



gstlearn

2022 - 2027



GEOLEARNING
CHAIRE /// Data Science for the Environment



HISTORY

RGeostats : a R package based on a C library



- Written by a single developer Didier Renard (bus factor 1)
- Implement traditional Geostatistics



MAIN OBJECTIVES OF THE LIBRARY

- Provide a large family of spatial (+ temporal) covariance functions (kernels) with parameters
- Provide tools to estimate parameters from spatial (+temporal) observations
- Provide tools to predict at unobserved locations (kriging, conditional expectation)
- Provide tools to assess uncertainty (conditional simulations in Gaussian framework)



AMBITIONS

- Replace RGeostats by **gstlearn**, a C++ 20 open-source library (BSD-3)
- Platform Agnostic (Windows, Linux, Mac OS)
- The reference for Geostatistical Package (R/Python)
- Easy to Use (**minigst**: the companion packages of gstlearn)
- Implement **modern** and **scalable** approaches of spatial statistics for inference
- Most Complete Set of Features for spatial and **spatio-temporal Gaussian processes**
- Robust (CI/CD) and efficient (multi-threading, ...) code

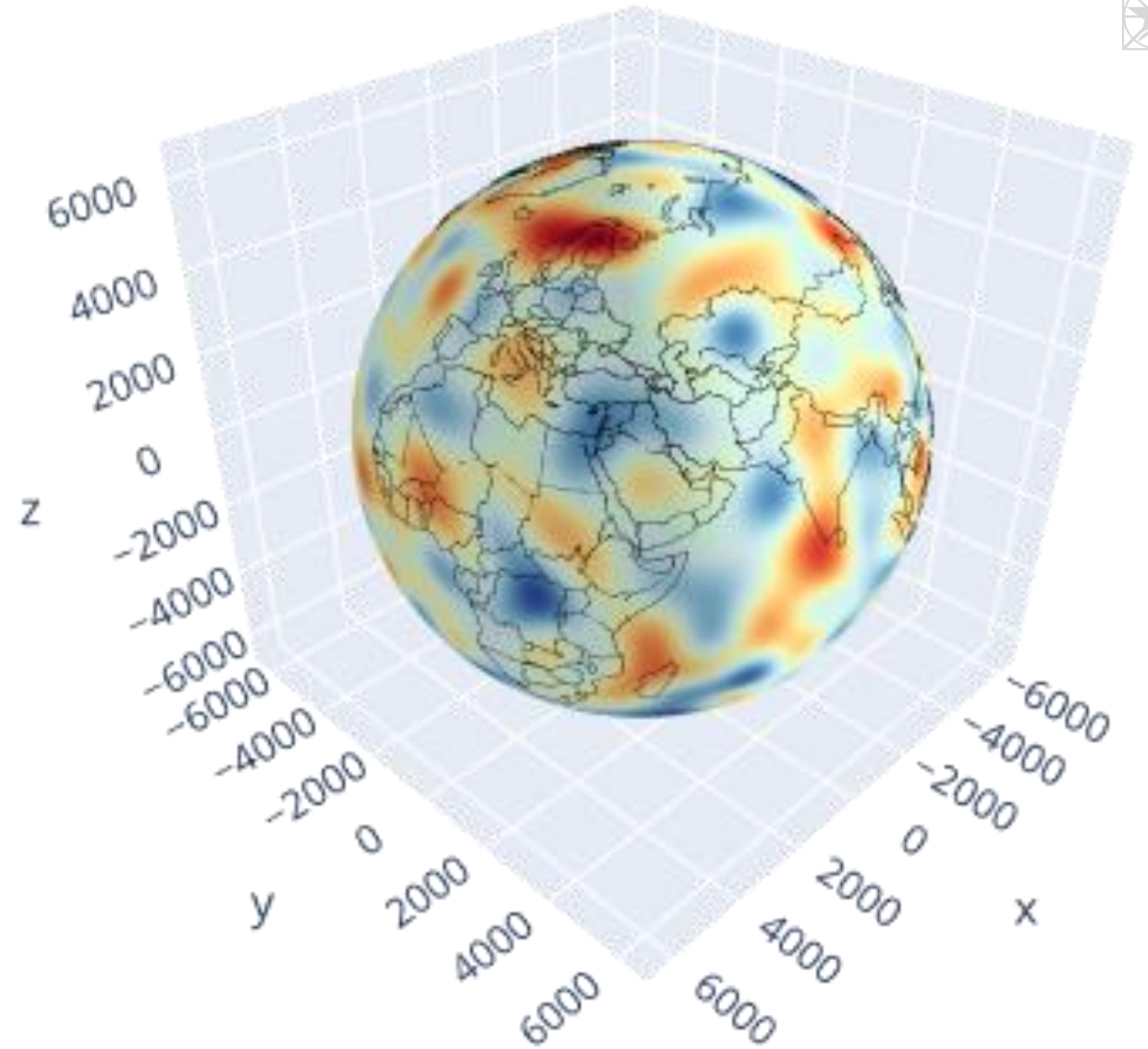


INFERENCE

- Maximum likelihood estimation
- Use of NLOpt library for optimization
- Gradient are computed by a mixing analytical gradients and auto-differentiation
- Vecchia approximation is implemented for very large data-sets (with ball-tree algorithm for efficient implementation)
- ...

NEW MODELS

- Gaussian processes on sphere

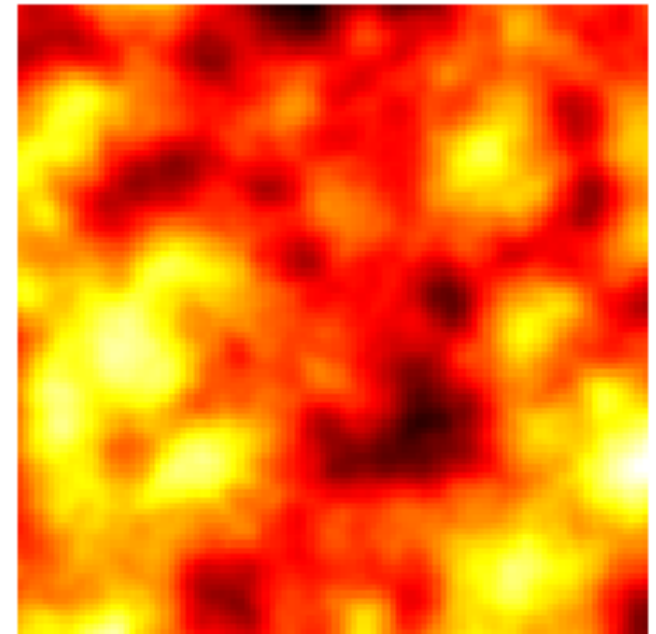
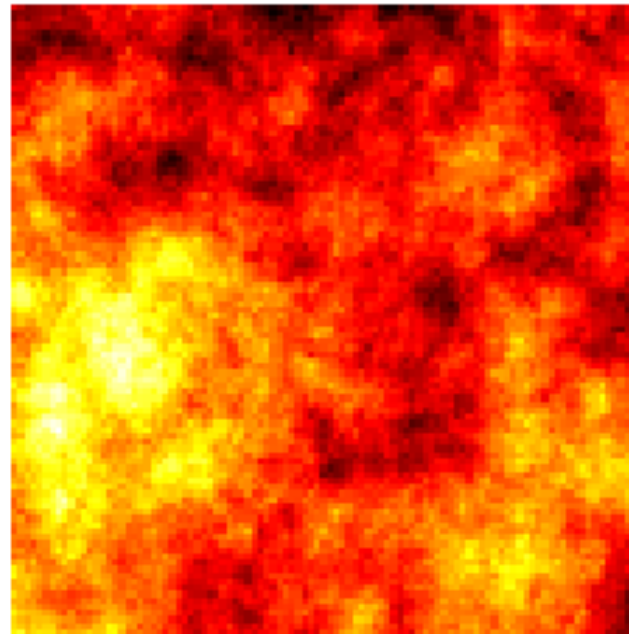


Lantuéjoul, C., Freulon, X., & Renard, D. (2019). Spectral simulation of isotropic Gaussian random fields on a sphere. *Mathematical Geosciences*, 51(8), 999-1020.



NEW MODELS

- Gaussian processes on sphere
- Multivariate Matern covariance

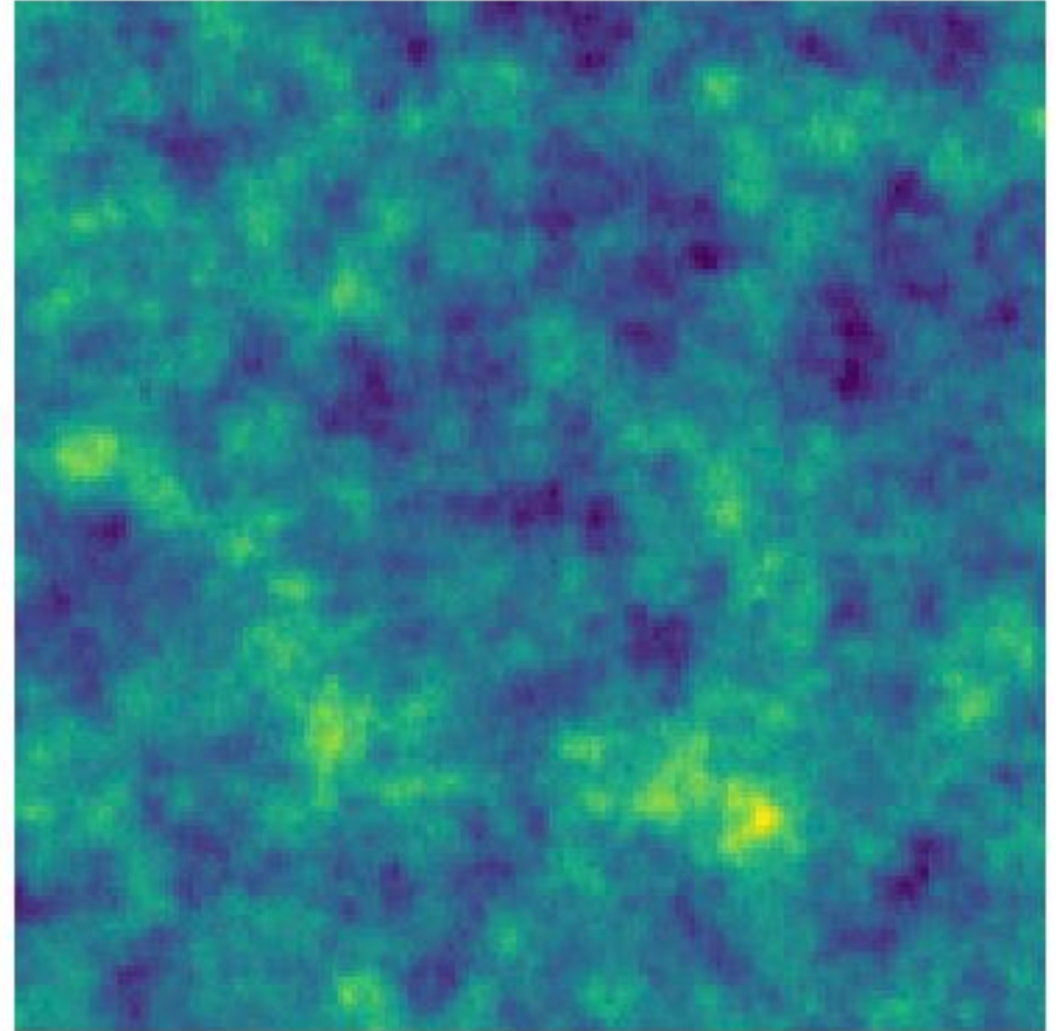


Schlather, M., Malinowski, A., Menck, P. J., Oesting, M., & Storkorb, K. (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software*, 63, 1-25.

NEW MODELS

- Gaussian processes on sphere
- Multivariate Matern covariance
- Multivariate Gneiting covariance

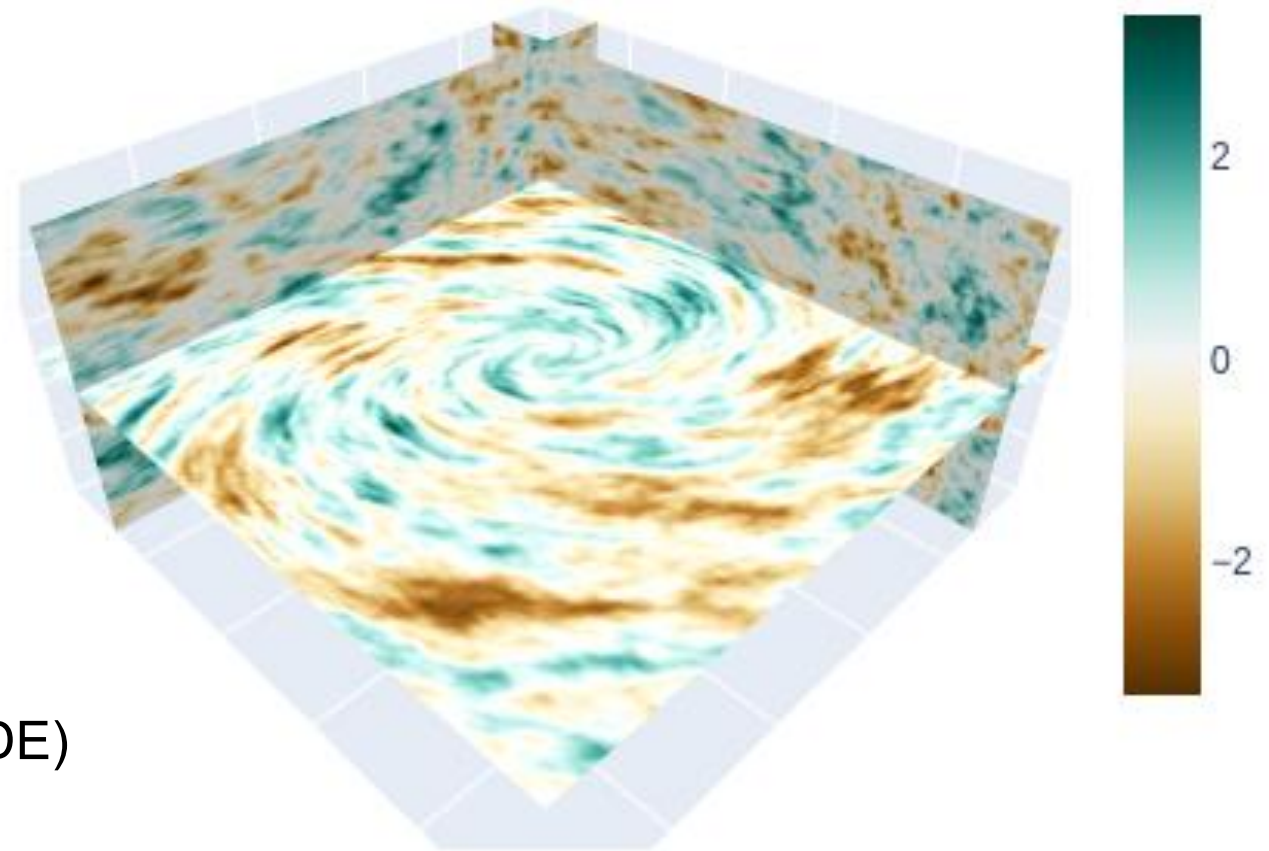
Time = 0



Allard, D., Emery, X., Lacaux, C., & Lantuéjoul, C. (2020). Simulating space-time random fields with nonseparable Gneiting-type covariance functions. *Statistics and Computing*, 30(5), 1479-1495.

NEW MODELS

- Gaussian processes on sphere
- Multivariate Matern covariance
- Multivariate Gneiting covariance
- Stochastic Partial Differential Equation (SPDE)

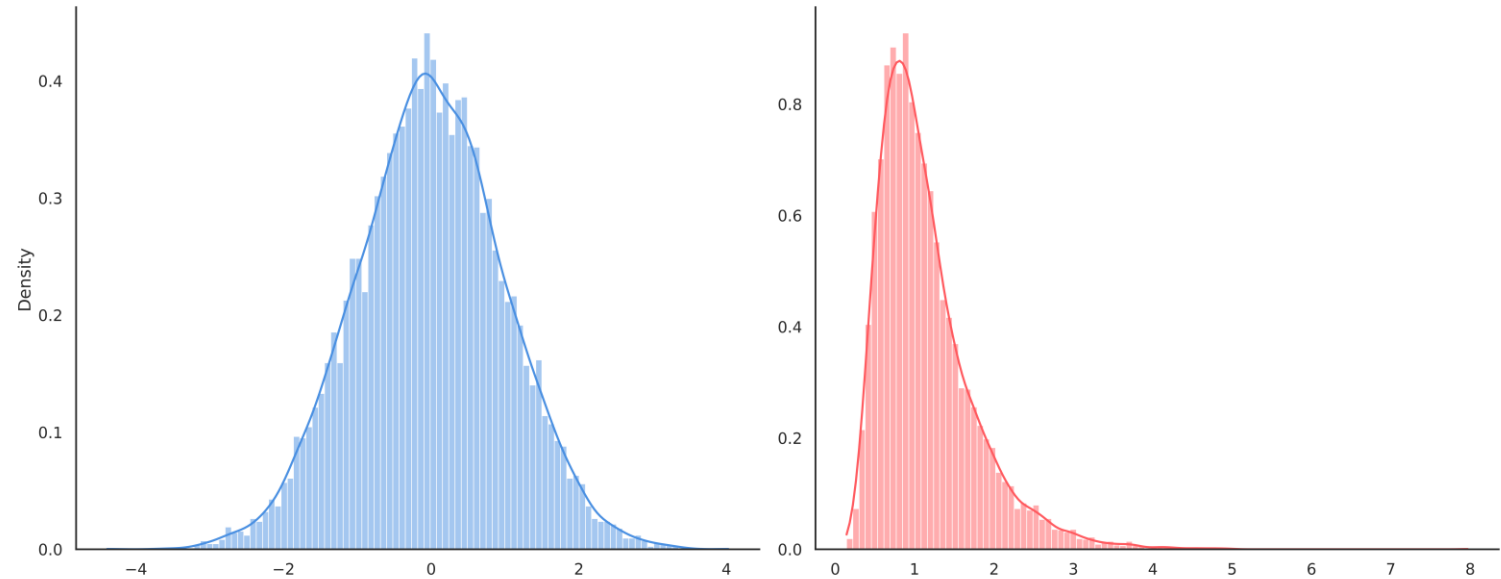


Pereira, M., Desassis, N., & Allard, D. (2022). Geostatistics for Large Datasets on Riemannian Manifolds: A Matrix-Free Approach. *Journal of Data Science*, 20(4).



NEW MODELS

- Gaussian processes on sphere
- Multivariate Matern covariance
- Multivariate Gneiting covariance
- Stochastic Partial Differential Equation (SPDE)
- Various transformations (Box-Cox, Yeo-Johnson, ...)





PIECE OF CODE

```
import minigt as mg
import pandas as pd

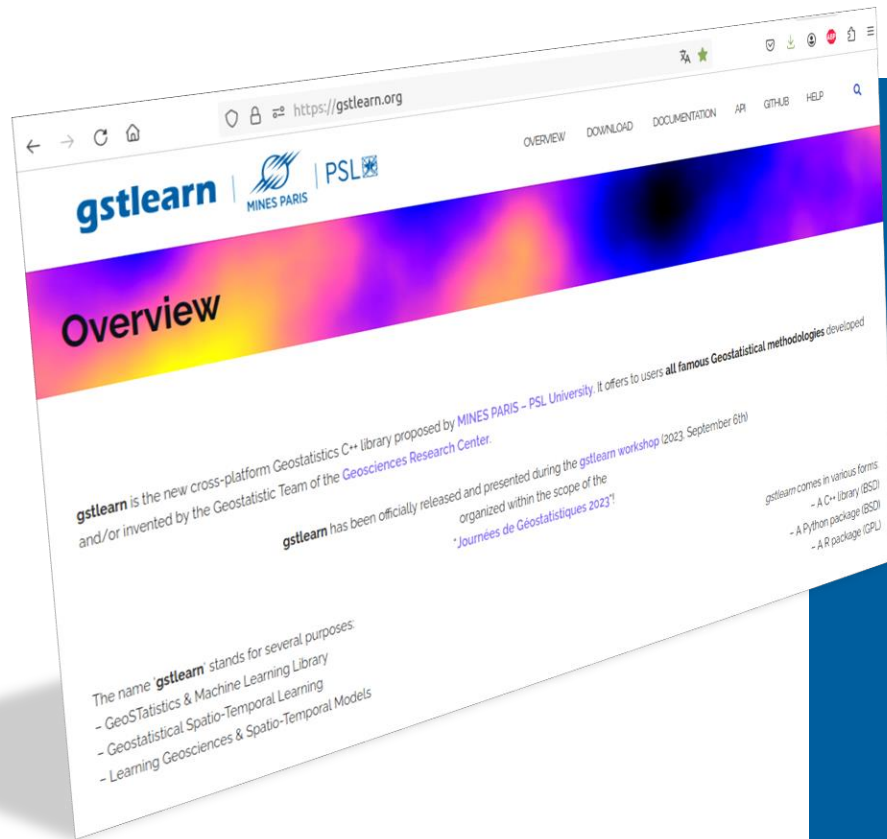
dat = pd.read_csv("data.csv")

#Data base and grid creation
db = mg.df_to_db(dat, coord_names= ["X", "Y"])
grid = mg.createCoveringGrid(db, [100, 100])
#Model creation and fit
model = mg.create_model("MATERN")
model.fit(data = db, variable_names = ["pollution"], nbVecchia = 20)
#Predict and display
model.predict(data = db, target = grid, variable_names = ["pollution"],
| | | | prefix = "kriging")
mg.dbplot_grid(grid, "kriging.pollution")
#Simulate and display
model.simulate(data = db, target = grid, variable_names = ["pollution"],
| | | | prefix = "simulation", nbsimus = 20)
mg.dbplot_grid(grid, "simulation.pollution.S1")
```

A LOT REMAINS TO BE DONE!

- Increase **minigst**
- GPU support for linear algebra
- Compatibility with GPyTorch
- Implement Spatial Extreme Values
- Write a QGIS plug-in
- ...

WEBSITE & LICENSES



URLs around **gstlearn**

<https://gstlearn.org>

C++ BSD3

<https://github.com/gstlearn/gstlearn>



<https://pypi.org/project/gstlearn>



Pypi

CRAN

<https://soft.mines-paristech.fr/cran/gstlearn.html>

BSD3

GPL3

<https://doi.org/10.5281/zenodo.18352742>

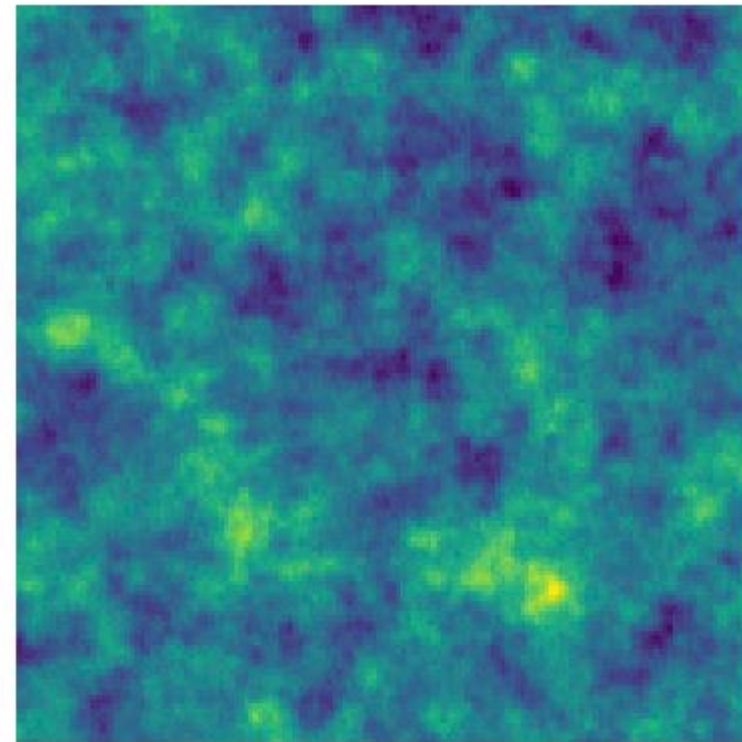
DOI 10.5281/zenodo.13343742

<https://github.com/gstlearn/minigst>

GSTLEARN LIBRARY FOR SPATIAL AND SPATIO-TEMPORAL GAUSSIAN PROCESSES

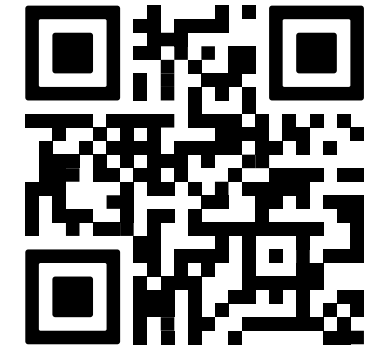
- Efficient parameter **inference** for a large class of models,
- Optimized **prediction** using conditional expectation (kriging),
- **Uncertainty** evaluation using conditional **simulations**.

Time = 0



Advected Gneiting covariance

gstlearn



Open-source
BSD-3 clause

