

EGU2020-6678

<https://doi.org/10.5194/egusphere-egu2020-6678>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



SciKit-GStat: A scipy flavored geostatistical analysis toolbox written in Python

Mirko Mälicke

Karlsruhe Institute of Technology (KIT), Institute of Water and River Basin Management, Hydrology, Freiburg, Germany
(mirko.maelicke@kit.edu)

Geostatistical and spatio-temporal methods and applications have made major advances during the past decades. New data sources became available and more powerful and available computer systems fostered the development of more sophisticated analysis frameworks. However, the building blocks for these developments, geostatistical packages available in a multitude of programming languages, have not experienced the same attention. Although there are some examples, like the gstat package available for the R programming language, that are used as a de-facto standard for geostatistical analysis, many languages are still missing such implementations. During the past decade, the Python programming language has gained a lot of visibility and became an integral part of many geoscientist's tool belts. Unfortunately, Python is missing a standard library for geostatistics. This leads to a new technical implementation of geostatistical methods with almost any new publication that uses Python. Thus, reproducing results and reusing codes is often cumbersome and can be error-prone.

During the past three years I developed scikit-gstat, a scipy flavored geostatistical toolbox written in Python to tackle these challenges. Scipy flavored means, that it uses classes, interfaces and implementation rules from the very popular scipy package for scientific Python, to make scikit-gstat fit into existing analysis workflows as seamlessly as possible. Scikit-gstat is open source and hosted on Github. It is well documented and well covered by unit tests. The tutorials made available along with the code are styled as lecture notes and are open to everyone. The package is extensible, to make it as easy as possible for other researchers to build new models on top, even without experience in Python. Additionally, scikit-gstat has an interface to the scikit-learn package, which makes it usable in existing data analysis workflows that involve machine learning. During the development of scikit-gstat a few other geostatistical packages evolved, namely pykrige for Kriging and gstools mainly for geostatistical simulations and random field generations. Due to overlap and to reduce development efforts, the author has made effort to implement interfaces to these libraries. This way, scikit-gstat understands other developments not as competing solutions, but as parts of an evolving geostatistical framework in Python that should be more streamlined in the future.