

reply:hess-2023-15-MS

July 2023

Dear authors, The paper is fine and I like it (see the bottom part of this document for some very minor suggestions). But I had severe difficulties when making sure that the data repository is accessible. T

Dear Jesus,

thank you again for editing our manuscript. Please find attached the revised version of the manuscript and our responses. We hope that the revisions to our manuscript and our responses address your wishes and look forward to hearing from you.

Yours, Falk

- he links in the paper do not work properly (check them in an “untrained” computer, i.e., any non-UFZ computer). In fact, the link in line 581 leads to <https://www.copernicus.org/>. Anyway, since the links led me to invalid webpages, I tried to search directly in Github.com. When I search for “geostatDB”, I get “This is the first release of the geostatDB package. geostatDB is an R package that provides access to the World Wide Hydrological Parameters Database (WWHYPDA). When I search for “geostatDB”, “GeoStat-Examples/GeoStat-DB”, or “GeoStat-Examples” I get no repository found. Finally, I “accidentally” got to the GeoStat-Examples repository when looking for GStools, but could not find the csv files. This discussion proves that I am a clumsy github user (largely a passive one). But I feel you should facilitate the work by providing the right indications on how to access the repository. Minor comments (take these as suggestions, you do not need to follow them).

The fact that you could not access the [GitHub](#) repository is indeed a problem since the data collected for this paper is one of its main assets. However, when we tried the link, it worked even on non-UFZ computers. To be absolutely sure that data can be accessed we made the link to the [Zenodo](#) repository better available. As we already comment in the “Code and data availability” section, the data on the [GitHub](#) repository are subject to change as more data become available. The [Zenodo](#) link on the other hand provides a compressed archived of the Python scripts and csv data files as they were used for this manuscript. They do provide a snapshot of the [GitHub](#) repository as it was at the time the manuscript was submitted.

- Line 15: “soils” may perhaps be better than “soils sites”
We changed this throughout the manuscript.
- Line 69: You may drop “In the next section,”
We changed this as suggested.
- Line 75: “we critically” better than “we will critically”
We changed this as suggested.
- Line 144: With your definitions (your definition of σ^2 is awkward to me), it is not correct $C(h)=\sigma^2 \rho(h)$, as $C(0)=s$. You may either point at the discontinuity of $C(h)$ at the origin or, perhaps simpler, write $C(h)=s-\gamma(h)$
We decided for $C(h) = s - \gamma(h)$ and changed it accordingly in the manuscript.
- I am somewhat disappointed that you did not follow the advice of referee 1 (which I supported in the decision letter) of acknowledging that since the variogram is usually the only reference to geology in geostat analyses. Your paper confirms beautifully that data does not suffice to choose the variogram. It is geology what does, which is why structural analysis is the most important step in variogram estimation. This is essential for the behavior at the origin (which you hardly discuss) and for the range (which you do). I accept that, for discussion purposes, you take the maximum length scale as prior. But, for each site, you should use the variability scale that geology tells you. In fact, this is the one I have always used rather than fitting to the sample variogram (for which there is no convergence proof). This is particularly true for the vertical direction (as proven by your discussion in lines 355-360). And this deserves a discussion.
We agree that this point and its ramifications deserves a dedicated discussion in the manuscript. As a result, we added a paragraph to the section discussing the limitations and possible pitfalls of our analysis. In this paragraph, we stress how a purely data-drive analysis cannot be a substitute for good domain knowledge and site-specific expertise in the determination of an appropriate variogram model and successful geostatistical investigation.
- Please, check figure 8. It looks like three, rather than two, Gaussians.
We agree that the kernel-density estimate may be even better approximated using a mixture model of three Gaussians. However, our aim here was to show that even with a comparably simple model, an excellent fit could be achieved. We revised the accompanying text to better convey this point.
- Line 365: The distinction between “measurement errors and unresolved variations in the measured variable below the measurement scale” is fundamental both conceptually (again, importance of structural analysis) and practically (Kriging and simulations are different). While not affecting

your analysis, you should revise the statement.

We revised the statement to make clear that the two causes are very different.