Gaussian process regression fast solver

A. Petrovskaia, C. Daniel, A. Uvarov, A. Usman

NLA course project
Skoltech 2017

Outline

Introduction

In numerical analysis, interpolation is an employed method of fitting a function by constructing new data points within the range of some discrete known set of points. So we basically use sampled points to estimate values at other unknown points.

Introduction

In numerical analysis, interpolation is an employed method of fitting a function by constructing new data points within the range of some discrete known set of points. So we basically use sampled points to estimate values at other unknown points.

There are numerous interpolation techniques, such as inverse distance method, natural neighbour, polynomial regression, splines and Gaussian process regression.

Gaussian regression

Gaussian regression, also known as Kriging, is a spatial prediction method that uses covariance structure to get a more precised, better prediction. It is an interpolator largely governed by the covariance prior of the data set.

Gaussian regression

Gaussian regression, also known as Kriging, is a spatial prediction method that uses covariance structure to get a more precised, better prediction. It is an interpolator largely governed by the covariance prior of the data set.
The model, majorly consist of two components, given by

$$y(x) = \mu + \epsilon(x)$$

where μ is the large scale variation and $\epsilon(x)$ the small scale spatial autocorrelation.

Gaussian regression

In regression model we know the error term is not important as it sums to zero. However, kriging extends to exploit the information about the spatial structure embedded in the covariance of the error term.

Gaussian regression

In regression model we know the error term is not important as it sums to zero. However, kriging extends to exploit the information about the spatial structure embedded in the covariance of the error term.

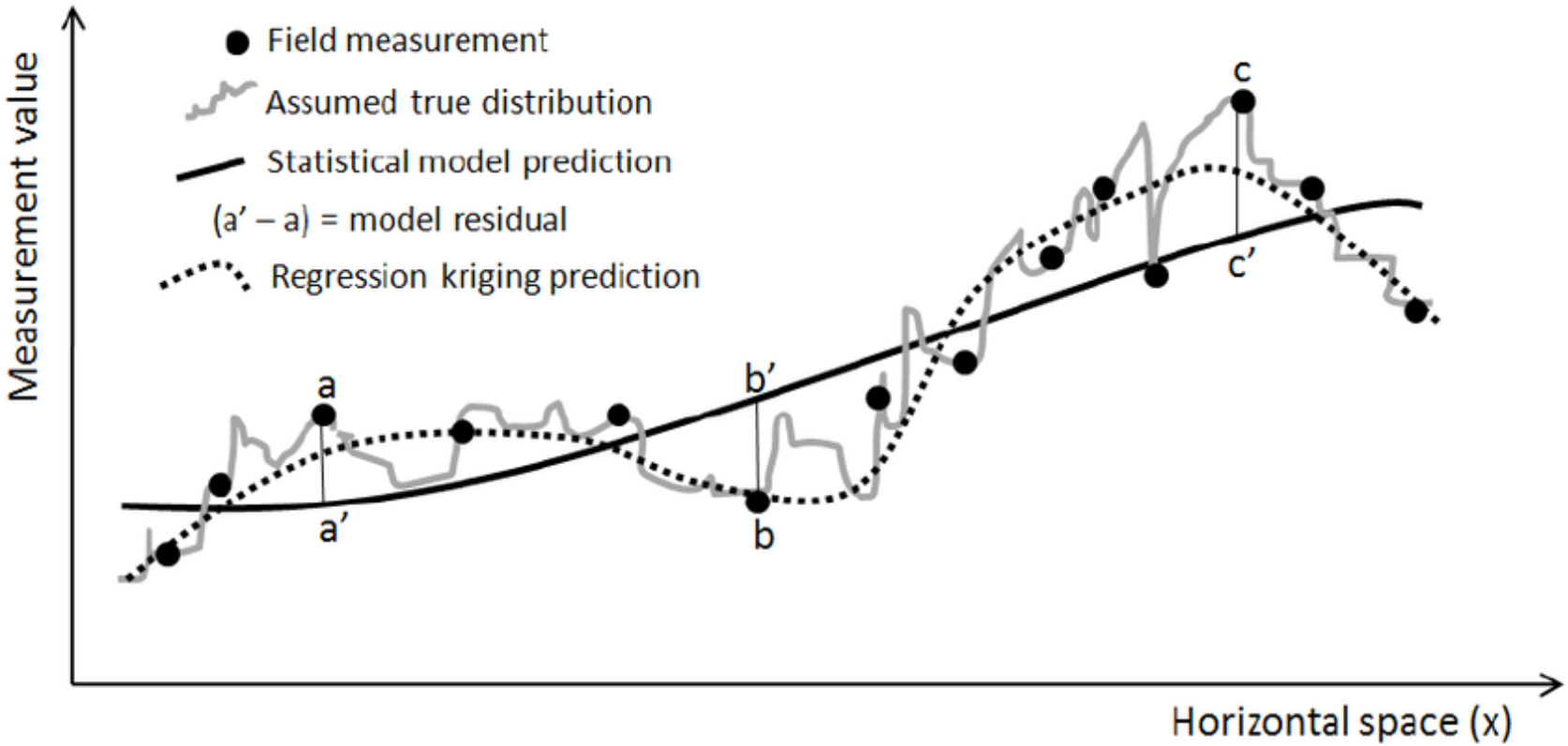


Рис.: Kriging Vs OLS interpolation
https://www.researchgate.net/figure/265075728_fig4

Model specification

Let x be a vector of sampled locations, y a vector of function values, z a vector of locations where we are interested in function value. Then, to find the function values y' at the points z, we do the following:

$$y' = \mu + K_{zx} K_{xx}^{-1}(y - \mu \mathbf{e}),$$

$(K_{xx})_{ij} = \exp(-\theta|x_i - x_j|^p)$, covariance of observed location

$(K_{zx})_{ij}$ covariance between the new locations and observed locations.

The parameters θ and p are estimated e.g. by method of maximum likelihood

Project goals

- Implement fast kriging on Python and test it on geodata

Project goals

- Implement fast kriging on Python and test it on geodata
- Write a plugin for QGIS realizing this solution

Fast kriging for evenly spaced grids

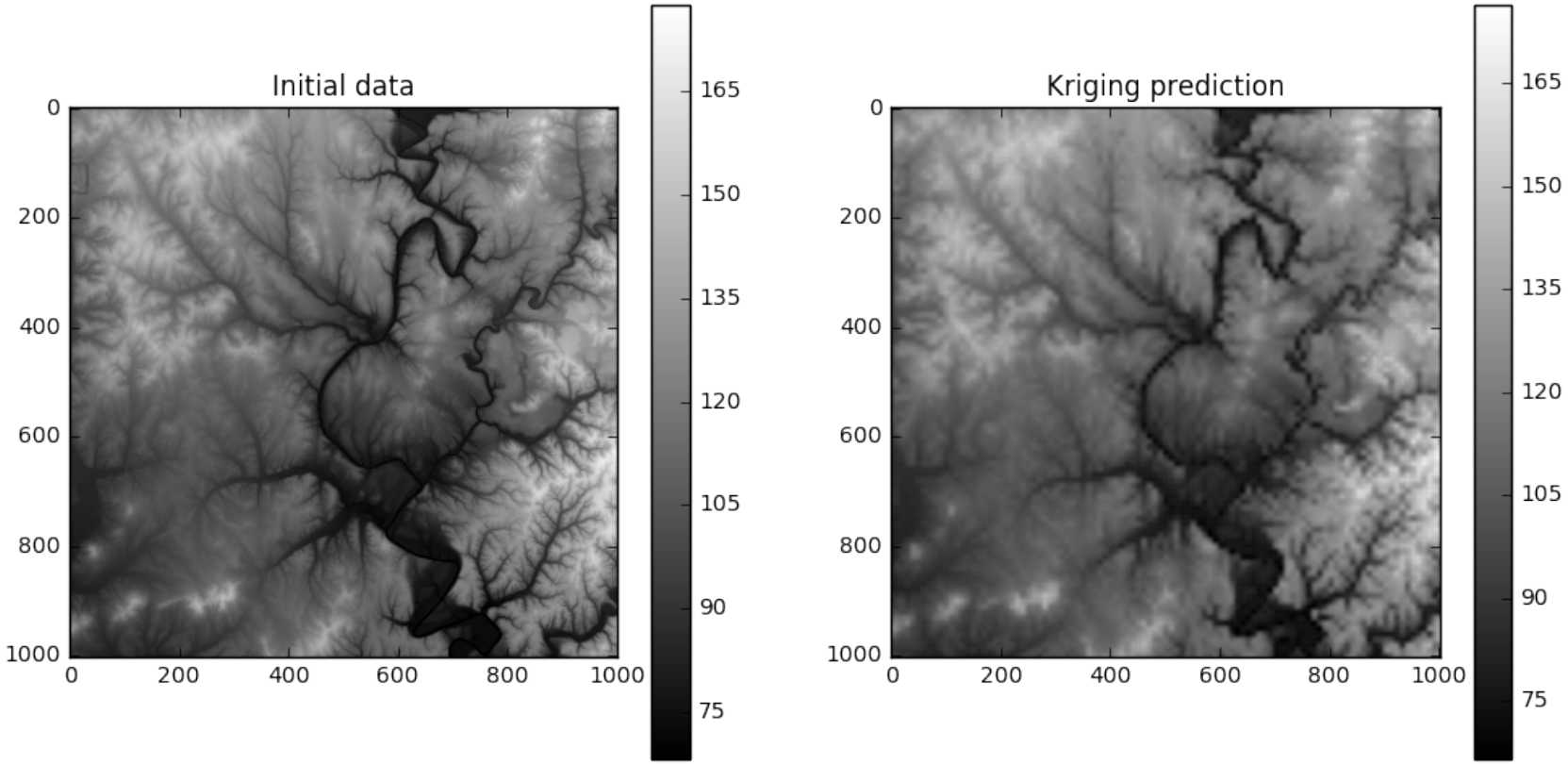
The structure of K_{xx} depends on selection of data points
Evenly spaced points yield a Toeplitz matrix (or BTTB for 2D):

$$(K_{xx})_{ij} = \exp(-\theta|x_i - x_j|^p) = (K_{xx})_{i+q,j+q}.$$

This allows us to perform fast matrix-vector product and thus solve linear systems faster

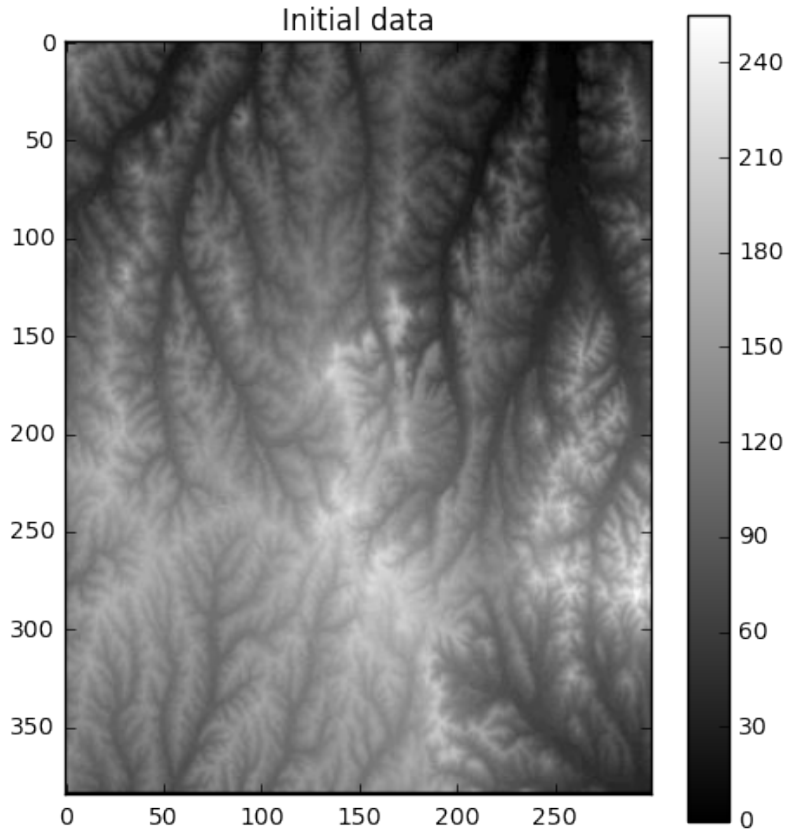
Results

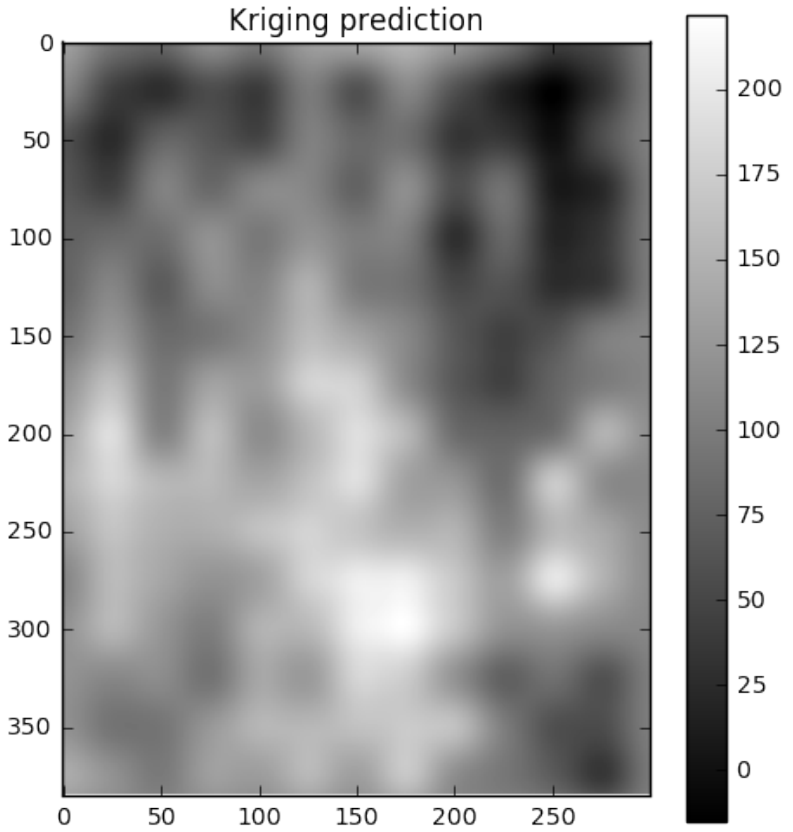
Prediction here is based on 100x100 points. For this test sample, $\|y' - y_{true}\|_2 / \|y_{true}\|_2 = 0.026$; $\sqrt{\sigma^2}/\mu = 0.132$



Results

Prediction here is based on 12x16 points. For this test sample, $\|y' - y_{true}\|_2 / \|y_{true}\|_2 = 0.193$; $\sqrt{\sigma^2}/\mu = 0.334$

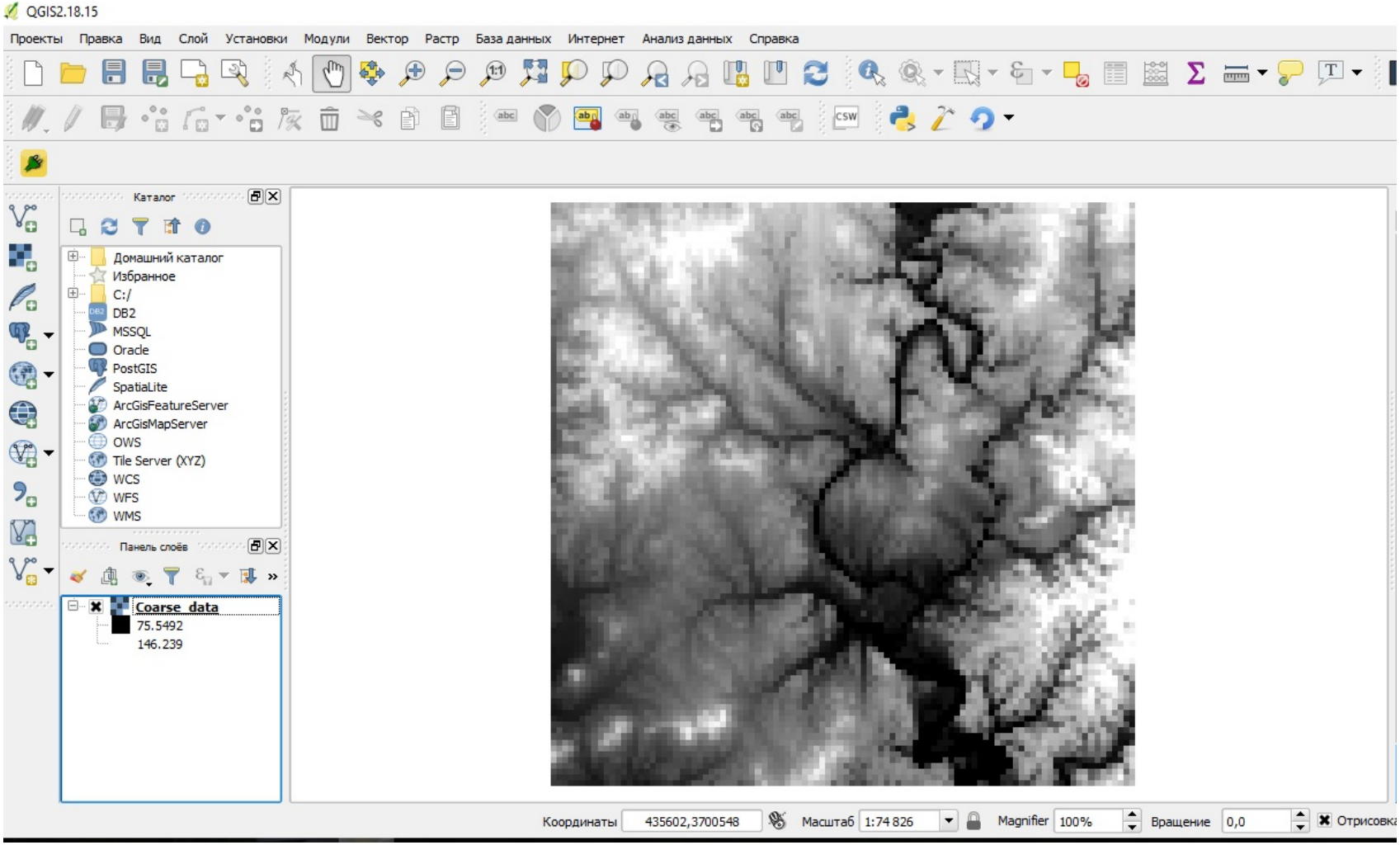




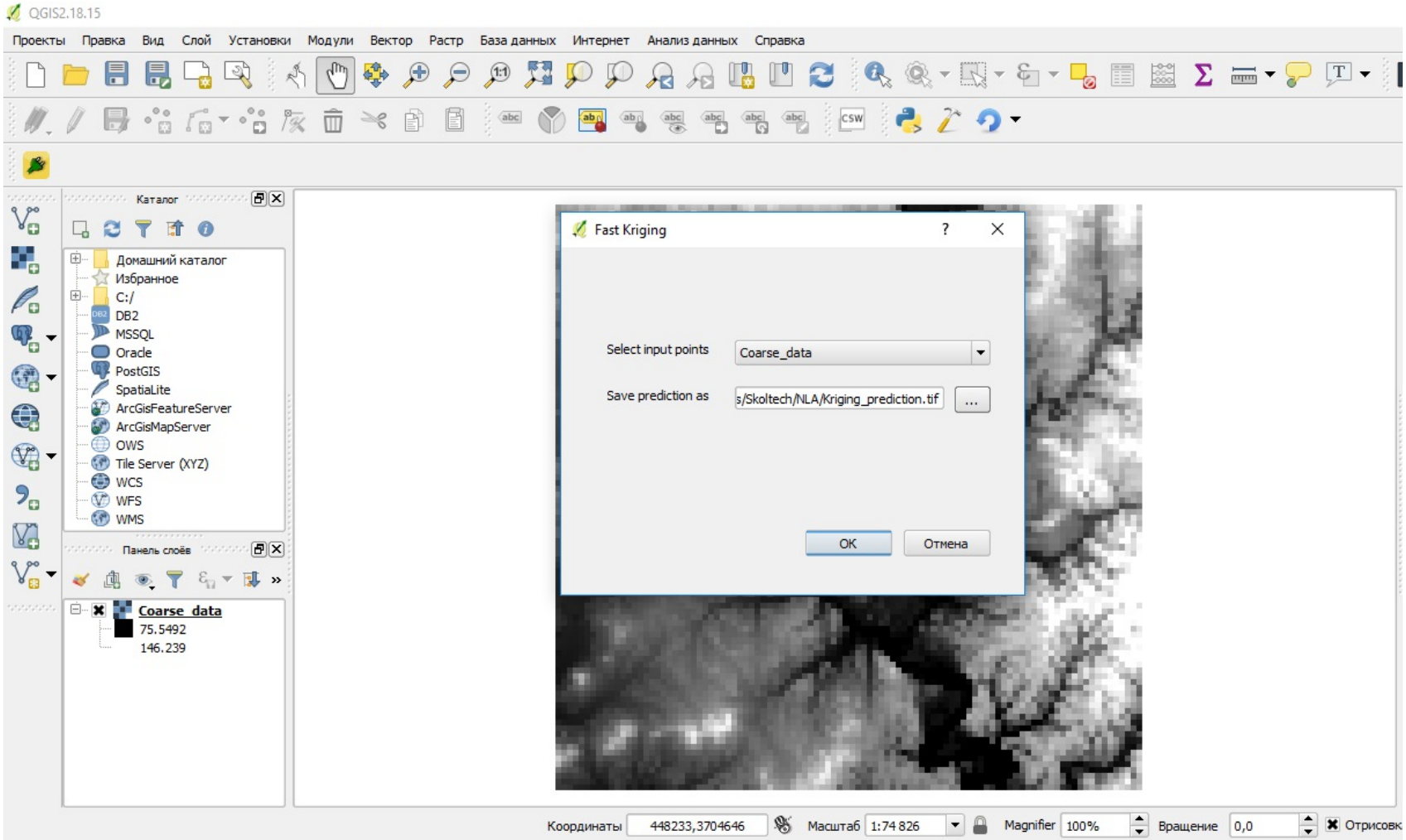
QGIS

QGIS is a free and open-source cross-platform desktop geographic information system application.
Its modular structure enables creation of plugins.
To our knowledge, existing kriging solutions don't take advantage of matrix structure

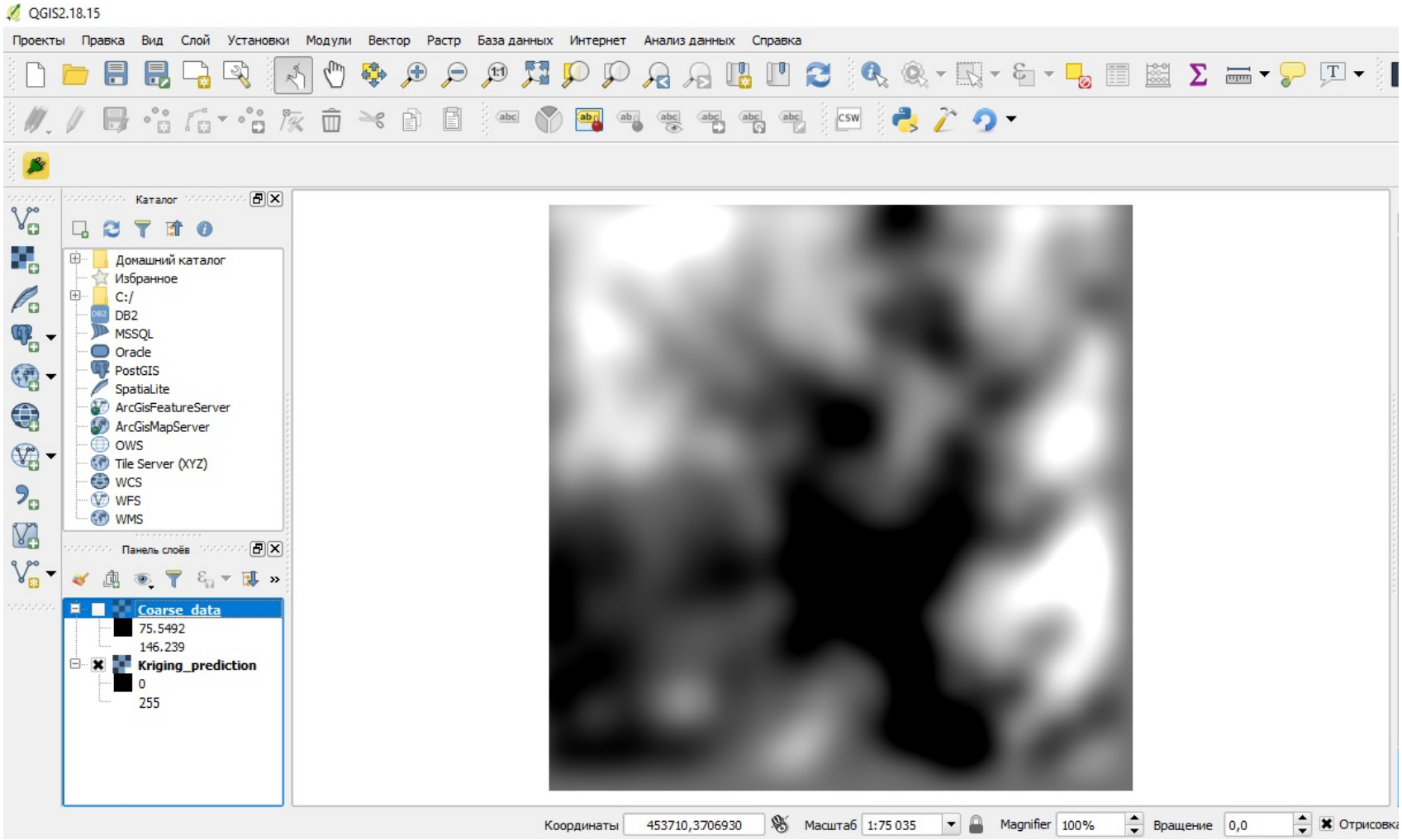
QGIS


QGIS2.18.15
Проекты Правка Вид Слой Установки Модули Вектор Растр База данных Интернет Анализ данных Справка
Каталог
Домашний каталог
Избранное
C:/
DB2
MSSQL
Oracle
PostGIS
SpatiaLite
ArcGisFeatureServer
ArcGisMapServer
OWS
Tile Server (XYZ)
WCS
WFS
WMS
Панель слоёв
Coarse_data
75.5492
146.239
Координаты 435602,3700548
Масштаб 1:74 826
Magnifier 100%
Вращение 0,0
Отрисовка

QGIS


QGIS2.18.15
Проекты Правка Вид Слой Установки Модули Вектор Растр База данных Интернет Анализ данных Справка
Каталог
Домашний каталог
Избранное
C:/
DB2
MSSQL
Oracle
PostGIS
SpatiaLite
ArcGisFeatureServer
ArcGisMapServer
OWS
Tile Server (XYZ)
WCS
WFS
WMS
Панель слоёв
Coarse_data
75.5492
146.239
Fast Kriging
Select input points
Coarse_data
Save prediction as
s/Skoltech/NLA/Kriging_prediction.tif
...
OK
Отмена
Координаты
448233,3704646
Масштаб
1:74 826
Magnifier
100%
Вращение
0,0
Отрисовк

QGIS


QGIS2.18.15
Проекты Правка Вид Слой Установки Модули Вектор Растр База данных Интернет Анализ данных Справка
Каталог
Домашний каталог
Избранное
C:/
DB2
MSSQL
Oracle
PostGIS
SpatiaLite
ArcGisFeatureServer
ArcGisMapServer
OWS
Tile Server (XYZ)
WCS
WFS
WMS
Панель слоёв
Coarse_data
75.5492
146.239
Kriging_prediction
0
255
Координаты 453710,3706930
Масштаб 1:75 035
Magnifier 100%
Вращение 0,0

Performance

- In theory, the speed must asymptotically be higher ($O(n \log n)$ per iteration versus $O(n^2)$)
- But for smaller n writing down the matrix explicitly and solving the system is faster (likely due to complicated structure of `LinearOperator`)
- However, for bigger n the naive implementation just runs out of memory for storage

Summary

- Gaussian regression aka kriging is a technique for interpolating large data sets while using fewer points
- When data is sampled evenly, the matrix for kriging is block Toeplitz with Toeplitz block, which enables faster matrix-vector product and memory-efficient storage
- We implemented the fast kriging algorithm in Python
- The code was converted into a plugin for QGIS

Thank you for your attention!