

# GeoStats.jl – High-performance geostatistics in Julia

Júlio Hoffmann ([juliohm@stanford.edu](mailto:juliohm@stanford.edu))<sup>1</sup>

<sup>1</sup> Stanford University

DOI: [10.21105/joss.00692](https://doi.org/10.21105/joss.00692)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 21 April 2018

Published: 24 April 2018

## Licence

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

[GeoStats.jl](#) is an extensible framework for high-performance geostatistics in Julia, as well as a formal specification of statistical problems in the spatial setting. It provides highly optimized solvers for estimation and (conditional) simulation of variables defined over general spatial domains (e.g. regular grid, point collection), and can utilize high-performance hardware for parallel execution such as GPUs and computer clusters.

Its unique design addresses the very important issue of scientific comparison between different geostatistical methods proposed by the research community. Unlike similar software (e.g. GSLIB, SGeMS, gstat), which implement algorithms for specific data and domain types with varying interfaces, GeoStats.jl introduces an abstraction layer with which users can define their problems precisely once, and switch between different solvers effortlessly. The same abstraction layer enables the development of higher-order routines that operate on solvers as first-class objects (e.g. cross-validation), a feature that gives researchers the ability to experiment with various geomodeling assumptions programatically.

Besides its technical contributions, the project aims to educate people outside of the field about state-of-the-art methods in geostatistics, their assumptions, and their limitations.

## Problem types

The framework currently defines two types of problems:

- **Estimation:** given spatial data and domain, estimate variable(s) at unseen locations, and provide whenever possible a variance (or uncertainty) map.
- **Simulation:** given domain and (optionally) spatial data, simulate multiple realizations of variable(s) matching previously existing data if present.

## Available solvers

As of version 0.6.0, the following solvers are available.

- Estimation solvers
  - Kriging (Matheron 1971)
  - Inverse Distance Weighting (Shepard 1968)
  - Locally Weighted Regression (Cleveland 1979)
- Simulation solvers
  - Direct Gaussian Simulation (Alabert 1987)
  - Sequential Gaussian Simulation (Isaaks 1990)
  - Fast Image Quilting (J. Hoffmann et al. 2017)

## Comparison methods

Solvers adhering to the interface proposed in the framework can be easily compared on a given problem with different comparison methods.

- Visual Comparison
- k-fold Cross-Validation

For example, the command `compare([solver1, solver2], problem, VisualComparison())` solves the problem with `solver1` and `solver2` and plots the solutions side by side.

## Example of usage

```
using GeoStats
using Plots

# data.csv:
#   x,      y,      station, precipitation
# 25.0, 25.0,   palo alto,           1.0
# 50.0, 75.0,  redwood city,          0.0
# 75.0, 50.0, mountain view,          1.0

# read spreadsheet file containing spatial data
geodata = readtable("data.csv", coordnames=[:x,:y])

# define spatial domain (e.g. regular grid, point collection)
grid = RegularGrid{Float64}(100, 100)

# define estimation problem for any data column(s) (e.g. :precipitation)
problem = EstimationProblem(geodata, grid, :precipitation)

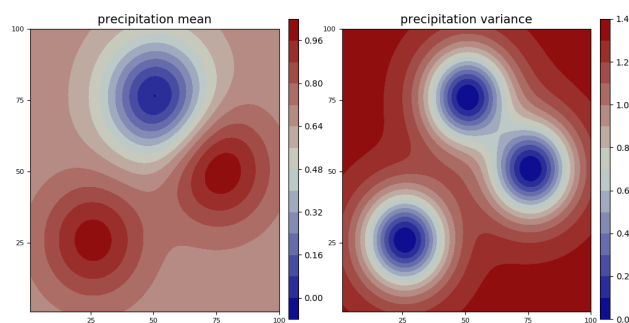
# choose a solver from the list of solvers
solver = Kriging(
    :precipitation => @NT(variogram=GaussianVariogram(range=35.))
)

# solve the problem
solution = solve(problem, solver)

# plot the solution
plot(solution)
```

## Usage in academia and industry

The solvers and tools implemented in the project have been used in both academic and industrial endeavours. To give an example, the [ImageQuilting.jl](#) (J. Hoffmann et al. 2017) solver has been used inside [ENI](#) to condition 3D process-based models to data acquired in Oil & Gas fields. It has also been used by researchers in Denmark interested in modeling groundwater resources (A. A. S. Barfod et al. 2017), and by researchers studying micromodels of porous medium in various research groups worldwide. Research colleagues at Stanford are currently using GeoStats.jl to model fractured reservoirs, mineral deposits, geothermal resources, and glaciers, among other spatial objects.



**Figure 1:** Estimation solution on a regular grid.

## Acknowledgements

I'd like to thank my research colleagues in the School of Earth Sciences at Stanford for their feedback, feature requests, and bug reports, as well as the Julia community for helping with internal design decisions.

## Funding

GeoStats.jl will always be open source and free of charge. If you are an organization that supports open source initiatives, please contact the author at [juliohm@stanford.edu](mailto:juliohm@stanford.edu) regarding funding opportunities.

## References

- Alabert, Francois. 1987. "The Practice of Fast Conditional Simulations Through the Lu Decomposition of the Covariance Matrix." *Mathematical Geology* 19 (5). Springer Nature:369–86. <https://doi.org/10.1007/bf00897191>.
- Barfod, Adrian A. S., Julien Straubhaar, Anne-Sophie Høyer, Júlio Hoffmann, Anders V. Christiansen, Ingelise Møller, and Jef Caers. 2017. "Hydrostratigraphic Modelling Using Multiple-Point Statistics and Airborne Transient Electromagnetic Methods." *Hydrology and Earth System Sciences Discussions*, August. Copernicus GmbH, 1–37. <https://doi.org/10.5194/hess-2017-413>.
- Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368). Informa UK Limited:829–36. <https://doi.org/10.1080/01621459.1979.10481038>.
- Hoffmann, Júlio, Céline Scheidt, Adrian Barfod, and Jef Caers. 2017. "Stochastic Simulation by Image Quilting of Process-Based Geological Models." *Computers & Geosciences* 106 (May). Elsevier BV:18–32. <https://doi.org/10.1016/j.cageo.2017.05.012>.
- Isaaks, Edward H. 1990. "The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data." PhD thesis.
- Matheron, Georges. 1971. "The Theory of Regionalized Variables and Its Applications." PhD thesis.
- Shepard, Donald. 1968. "A Two-Dimensional Interpolation Function for Irregularly-Spaced Data." *Proceedings of the 1968 23rd ACM National Conference*. ACM Press. <https://doi.org/10.1145/800186.810616>.