review:WR034441

May 2023

1 Second reviewer

The paper presents present a data-driven approach to a classical problem in subsurface hydrology, the estimation of parameters characterizing the variogram of subsurface properties. The proposed method advocates the use of Bayesian inference to set up a prior distribution for models that describe spatial correlations (covariance or variogram). A remarkable data set is examined, and available data are classifiable into primary (point measurements), secondary (empirical variogram functions), or tertiary (statistical estimates of subsurface properties) data. Data were processed to avoid overlaps and over-representation.

The first available result from the manuscript is the comparison between different variogram model functions, that could be improved in my view (see comment #2). The scale dependency of the hydraulic conductivity is examined next, confirming earlier literature results. An example application of the Bayesian approach to the estimate of correlation length given the maximum length scale is then presented. Other variogram parameters examined are anisotropy, nugget effect, and shape parameter.

The discussion addresses important issues such as the unbiasedness of the data set employed, among others.

The paper looks as a mature contribution; given the topic and the type of paper, I see however some room for further improvement. Results are of interest to the readership of Hydrology and Earth System Sciences. The methods are adequate, the paper subdivision into sections sound, and the figures illustrative. I recommend minor revisions for the reason explained below.

We appreciate the reviewers comments and overall supportive feedback on our study.

The manuscript examines only stationary variograms, I suggest to mention that nonstationary variograms (see, e.g.,Di Federico and Neuman, 1997) were excluded from the analysis. Due to some comments of another reviewer, we revised our analysis and included a truncated power law variogram.

The comparison among variograms having a different numbers of degrees of freedom (section 3.1) could be rendered more qualitative by model identification criteria (AOC, AIC_c, KIC, . . .), incorporating the number of parameters involved and the principle of parsimony. The same holds probably for other comparisons performed. We agree with the reviewer that the use of a model-selection criterion can formalize the comparison we have done in the manuscript to far. Models with more degrees of freedom have more flexibility and should therefore be better able to capture an observed behavior. In case of variogram model, the Matern and the Stable model should therefore outperform the other models in terms of goodness-of-fit. In the current version, we only performed a qualitative analysis by observing the overall similar accuracy and noting how this would not justify the use of a more sophisticated model like the aforementioned Matern and Stable model. Using a criterion like AOC, AIC_c, KIC, etc. would make this analysis more quantitative.

How crucial is the assumption of two independent Gaussian distributions in Section 3.2 ? Could they develop a more general theory without it, maybe subject to other limitations? The assumption of two different Gaussian distribution is not crucial to the approach presented there at all. In fact, the parametric model for the residuals around the regression line was chosen on the spot after visually inspecting them. As we explain in the discussion section, many different parametric models may be possible depending on the situation. In fact, if enough data points are available a completely non-parametric approach is possible as well. To summarize our approach here again, we would describe it as follows. First, the residuals around the regression line are representing the uncertainty one has with respect to the regression model. From a Bayesian perspective, the can be used to estimate a prior probability. We do this by first visually inspecting the results of a kernel-density estimation (KDE), ie., a non-parametric estimation procedure. KDE is a powerful estimator, but it always produces very smooth densities which may bamboozle practitioners into over interpreting its results. To avoid overconfidence, we therefore only use KDE to find a good parametric model that could describe the empirically observed distribution of the residuals. In our case, a mixture mode using two Gaussian distributions seemed like a good choice. When we fitted such a model to the residuals and compared it to the KDE, we saw a excellent overlap. Of course this agreement has to be interpret with care since the KDE is not the ultimate benchmark of truth, for the reasons outline above. But having two different estimation procedure give very similar results certainly adds confidence that the both express some underlying truth. This single example is explicitly presented as a proof of concept for how to use the data provided in our study for the derivation of prior distributions in a Bayesian context. As we state in the manuscript, using other data and/or other variogram functions may lead to a somewhat different regression analysis with different residuals. In our opinion, there is probably no general theory on what parametric model, if any, to use for the description of the prior distribution. Every case may be different and practitioners are advised to use their judgement in adapting this approach to their situations. In the revised manuscript, we now explain this reasoning in more detail to better convey this important idea.