Comment on gmd-2021-174

Anonymous Referee #1

Referee comment on "SciKit-GStat 1.0: a SciPy-flavored geostatistical variogram estimation toolbox written in Python" by Mirko Mälicke, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-174-RC1, 2021

General comments

The author presents a variogram estimation Python package SciKit-GStat based on SciPy. Variogram is the core function for geostatistical methods to describe spatial covariance of the data and is the determination factor for the application of geostatistics such as Kriging interpolation. SciKit-GStat includes many commonly used variogram related algorithms of variogram estimators, theoretical spatial models and distance lag binning, also advanced models of directional variogram and space-time variogram are implemented. Although the limitation of Kriging methods implementation exists in the package, the interface to other geostatistical packages such as gstools is provided for final steps of geostatistical analysis. SciKit-GStat was well coded and documented as an efficient and ease-used variogram toolbox.

A major revision is needed to improve the manuscript before the publication.

Specific comments

- The pancake dataset was used as example data for the demonstration of the functions in SciKit-GStat and the author described the advantage of it in appendix A. But I still think a complicated real-world data example such as precipitation should be given to support the powerful of the package. Also, such case can give more realistic variogram usage clues to the users.
- The author declared that the SciKit-GStat version is 1.0, but only version 0.6 (or 0.6.6) of it could be found in github and online document website.
- Why non-Gaussian geostatistics are not covered in SciKit-GStat?
- Why the procedures that can fit a model directly based on unbinned data are not implemented in SciKit-GStat?

Minor comments

- Color bars should be plotted in figure 1, 4, and 8.
- Line 99, change "SciKit-Gstat" to "SciKit-GStat".
- Line 212, remove one "all".
- Figure 3 was not explained clearly.
- Figure 6, it's better to set transparent color to the surface part so the distribution of the scatter data could be clearer.
- The data source of figure 6 was not described.