Fixed Rank Kriging
for
Continuous Gamma Radiation Data

November 15, 2016

Giorgio Paulon & Jennifer Starling

Here we apply the Fixed Rank Kriging method, as described by Katzfuss and Cressie in their 2011 paper "Tutorial on Fixed Rank Kriging of CO2 Data", to continuous spatial data consisting of gamma radiation measurements over the University of Texas at Austin campus.

1. Introduction

Kriging is a general technique for interpolating a response variable over a geospatial region, where the response has been observed at a limited number of locations within the region, and measurement error is present. Kriging allows us to predict the response on a fine mesh grid over the region. There are various methods of Kriging. The most basic technique, 'Ordinary Kriging', uses a weighted average of surrounding observations to predict the estimated response at the new location. A semi-variogram is used to optimize the weights.

Traditional kriging methods are computationally intractible in the big data setting, as they require inversion of the covariance matrix of the observed data set. With n observations, the efficiency of the inversion is $O(n^3)$. With data sets ranging from n in the tens of thousands to millions, covariance matrix inversion quickly becomes unscalable.

Katfuss and Cressie describe a variation on the traditional techniques called Fixed Rank Kriging (FRK). This technique was originally presented by Cressie and Johnanneson (2008). Fixed Rank Kriging relies on dimension reduction by way of basis functions; the response is modeled using a combinations of basis functions and fine-scale variance. This results in a covariance matrix which is the dimension of the number of basis functions used. In order for Fixed Rank Kriging to be computationally efficient compared to traditional kriging methods, the number of basis functions must be considerably smaller than the number of total observations in the data set. This is easily managed; Katzfuss and Cressie model CO2 readings over the entire globe using 396 basis functions of varying resolutions.

Here we apply the Fixed Rank Kriging data to a large data set of gamma-radiation readings from the University of Texas at Austin campus. The data is spatio-temporal in nature; here we focus on modeling only the spatial aspect of the data. We follow the analysis steps laid out in the Katzfuss and Cressie tutorial, including an exploratory analysis of the data.

Kriging is intended for predicting response values at new locations. In this work, we use Kriging to smooth the locations already observed, as well as predict responses over a grid of new locations.

2. Method

Method goes here.

Basis Function Generation

The choice of the basis functions is a very important step in the model specification for FRK. In fact, the matrix S allows us to represent the covariance structure as a linear combination of some basis function $S_1(\boldsymbol{u}), \dots, S_r(\boldsymbol{u})$, which results in a loss of information with respect to the full covariance representation.

The choice of the basis functions has to combine two goals. First of all, we should choose a number of basis functions $r \ll n$ in order to see an actual gain in terms of computational efficiency. Moreover, the basis functions have to be multiresolutional, that is, they should be allowed to capture multiple scales of variation in the covariance structure. In practice, there are a few smooth basis functions with large support (the limit case is the constant basis function, that is already implied in the centering step), and many spiky basis functions with small support.

The choice of the basis functions involves three separate problem: the choice of the type, the number r and the locations. The basis functions do not have to be necessarily orthogonal. In this work, we use bisquare functions, i.e. functions of the form

$$f(r) = \begin{cases} \left[1 - \left(\frac{r}{c}\right)^2\right]^2 & r \leq c \\ 0 & r > c \end{cases}$$

where c represents the resolution of the function and r is the euclidean distance of the coordinate from the center of the function. The number of basis functions r is chosen heuristically, in such a way that it can represent well the domain but that it does not compromise the performance of the algorithm. As far as the locations are concerned, they should cover as much as possible the spatial domain of interest (the prediction grid), and they should not overlap for different basis functions.

In this work we use a total of $r =$ basis functions with three different resolutions. In particular, $r_1 = 9$ functions have a low resolution $c_1 = 5 \cdot 10^{-3}$, $r_2 = 16$ functions have an intermediate resolution $c_2 = 2 \cdot 10^{-3}$ and $r_3 = 25$ functions have a high resolution $c_3 = 10^{-3}$.

Check and try again to use resolution = 1.5 times the shortest distance between the center of any of the functions with that resolution [see Cressie, pag 6].

Estimation of σ_ϵ^2 via Semivariogram

Stuff.

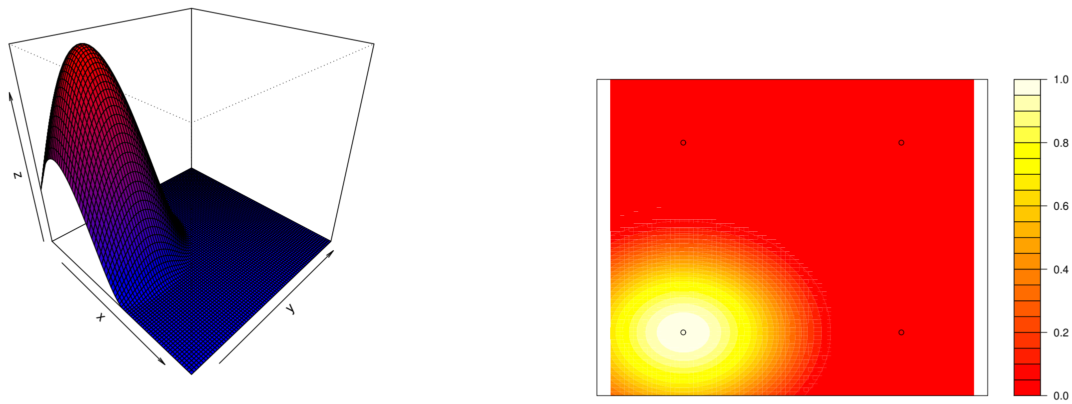

Figure 1:

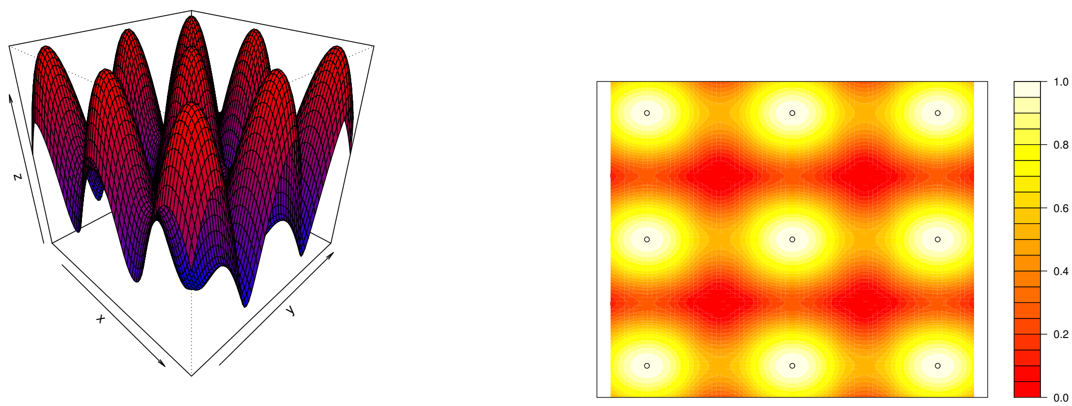

Figure 2:

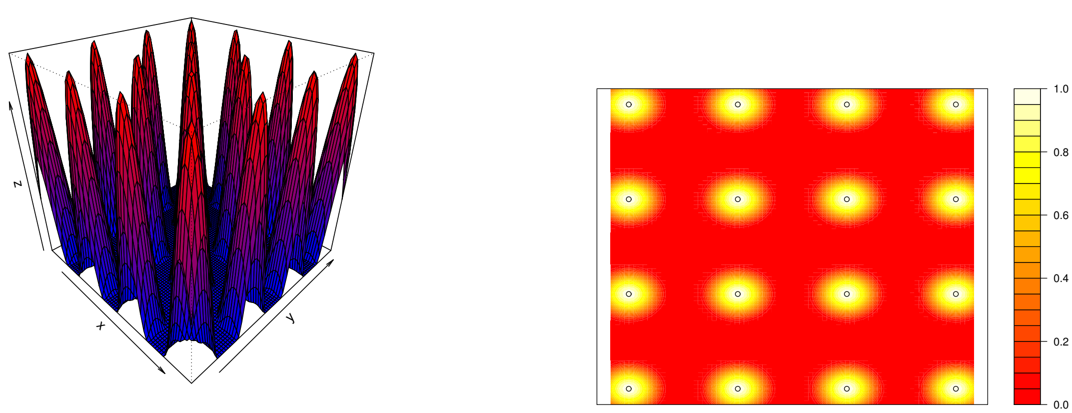

Figure 3:

Estimation of σ_ψ^2 and K via EM Algorithm

Stuff.

Fixed Rank Kriging: Smoothing and Prediction

Stuff.

3. Results

Results go here.

4. Discussion

Discussion goes here.

DO NOT FORGET TO CHECK NORMALITY OF RESIDUALS!!!!!!!

5. R Code Appendix

Documentation Source

All project documentation and source code is available in the following github repository.

https://github.com/jstarling1/spatialsmoothing

R Code: Main Launcher File

R Code: Main Launcher File