. As regards the lack of clear distinction between varied hydraulic properties, we agree with the reviewer that, as far as the analyses are concerned, this is both a strength and weakness. Any statistical analysis has to find a compromise between sample size and accuracy. Using the whole sample means that no within sample effects can be discovered. Split the sample according to different criteria, and the resulting sub-samples may be too small to perform a viable statistical analysis. For our paper, we only distinguished between soil and aquifer site. Splitting for instance the aquifer data further into conductivities, transmissivites and permeabilities, would facilitate the investigation of possible differences between them. Alas, since most of data was drawn from measurements of hydraulic conductivity, the samples for transmissivites and permeabilities would have been small, thus, making comparisons more elusive. Still, the data set is openly available and practitioners interested in this topic are free to do such an analysis by themselves. The same problem of reduced sample size also applies to comparisons

between, say, different settings and/or measurement techniques. This was particularly unfortunate, since we were very interested in this topic ourselves. Alas, many papers did not report these additional information, which exacerbated the aforementioned problem even further. To better highlight this problem we revised the conclusions portion of the manuscript accordingly.

To me, one of the more interesting results of the analysis is confirmation of a scale effect that reveals itself, precisely, when one throws data from varied sites and settings into a single basket, as do the authors. It would therefore behove the authors to take the additional step of attempting to fit a truncated power variogram to their data, thereby potentially showing that such a variogram captures a greater range of behaviors with fewer parameters than do the stationary variograms they consider. Not only might a truncated power variogram prove to be more general and parsimonious than do standard stationary variogram, but they would be much more suitable than the latter when one needs to predict, using a transport model, how a plume of contaminant might migrate and behave outside the domain originally used to define the variogram.

I think that with additional work along the lines I have just proposed, this paper could stand out as a much more significant contribution to the literature than it does now.

We agree that using the available data set with a truncated power variogram can help to investigate the scaling behavior of subsurface media indicated by a number of different studies. We therefore re-did the analysis using this variogram function. Our results show that a truncated power law variogram achieves a similar goodness of fit compared to the other variogram functions but does not show any scale dependencies of its parameters. Depending on the application, such a lack of scale dependency can be considered an asset. We will include these results and some discussion on this topic into the manuscript.