

**Fixed Rank Kriging
for
Continuous Gamma Radiation Data**

December 5, 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 Johnnneson (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

We follow the method described by Katzfuss and Cressie, outlined as follows.

- Data exploration and de-trending.
- Estimation of basis functions.
- Estimation of σ_ϵ^2 via semivariogram.
- Estimation of covariance matrix K and σ_ψ^2 via EM algorithm.
- Fixed-rank Kriging calculations.
- Estimation of FRK variance.

All components of these steps are described in detail below in **Method Description**.

Method Description

We briefly here describe the model underlying the Fixed Rank Kriging (FRK), highlighting the differences between the latter and the ordinary Kriging.

Let us suppose that we are interested in making inference on a hidden spatial process $\{Y(\mathbf{s}), \mathbf{s} \in D\}$, where D is a spatial domain identified by latitude and longitude, i.e. $D \in \mathbb{R}^2$. What is actually observed are the n noisy data measurements

$$Z(\mathbf{s}_i) = Y(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad i = 1, \dots, n, \quad (1)$$

where $\epsilon(\mathbf{s}_i)$ is the measurement error which is free to vary around the grid. The problem can be seen as a smoothing of the observed values $Z(\mathbf{s}_i)$ that takes into account the covariance structure of the data itself. The measurement error is usually estimated by the data. A general assumption is that $\epsilon(\cdot)$ is independent of $Y(\cdot)$ and distributed as

$$\epsilon \sim N(\mathbf{0}, \sigma_\epsilon^2 V_\epsilon).$$

Moreover, if one assumes that the measurement error variances are not different in different parts of the domain D , then $V_\epsilon = \mathcal{I}_n$, where \mathcal{I}_n denotes the identity matrix of dimension $n \times n$. The latter assumption will be implicit in the following of this work. The global scaling variance σ_ϵ^2 is estimated via the semivariogram of the data near the origin, as we will see later.

The hidden process $Y(\cdot)$ is supposed to be decomposed in a linear term $\mu(\cdot)$ and in a random spatial variation $\nu(\cdot)$, that is,

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \nu(\mathbf{s}_i) = \mathbf{t}(\mathbf{s}_i)^T \boldsymbol{\alpha} + \nu(\mathbf{s}_i), \quad i = 1, \dots, n. \quad (2)$$

The linear term can be thought as a linear regression, where $\mathbf{t}(\cdot)$ is a vector of known space-dependent covariates. It is fundamental to detrend the data prior to apply the Kriging techniques,

since we want to recover the pure covariance structure given by the spatial dimension, and not by any other variable. In absence of significant covariates, the process has at least to be centered (detrended by the grand mean). The spatial variation term is the key goal of spatial statistics, that is to infer on the underlying covariance structure of the data. We define the spatial covariance function

$$C(\mathbf{u}, \mathbf{v}) = \text{Cov}(\nu(\mathbf{u}), \nu(\mathbf{v})), \quad \mathbf{u}, \mathbf{v} \in D.$$

One can also rewrite the model by grouping the two hierarchical levels (1) - (2) as

$$\mathbf{Z} = T\boldsymbol{\alpha} + \boldsymbol{\delta},$$

where

$$\Sigma = \text{Cov}(\boldsymbol{\delta}) = C + \sigma_e^2 \mathcal{I}_n = \begin{cases} C(\mathbf{s}_i, \mathbf{s}_i) + \sigma_e^2 & i = j \\ C(\mathbf{s}_i, \mathbf{s}_j) & i \neq j. \end{cases}$$

Ordinary Kriging is a smoothing technique which is based on the following formula

$$\hat{Y}(\mathbf{s}^*) = \mathbf{t}(\mathbf{s}^*)^T \hat{\boldsymbol{\alpha}} + \mathbf{k}(\mathbf{s}^*)^T (\mathbf{Z} - T\hat{\boldsymbol{\alpha}}), \quad \mathbf{s}^* \in \mathbb{R}^2 \quad (3)$$

where

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (T^T \Sigma^{-1} T)^{-1} T^T \Sigma^{-1} \mathbf{Z} \\ \mathbf{k}(\mathbf{s}^*)^T &= \mathbf{c}(\mathbf{s}^*)^T \Sigma^{-1} \\ \mathbf{c}(\mathbf{s}^*) &= (C(\mathbf{s}^*, \mathbf{s}_1), \dots, C(\mathbf{s}^*, \mathbf{s}_n))^T. \end{aligned}$$

Therefore, the computation of ordinary Kriging requires the inversion of the $n \times n$ covariance matrix, which is computationally unfeasible in a big data framework, where n can be up to one million. FRK, instead, tries to seek an approximation of this matrix on a low-dimensional subspace. We specify, in particular, a form of the C matrix that makes the inversion of Σ easier. We hypothesize a Spatial Random Effects model for the random spatial variation term $\nu(\cdot)$, that is

$$\nu(\cdot) = \mathbf{S}(\cdot)^T \boldsymbol{\eta} + \xi(\cdot), \quad (4)$$

where

$$\begin{aligned} \boldsymbol{\eta} &\sim N_r(\mathbf{0}, K) \\ \xi(\cdot) &\sim N(0, \sigma_\xi^2 V_\xi(\cdot)), \end{aligned} \quad (5)$$

where K is a $r \times r$ matrix. In this latter expression $\xi(\cdot)$ represents the fine scale variation process that accounts for the error introduced by the dimension reduction. In the following we assume that $V_\xi(\cdot)$ is constant, whereas K and σ_ξ^2 are estimated by the data via the EM algorithm presented in the following section. It is easy to see that a consequence of (4) - (5) is that the matrix of the covariance structure C can be written as

$$C(\mathbf{u}, \mathbf{v}) = \mathbf{S}(\mathbf{u}) K \mathbf{S}(\mathbf{v})^T + \sigma_\xi^2, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^2$$

where $\mathbf{S}(\mathbf{u})$ represents the basis function approximation in the point \mathbf{u} . Thus, the covariance structure of the observed data is $\Sigma = SKS^T + D = SKS^T + (\sigma_\epsilon^2 + \sigma_\xi^2)\mathcal{I}_n$, where S is the $n \times r$ matrix of the basis functions evaluated at each data point.

Let us define $\sigma^2 = \sigma_\epsilon^2 + \sigma_\xi^2$. The inverse of the covariance matrix can now be efficiently computed using the Sherman-Morrison-Woodbury formula

$$\Sigma^{-1} = \frac{1}{\sigma^2} - \frac{1}{\sigma^4} S \left(K^{-1} + \frac{1}{\sigma^2} S^T S \right)^{-1} S^T,$$

which requires only the inversion of the $r \times r$ matrix K . The Kriging equation becomes then

$$\hat{Y}(\mathbf{s}^*) = \mathbf{t}(\mathbf{s}^*)^T \hat{\alpha} + \mathbf{S}(\mathbf{s}^*)^T KS^T \Sigma^{-1} (\mathbf{Z} - T\hat{\alpha}), \quad \mathbf{s}^* \in \mathbb{R}^2 \quad (6)$$

and the standard error is

$$\begin{aligned} \sigma_k(\mathbf{s}^*) = & [\mathbf{S}(\mathbf{s}^*)^T KS(\mathbf{s}^*) - \mathbf{S}(\mathbf{s}^*)^T KS^T \Sigma^{-1} SKS(\mathbf{s}^*) + \\ & + (\mathbf{t}(\mathbf{s}^*) - T^T \Sigma^{-1} SKS(\mathbf{s}^*))^T (T^T \Sigma^{-1} T)^{-1} (\mathbf{t}(\mathbf{s}^*) - T^T \Sigma^{-1} SKS(\mathbf{s}^*))]^{1/2}. \end{aligned} \quad (7)$$

This approach reduces the computational burden of the algorithm from $\mathcal{O}(n^3)$ to $\mathcal{O}(r^2 n)$.

Data Exploration and De-trending

We include several preparatory steps before formally beginning the Fixed Rank Kriging process.

De-Trending

The first step is to select a subset of the original data, limited to the latitudes and longitudes containing the University of Texas campus. The data set contains several outliers, including measurements which appear to be from San Antonio. These outliers are excluded.

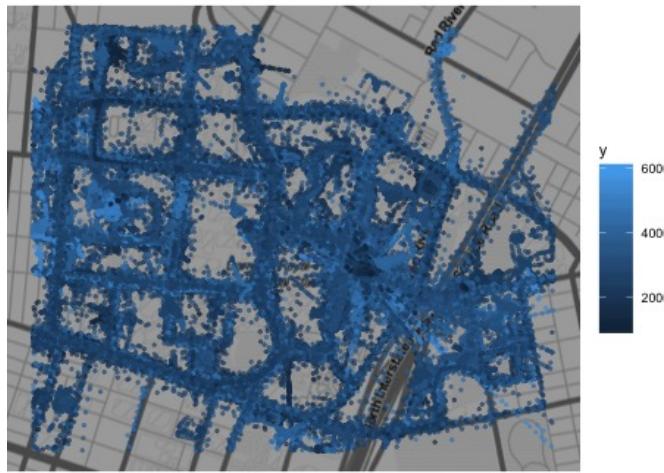


Figure 1: Observed data

Second, Katzfuss and Cressie suggest de-trending to adjust for any large-scale spatial variation and any key covariates. We plotted the raw data versus longitude, latitude, and temperature. The data do not show any trend by latitude or longitude, ruling out deterministic trends. There is a trend by temperature, which is not surprising as lower temperatures may affect the accuracy of the radiation measurements.

We chose not to de-trend the data by temperature, as this introduced strange patterns in the data when plotted by latitude and longitude.

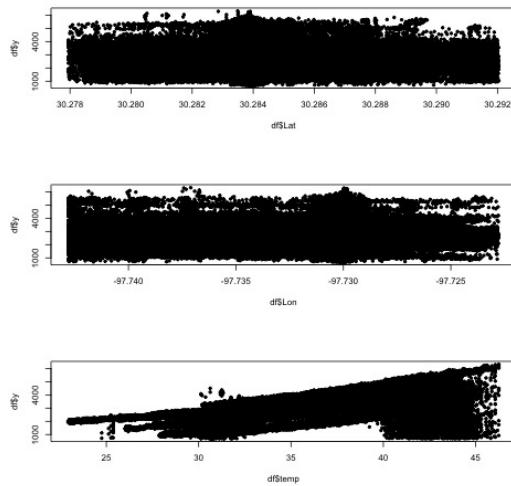


Figure 2: Data by Latitude, Longitude and Temperature

Transformation to Normality

Third, Katzfuss and Cressie recommend that the data be checked for approximate normality and transformed if necessary. The radiation data is in the form of counts (intensity measurements) and is likely Poisson, so we use an Anscombe transformation to induce normality. A histogram of the normalized observations ($y.norm$) confirm this transformation was successful. Note that Anscombe back-transformations can lead to a biased result, and so we use the following transform and back-transform.

$$x \rightarrow 2\sqrt{(x + \frac{3}{8})}$$
$$y \rightarrow \left(\frac{y}{2}\right)^2 - \frac{1}{8} \quad (\text{instead of the direct algebraic inverse, where } \frac{3}{8} \text{ is subtracted.})$$

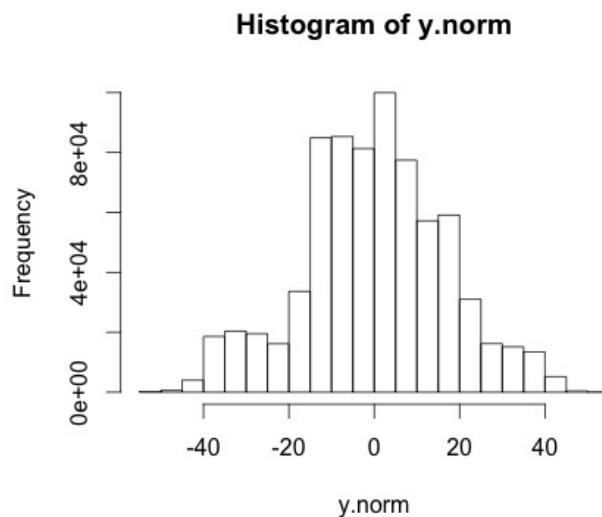


Figure 3: Anscombe-Transformed Responses

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 $S_1(\mathbf{u}), \dots, S_r(\mathbf{u})$, 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 $r = 50$ 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].

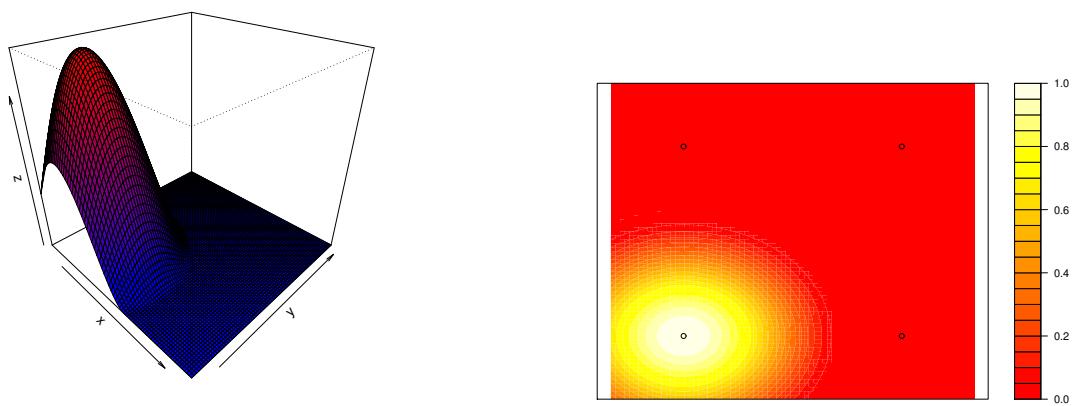


Figure 4:

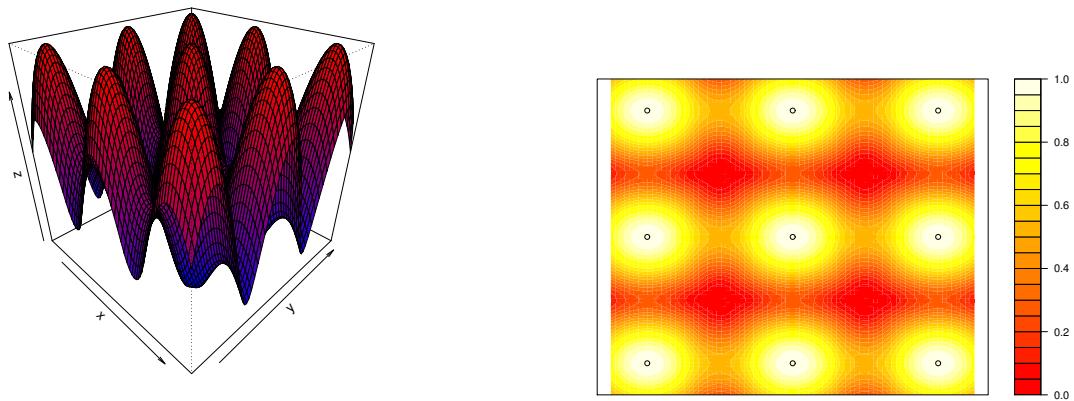


Figure 5:

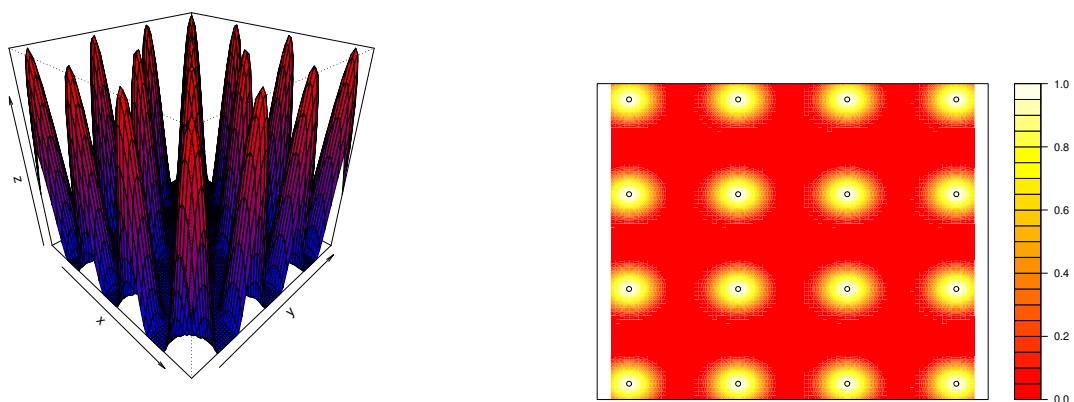


Figure 6:

Estimation of σ_ϵ^2 via Semivariogram

The FRK model specifies that n measurements are modeled as $Z(s_i) = Y(s_i) + \epsilon(s_i)$, where s_i are locations in the region of interest, and $\epsilon(s_i)$ is additive measurement error.

As detailed above, the measurement error is assumed to be independent of $Y(s_i)$ and normally distributed as $\epsilon(s_i) \sim N(0, \sigma_\epsilon^2 \nu_\epsilon(s_i))$. The $\nu_\epsilon(s_i)$ component is a function that is assumed to be known, which describes the random spatial variation. As with the CO₂ data, we assume $\nu_\epsilon(s_i)$ is a constant function equal to 1 at every location, as we assume that random spatial variation is constant over the campus.

This leaves us with σ_ϵ^2 to estimate.

Katzfuss and Cressie note that σ_ϵ^2 may be specified in advance, if there is previous knowledge about measurement error for the instrument being used. We do not assume prior knowledge of σ_ϵ^2 .

The semivariogram estimation method presented in Katzfuss and Cressie is known as the 'Robust Cressie Variogram.'

$$\bullet 2\gamma(h) = \left(\frac{1}{|N(h)|} \sum_{N(h)} \left| \frac{\tilde{Z}(s_i)}{\sqrt{\nu_\epsilon(s_i)}} - \frac{\tilde{Z}(s_j)}{\sqrt{\nu_\epsilon(s_j)}} \right|^5 \right)^4 / (0.457 + 0.494/|N(h)|)$$

where:

$\tilde{Z}(s_i)$ are the normalized, de-trended data.

$N(h) = (i, j) : \|s_i - s_j\| \in [h - \delta, h + \delta]$,

the set of all indices for location pairs roughly h distance apart.

- Then straight line $\hat{\gamma}(h) = \hat{\gamma}(0) + bh$ is fit, and $\sigma_\epsilon^2 = \hat{\gamma}(0)$.

(For further detail on this methodology, see Cressie and Hawkins (1980), Cressie (1985), and Kang et al. (2010)).

We use this method to estimate a semi-variogram for the data, and then fit a straight line to the semivariogram and let $\hat{\sigma}_\epsilon^2$ be equal to the intercept. Our resulting estimate is $\hat{\sigma}_\epsilon^2 = 106.3635$.

We did remove the police station measurements in the calculation of this semivariogram. The police car carrying the measurement device parked in the police station lot overnight, so there were a disproportionate number of measurements at this location, including specifically nighttime measurements where cooler temperatures mean the measurement device is more prone to error.

This was considerably higher than the estimate obtained for the Katzfuss and Cressie simulated

CO₂ data. However, their data was simulated with $\sigma^2_\epsilon = 0.25$, and our normalized data has a range of -51 to 54, with a total variance of 287, so our result is reasonable. To check reasonableness, we split the region of interest into four quadrants and calculated semivariograms for each quadrant to ensure that they yielded comparable $\hat{\sigma}^2_\epsilon$ estimates. Estimates were similar in all regions, and did not differ significantly from the overall estimate. The data is known to have some outliers which affect the variance.

Estimation of σ_ξ^2 and K via EM Algorithm

The next step of the Fixed Rank Kriging process is to estimate the parameters σ_ξ^2 and K .

As described above, K is the ($r \times r$) covariance matrix, where r is the number of basis functions selected. K represents the covariance in the spatial-random-effects (SRE) model.

σ_ξ^2 is the fine-scale variance in the SRE model, accounting for error introduced to the model through the dimension reduction.

Katzfuss and Cressie suggest two methods for estimating these parameters: Method of Moments and the EM Algorithm. We have implemented the EM algorithm to estimate these values.

- Begin with default choices suggested in Katzfuss and Cressie:

$$\begin{aligned} K^{[0]} &= (.9)\nu^2 I_r \\ \sigma_\xi^{2[0]} &= (.1)\nu^2 I_r \\ \text{where } \nu^2 &= \tilde{Z}'\tilde{Z}/n. \end{aligned}$$

- Iteratively update the two parameters until convergence is met:

$$\begin{aligned} K^{[t+1]} &= K^{[t]} - K^{[t]}S'\Sigma^{[t]-1}SK^{[t]} + \left(K^{[t]}S'\Sigma^{[t]-1}\tilde{Z}\right)\left(K^{[t]}S'\Sigma^{[t]-1}\tilde{Z}\right)' \\ \sigma_\xi^{2[t+1]} &= \sigma_\xi^{2[t]} + \left(\sigma_\xi^{2[t]}\right)^2 \text{tr}\left(\Sigma^{[t]-1}[\tilde{Z}\tilde{Z}'\Sigma^{[t]-1} - I_n]V_\xi\right) \end{aligned}$$

Computational notes:

- $\Sigma^{[t]-1}$ is shorthand for $(\Sigma^{[t]})^{-1}$
- As discussed previously, we used the Sherman-Morrison-Woodbury formula to calculate $\Sigma^{[t]-1}$, to efficiently compute the inverse.
- So we let $D^{[t]} = \sigma^2 V_\xi + \sigma_\epsilon^2 V_\epsilon$.
- Then $\Sigma^{[t]-1} = D^{[t]-1} - D^{[t]-1}S[K^{[t]-1} + SD^{[t]-1}S']^{-1}D^{[t]-1}$

We can use these results to re-write the $\sigma_\xi^{2[t+1]}$ update in a more efficient way:

$$\begin{aligned}\sigma_\xi^{2[t+1]} &= \sigma_\xi^{2[t]} + \left(\sigma_\xi^{2[t]}\right)^2 \text{tr}([S'D^{[t]-1}V_\xi D^{[t]-1}S][K^{[t]-1} + S'D^{[t]-1}S]^{-1})/n \\ &\quad - \left(\sigma_\xi^{2[t]}\right)^2 \text{tr}(D^{[t]-1}V_\xi)/n + \left(\sigma_\xi^{2[t]}V_\xi\Sigma^{[t]-1}\tilde{Z}\right)' V_\xi^{-1} \left(\sigma_\xi^{2[t]}V_\xi\Sigma^{[t]-1}\tilde{Z}\right)/n\end{aligned}$$

Results:

We obtain an estimated $\hat{\sigma}_\xi^2 = 174.4245$. The R code saves $\hat{\sigma}_\xi^2$ and \hat{K} as part of the EM function output.

Fixed Rank Kriging: Smoothing and Prediction

We implemented the Fixed Rank Kriging process using the equations described previously.

$$\hat{Y}(\mathbf{s}^*) = \mathbf{t}(\mathbf{s}^*)^T \hat{\alpha} + \mathbf{S}(\mathbf{s}^*)^T K S^T \Sigma^{-1} (\mathbf{Z} - T \hat{\alpha}), \quad \mathbf{s}^* \in \mathbb{R}^2 \quad (8)$$

and the standard error is

$$\begin{aligned}\sigma_k(\mathbf{s}^*) &= [\mathbf{S}(\mathbf{s}^*)^T K S(\mathbf{s}^*) - \mathbf{S}(\mathbf{s}^*)^T K S^T \Sigma^{-1} K S(\mathbf{s}^*) + \\ &\quad + (\mathbf{t}(\mathbf{s}^*) - T^T \Sigma^{-1} K S(\mathbf{s}^*))^T (T^T \Sigma^{-1} T)^{-1} (\mathbf{t}(\mathbf{s}^*) - T^T \Sigma^{-1} K S(\mathbf{s}^*))]^{1/2} \quad (9)\end{aligned}$$

where $\Sigma = K S^T + D = K S^T + (\sigma_\epsilon^2 + \sigma_\xi^2) I_n$.

The Kriging function returns a grid of locations and estimated responses, along with a vector of kriging variances. Results are included in subsequent sections.

3. Results

Using a uniform set of basis functions, predicted kriging values are shown in the heatmap below, followed by a plot of the kriging variance.

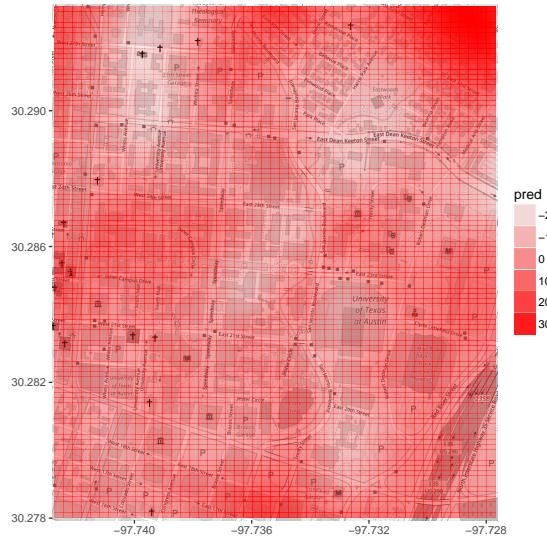


Figure 7: Predicted values, uniform basis

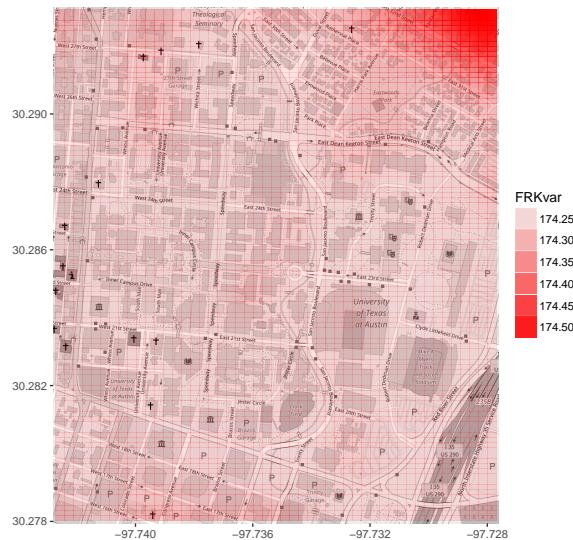


Figure 8: FRK variance, uniform basis

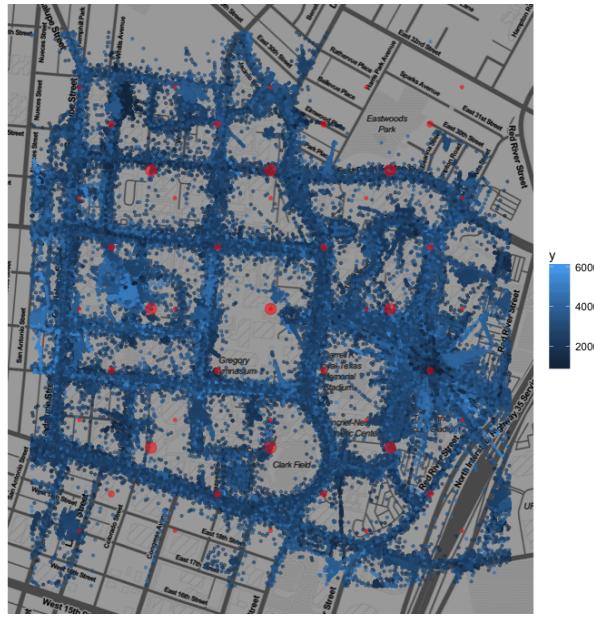


Figure 9: Observed data and basis function centers

We noticed an area in the top right of very high variance. This issue is caused by very sparse observed data coverage in the top right area of the grid; there is one blue point by itself which is near East 32nd Street. There are no other close points informing the kriging estimate in this area of the grid, and compounding the issue, there are two spiky basis centers quite close to this point.

To address this issue, we adjusted the basis centers by removing the two spiky centers in this corner. We justified this by noting that kriging is not intended to extrapolate into new regions without surrounding estimates to inform the prediction. Of course a lone point is going to have a high kriging variance, when none of its neighbors are near. The result of this adjustment is below.

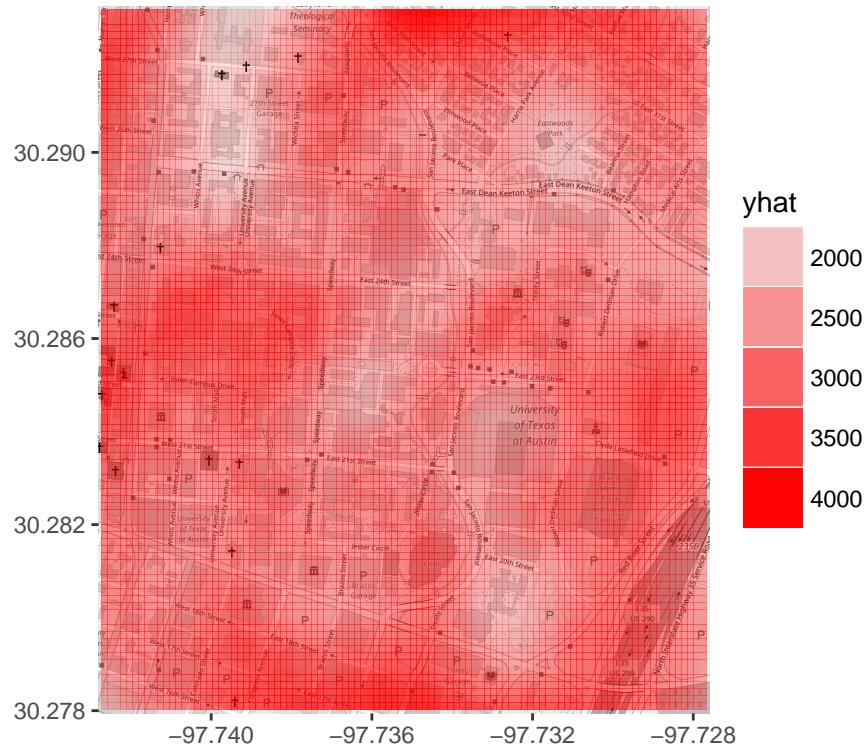


Figure 10: Predicted values, adjusted basis

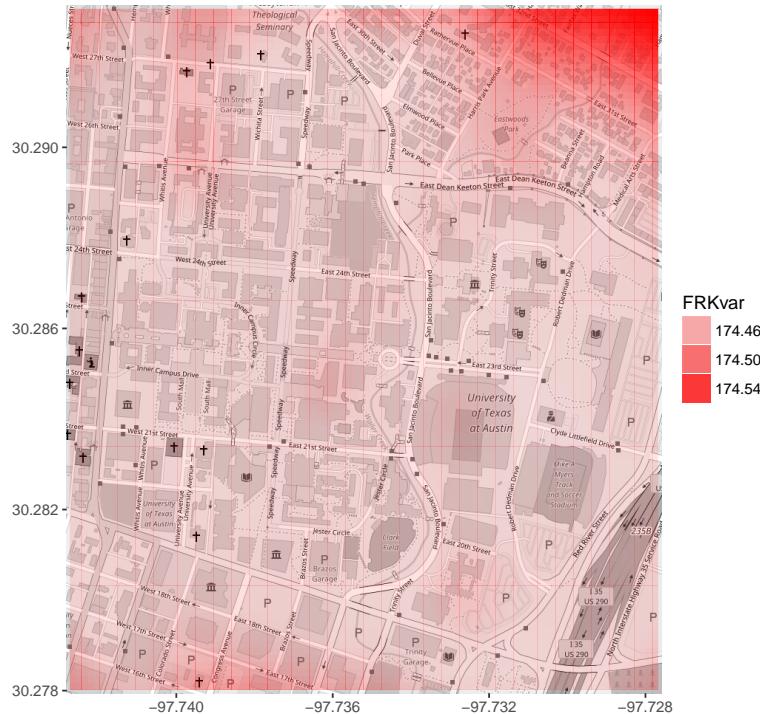


Figure 11: FRK variance, adjusted basis

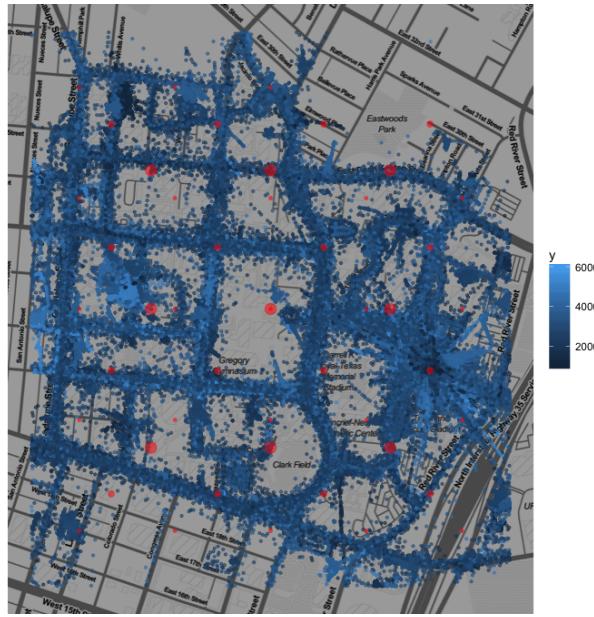


Figure 12: Observed data and adjusted basis function centers

The result of this adjustment yields desirable results for the predicted data. The kriging variance is nearly uniform across the predicted grid of locations, with the area of high variance in the top right region reduced in size.

Finally, we include a histogram of the observed versus predicted residual for the kriging function applied to the observed locations. These values should be normally distributed to indicate goodness of model fit.

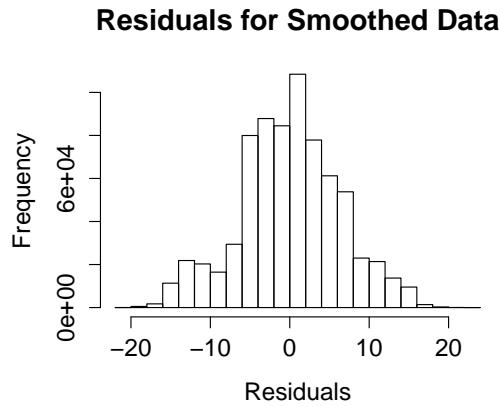


Figure 13: Residuals

4. Discussion

Our Fixed Rank Kriging results looked to be reasonable based on comparison with the underlying observed data, our knowledge of potential higher-radiation areas on campus, and based on the normality of the histogram of the residuals. Our code was computationally efficient, executing the entire process including semivariogram estimation, basis function generation, EM algorithm calculations, and Fixed Rank Kriging computations in 18 minutes.

To increase efficiency over a large data set, our implementation uses pre-caching and sparse matrix functionality to the extent possible. D^{-1} is stored as a sparse matrix since it is diagonal. Computations involving matrix multiplications that always appear together in updates are pre-cached, such as $K^{[t]}S'\Sigma^{[t]-1}$.

One limitation in our code is the calculation of the Fixed Rank Kriging variance in the case where the code is run for smoothing the observed values, instead of predicting over a grid of new values. We omitted the computation of the Fixed Rank Kriging variance in the case where the function is smoothing, ie 100% of the locations to be predicted correspond with already-observed locations. The variance computation becomes very computationally intensive in this case, and a future area for expansion is to implement the FRK variance calculation in C++.

5. R Code Appendix

Documentation Source

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

<https://github.com/jstarling1/spatialsMOOTHING>

The R code launcher is available in the *R Code_Launcher* directory.