A Study of Fixed Rank Kriging for Large Spatial Data Sets

Xiao Tan and Zhenbang Wang

Abstract

In 'big data' era, making optimal spatial prediction(kriging) can be challenging. Motivated by the fact that traditional method involves inverting $n \times n$ covariance matrix which is computational expensive and problematic when n is very large, we study a flexible family of non-stationary covariance functions is defined by using a set of basis functions that is fixed in number $r \ll n$, which leads to significant time reduction spatial prediction method that we call fixed rank kriging (**FRK**) [Cressie and Johannesson, 2008]. We applied **FRK** to make prediction on a synthetic CO2 data and meuse data, which includes hundreds of thousands observations.

1 Introduction

In geostatistics, kriging which gives the best linear unbiased prediction under appropriate assumptions, has been systematically studied and successfully applied to achieve data interpolation, especially in earth and environmental sciences. Kriging captures the information in spatial data through the spatial variability (i.e. covariance function) to obtain spatial prediction. Associated spatial prediction map is available and it is extremely informative if we wish to study the general trend of some spatial data. In [Cressie and Johannesson, 2008], the **FRK** methodology was proposed and well studied to speed up the computation of original kriging by modeling covariance function through basis function.

Comparing to past, spatial datum are more ubiquitous and accessible in this "Big Data" age. However, the benefits of this ubiquity also come with the challenge for original kriging methodology. The computation burden of straightforward kriging is extremely unacceptable for massive dataset such as data from satellites when it comes to obtain meaningful results in reasonable time. Obtaining spatial prediction when equations in original kriging requires solving the inverse of $n \times n$ variance-covariance matrix $\mathbf{\Sigma}$. The computational cost of getting $\mathbf{\Sigma}^{-1}$ in straightforward kriging is $\boldsymbol{O(n^3)}$. While in proposed **FRK** method, the computational cost of obtaining $\mathbf{\Sigma}^{-1}$ reduced to $\boldsymbol{O(nr^2)}$, which is linear on the sample size(n) of your data.

In section 2, we presents the kriging methodology and gives the equations that define **FRK**. In section 3, how to estimate covariance matrix $\boldsymbol{K}$ and the choice of basis function were discussed. The application of **FRK** methodology to a global CO_2 dataset was achieved in section 4.

2 Kriging: optimal linear spatial prediction

2.1 The kriging overview

Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{D} \in \mathcal{R}^d\}$ be real-valued spatial process. Based on the observed spatial process $\boldsymbol{Z}(\cdot)$ that contain measurement error, we are interested in making inference on Y-process.

$$Z(\boldsymbol{s}) = Y(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) \quad \boldsymbol{s} \in D \tag{2.1}$$

where $\epsilon(\boldsymbol{s})$ is for measurement error and $\{\epsilon(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{D}\}$ is spatial white noise process with mean 0 and variance $\text{var}\{\epsilon(\boldsymbol{s})\} = \sigma^2 \upsilon(\boldsymbol{s}) \in (0, \infty)$, $\boldsymbol{s} \in \boldsymbol{D}$, for $\sigma^2 > 0$ and $\upsilon(\cdot)$ is known.
Also, we only know $\boldsymbol{Z}(\cdot)$ at finite number of spatial locations $\{\boldsymbol{s}_1, \boldsymbol{s}_2, ...\boldsymbol{s}_n\}$, then the observed data vector is

$$\boldsymbol{Z} = (Z(\boldsymbol{s}_1), ..., Z(\boldsymbol{s}_n)) \tag{2.2}$$

We are interested in making inference on hidden Y process. Linear mean structure is assumed for Y process, i.e.

$$Y(\boldsymbol{s}) = \boldsymbol{t(s)}'\boldsymbol{\alpha} + \upsilon(\boldsymbol{s}) \quad \boldsymbol{s} \in D \tag{2.3}$$

where $\boldsymbol{t(\cdot)} = (t_1(\cdot), ...t_p(\cdot))'$ are a process of known covariates, $\boldsymbol{\alpha} = (\alpha_1, ...\alpha_p)'$ are the unknown coefficients, $\upsilon(\boldsymbol{s}) : \text{E}\{\upsilon(\boldsymbol{s})\} = 0,\ 0 < \text{Var}\{\upsilon(\boldsymbol{s})\} < \infty, \boldsymbol{s} \in D$.

Generally non-stationary spatial covariance function,

$$\text{cov}\{\upsilon(\boldsymbol{u}), \upsilon(\boldsymbol{v})\} = C(\boldsymbol{u}, \boldsymbol{v}) \qquad \boldsymbol{u}, \boldsymbol{v} \in D \tag{2.4}$$

(2.1)-(2.4) above imply a general linear mixed model,

$$\boldsymbol{Z} = \boldsymbol{T\alpha} + \boldsymbol{\epsilon} \qquad \boldsymbol{\delta} = \boldsymbol{V} + \boldsymbol{\epsilon} \tag{2.5}$$

where $\boldsymbol{T}$ is n×p matrix of covariates $(\boldsymbol{t(s_1)}, ...\boldsymbol{t(s_n)})'$, $\boldsymbol{\alpha} = (\alpha_1, ...\alpha_p)'$ are the unknown coefficients $\boldsymbol{\delta}$ addition of two independent, zero-mean components, then we have $\text{E}(\boldsymbol{\delta}) = \boldsymbol{0}, \text{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma} = (\sigma_{ij})$

$$\sigma_{ij} = \begin{cases} C(\boldsymbol{s}_i, \boldsymbol{s}_j) + \sigma^2 \upsilon(\boldsymbol{s}_j) & i = j \\ C(\boldsymbol{s}_i, \boldsymbol{s}_j) & i \neq j \end{cases}$$

Put it in the matrix form, we have

$$\boldsymbol{\Sigma} = \boldsymbol{C} + \sigma^2 \boldsymbol{V} \tag{2.6}$$

where $\boldsymbol{C} = (C(\boldsymbol{s}_i, \boldsymbol{s}_j))$ and $\boldsymbol{V} = \boldsymbol{diag}\{\upsilon(\boldsymbol{s}_1)...\upsilon(\boldsymbol{s}_n)\}$. Note that no assumptions such as stationarity and isotropy have been made on convariance function.

We observed Z process at finite number of locations$(s_1, ...s_n)$. However, we are interested in making inference on Y-process, i.e. making point prediction on unobserved location $s_0, s_0 \in D$. [Cressie, 1992], gave the formula of kriging predictor $Y(s_0)$

$$\hat{Y}(\boldsymbol{s}_0) = \boldsymbol{t}(\boldsymbol{s}_0)'\hat{\boldsymbol{\alpha}} + \boldsymbol{k}(\boldsymbol{s}_0)'(\boldsymbol{Z} - \boldsymbol{T}\hat{\boldsymbol{\alpha}}) \tag{2.7}$$

where

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} \tag{2.8}$$

$$k(\boldsymbol{s}_0)' = \boldsymbol{c}(\boldsymbol{s}_0)'\boldsymbol{\Sigma}^{-1} \tag{2.9}$$

where $\boldsymbol{c}(\boldsymbol{s}_0) = (C(\boldsymbol{s}_0, \boldsymbol{s}_1), ..., C(\boldsymbol{s}_0, \boldsymbol{s}_n))'$

The kriging standard error is the root-mean-squared prediction error of $\hat{Y}(\boldsymbol{s}_0)$, $\sigma_k(\boldsymbol{s}_0) = \left[E\{Y(\boldsymbol{s}_0) - \hat{Y}(\boldsymbol{s}_0)\}^2\right]^{\frac{1}{2}}$

$$\sigma_k(\boldsymbol{s}_0) = \{\boldsymbol{C}(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{k}(\boldsymbol{s}_0)'\boldsymbol{\Sigma}\boldsymbol{k}(\boldsymbol{s}_0) + (\boldsymbol{t}(\boldsymbol{s}_0) - \boldsymbol{T}'\boldsymbol{k}(\boldsymbol{s}_0))'(\boldsymbol{T}'\Sigma^{-1}\boldsymbol{T})^{-1}(\boldsymbol{t}(\boldsymbol{s}_0 - \boldsymbol{T}'\boldsymbol{k}(\boldsymbol{s}_0))\}^{\frac{1}{2}} \tag{2.10}$$

Through equations from (2.7)-(2.10), we could make kriging prediction map and kriging standard error map based on $\hat{Y}(\boldsymbol{s}_0)$ and $\sigma_k(\boldsymbol{s}_0)$, $s_0 \in D$. However, they involve solving the inverse of Σ, which has computational cost of $O(n^3)$. What we discussed above is feasible when the number of observations(n) is small. When n is large, the computation will not be handled in a reasonable time. In the following subsection, we will propose a new method that can significantly reduce computational cost by on the choice of covariance function.

2.2 Spatial covariance function

From (2.3), we have

$$Y(\boldsymbol{s}) = \boldsymbol{t}(\boldsymbol{s})'\boldsymbol{\alpha} + v(\boldsymbol{s}) \quad \boldsymbol{s} \in D$$

By setting $v(\boldsymbol{s}) = \boldsymbol{S}(\boldsymbol{s})'\boldsymbol{\eta}$, we have

$$Y(\boldsymbol{s}) = \boldsymbol{t}(\boldsymbol{s})'\boldsymbol{\alpha} + \boldsymbol{S}(\boldsymbol{s})'\boldsymbol{\eta}$$

which is called a spatial mixed effects linear model. Intuitive idea behind this is to model the covariance function $\boldsymbol{C}(\boldsymbol{u}, \boldsymbol{v})$ through a set of r(not necessarily orthogonal) basis functions.

$$\boldsymbol{S}(\boldsymbol{u}) = (S_1(\boldsymbol{u}), ...S_r(\boldsymbol{u}))' \qquad \boldsymbol{u} \in \mathcal{R}^d \tag{2.11}$$

where r is fixed. [Cressie and Johannesson, 2006] discussed that cov $\{Y(\boldsymbol{u}), Y(\boldsymbol{v})\}$ could be modeled as

$$\boldsymbol{C}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{S}(\boldsymbol{u})'\boldsymbol{K}\boldsymbol{S}(\boldsymbol{v}) \qquad \boldsymbol{u}, \boldsymbol{v} \in \mathcal{R}^d \tag{2.12}$$

where $\boldsymbol{K}$ is any $r \times r$ positive definite matrix. Also note that $\boldsymbol{\eta}$ is an r-dimensional vector with var$(\boldsymbol{\eta}) = \boldsymbol{K}$

2.3 Fixed rank kriging

Recall in (2.6), we have $\boldsymbol{\Sigma} = \boldsymbol{C} + \sigma^2\boldsymbol{V}$. By modeling covariance $\boldsymbol{C}(\boldsymbol{u}, \boldsymbol{v})$ function through basis, we have

$$\boldsymbol{\Sigma} = \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}' + \sigma^2\boldsymbol{V} \tag{2.13}$$

where $\boldsymbol{K}$ is a unknown, positive definite $r \times r$ matrix, $\boldsymbol{S}$ is a $n \times r$ matrix whose (i, l) element is $S_l(\boldsymbol{s}_i)$, assumed known $\boldsymbol{V}$ is a diagonal matrix with entries given by the measurement error variance; assumed known $\sigma^2 > 0$ Also note that

$$\boldsymbol{c}(\boldsymbol{s}_0)' = \text{cov}\{Y(\boldsymbol{s}_0), \boldsymbol{Z}\} = \boldsymbol{S}(\boldsymbol{s}_0)'\boldsymbol{K}\boldsymbol{S}' \tag{2.14}$$

As we mentioned before, we are interested in making inference on Y process, i.e. obtaining $\hat{Y}(\boldsymbol{s}_0)$(2.7) and $\sigma_k(\boldsymbol{s}_0)$(2.10). Through modeling covariance function in (2.12), we could do kriging

prediction with appropriate choice of matrix $\boldsymbol{K}$ and basis function $\boldsymbol{S}(\boldsymbol{u})$. We discussed the choice of $\boldsymbol{K}$ and $\boldsymbol{S}$ in section 3.

As we mentioned before, the original way to achieve kriging prediction involves calculating the inverse $n \times n$ matrix $\boldsymbol{\Sigma}$. The advantage of modeling covariance function in (2.12) is that it provides us an alternative way to compute $\hat{Y}(\boldsymbol{s}_0)$ (2.7) and $\sigma_k(\boldsymbol{s}_0)$(2.10), which only involves the inverse of $r \times r$ matrix when calculating $\boldsymbol{\Sigma}^{-1}$.

Recall (2.13) $\boldsymbol{\Sigma} = \boldsymbol{SKS}' + \sigma^2\boldsymbol{V}$, then

$$\boldsymbol{\Sigma}^{-1} = \sigma^{-1}\boldsymbol{V}^{-1/2}\{\boldsymbol{I} + (\sigma^{-1}\boldsymbol{V}^{-1/2}\boldsymbol{S})\boldsymbol{K}(\sigma^{-1}\boldsymbol{V}^{-1/2}\boldsymbol{S})'\}\sigma^{-1}\boldsymbol{V}^{-1/2} \tag{2.15}$$

According to Sherman-Morrison-Woodbury formula [Henderson and Searle, 1981], we have following results, for any $n \times r$ matrix $\boldsymbol{P}$,

$$\boldsymbol{I} + \boldsymbol{PKP}' = \boldsymbol{I} + (\boldsymbol{I} + \boldsymbol{PKP}')\boldsymbol{PK}(\boldsymbol{I} + \boldsymbol{P}'\boldsymbol{PK})^{-1}\boldsymbol{P}'$$

Multiplying by $(\boldsymbol{I} + \boldsymbol{PKP}')^{-1}$ yields

$$(\boldsymbol{I} + \boldsymbol{PKP}')^{-1} = \boldsymbol{I} - \boldsymbol{P}(\boldsymbol{K}^{-1} + \boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}'$$

We have

$$\boldsymbol{\Sigma}^{-1} = (\sigma^2\boldsymbol{V})^{-1} - (\sigma^2\boldsymbol{V})^{-1}\boldsymbol{S}\{\boldsymbol{K}^{-1} + \boldsymbol{S}'(\sigma^2\boldsymbol{V})^{-1}\boldsymbol{S}\}^{-1}\boldsymbol{S}'(\sigma^2\boldsymbol{V})^{-1} \tag{2.16}$$

In (2.16), it only involves the inverse of fixed rank $r \times r$ positive definite matrix $\boldsymbol{K}$ and the diagonal matrix $\boldsymbol{V}$. Based on (2.17), an efficient spatial kriging prediction is obtained.

$$\hat{Y}(\boldsymbol{s}_0) = \boldsymbol{t}(\boldsymbol{s}_0)'\hat{\boldsymbol{\alpha}} + \boldsymbol{S}(\boldsymbol{s}_0)'\boldsymbol{KS}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{T}\hat{\boldsymbol{\alpha}}) \tag{2.17}$$

$$\begin{aligned}\sigma_k(\boldsymbol{s}_0) =& \{\boldsymbol{S}(\boldsymbol{s_0})'\boldsymbol{KS}(\boldsymbol{s_0}) - \boldsymbol{S}(\boldsymbol{s_0})'\boldsymbol{KS}'\boldsymbol{\Sigma}^{-1}\boldsymbol{SKS}(\boldsymbol{s_0}) + \\ & (\boldsymbol{t}(\boldsymbol{s_0}) - \boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{SKS}(\boldsymbol{s_0}))'(\boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{T})^{-1}(\boldsymbol{t}(\boldsymbol{s_0}) - \boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{SKS}(\boldsymbol{s_0})))\}^{-1/2}\end{aligned} \tag{2.18}$$

where $\hat{\boldsymbol{\alpha}} = (\boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}$ $\quad$ $\boldsymbol{\Sigma}$ in (2.16)

(2.17) - (2.18) are the spatial kriging prediction for **FRK** method [Cressie and Johannesson, 2006]. The **FRK** spatial kriging prediction method still provides prediction map as the original method. However, the computational cost has been reduced tremendously since it only involves solving $r \times r$ matrix not $n \times n$ matrix as before. The original computational cost is $\boldsymbol{O}(n^3)$ and now it decreased to $\boldsymbol{O}(nr^2)$ that is linear on sample size n. Details about how to compute computational burden discussed in [Cressie and Johannesson, 2008].

The estimation of fixed rank positive matrix $\boldsymbol{K}$ from data and the choice of basis functions $\{S_l(\cdot)\}$ become the crucial problems to apply **FRK** method which provides flexibility in spatial covariance function, followed by computationally efficient in kriging predictors and kriging standard error estimation in very large dataset.

In the following section, we will discuss more details about the estimation of $\boldsymbol{K}$ and the basis functions $\{S_l(\cdot)\}$.

3 The class of covariance functions

Recall from (2.12), the covariance function is given by

$$\boldsymbol{C}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{S}(\boldsymbol{u})'\boldsymbol{K}\boldsymbol{S}(\boldsymbol{v}) \qquad \boldsymbol{u},\boldsymbol{v} \in \mathcal{R}^d$$

where $\boldsymbol{K}$ is any $r \times r$ positive definite matrix estimated from the data and $\boldsymbol{S}(\cdot)$ is an $r \times 1$ vector made up of basis functions $S_1(\cdot), \cdots, S_r(\cdot)$ where r is fixed.

3.1 Choice of basis functions

Since we do not pose any assumption on basis function (e.g. orthogonality), so the following basis functions are free to use including smoothing spline basis functions, wavelet basis functions, radial basis functions. While is worth mentioning that from computational prospective, we recommend choosing class of basis functions such that the evaluation of $\boldsymbol{S}^{'}\boldsymbol{V}^{-1}\boldsymbol{S}$ is fast. By using bisquare class or wavelet class, the computational cost will be reducted from $O(nr^2)$ to $O(kr^2)$ where $k << n$. In the experiments, we will use bisquare basis function (i.e $\phi(\boldsymbol{s}_1, \boldsymbol{s}_2) = A\left(1 - \left(\frac{||\boldsymbol{s}_1 - \boldsymbol{s}_2||^2}{R}\right)\right)^2 I(||\boldsymbol{s}_1 - \boldsymbol{s}_2|| < R))$ to make kriging where R is the range of support of the bisquare function [Zammit-Mangion and Cressie, 2017].

3.2 Estimation of Covariance matrix K

Following the classical approach in [Matheron, 1963], an empirical method of moments estimator $\hat{\Sigma}_{\text{MM}}$ is first obtained for Σ. Since $\hat{\Sigma}_{\text{MM}}$ is noisy and may not be positive definite, so one chooses a $\boldsymbol{K} \in \mathcal{R}^{r\times r}, \sigma^2 \in (0, +\infty)$ such that $\Sigma(\hat{\boldsymbol{K}}, \hat{\sigma}^2)$ is close to $\hat{\Sigma}_{\text{MM}}$. So, we can write this in terms of optimization problem which is given by

$$\min_{\boldsymbol{K}\in\mathcal{R}^{r\times r}, \sigma^2\in(0,+\infty)} ||\Sigma(\hat{\boldsymbol{K}}_{\text{MM}}, \hat{\sigma}^2) - \hat{\Sigma}||_F^2 \tag{3.1}$$

Finally, the resulting $\Sigma(\hat{\boldsymbol{K}}, \sigma^2)$ is substituted into the kriging equations (2.17) and (2.18). For clarity, we omit the detail process for solving above optimization problem. The detail can be found in [Cressie and Johannesson, 2008].

4 Experiments

4.1 Synthetic data

4.1.1 Model

[Katzfuss and Cressie, 2011] gives detailed tutorial on the generic recipe for FRK using a global dataset of CO2 measurements. Starting with linear model $\boldsymbol{Z} = \boldsymbol{Y} + \boldsymbol{\epsilon}$, by breaking down $\boldsymbol{Y}$ into

deterministic component $\boldsymbol{X\beta}$ and random spatial-variation component $\boldsymbol{S\eta} + \boldsymbol{\xi}$, so the model is

$$\boldsymbol{Z} = X\boldsymbol{\beta} + S\boldsymbol{\eta} + \boldsymbol{\xi} + \boldsymbol{\epsilon}$$

The covariance matrix is that

$$\Sigma = S^{'}KS + \sigma_{\xi}^2 V_{\xi} + \sigma_{\epsilon}^2 V_{\epsilon}$$

In terms of hierarchical form with respect to the unknown parameter, we have that

$$\begin{aligned}
&\boldsymbol{Z}|\boldsymbol{Y}, \sigma_{\epsilon}^2 \sim N_n(\boldsymbol{Y}, \sigma_{\epsilon}^2 V_{\epsilon}) \\
&\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma_{\xi}^2 \sim N_m(\boldsymbol{X\beta} + S\boldsymbol{\eta}, \sigma_{\xi}^2 V_{\xi}) \\
&\boldsymbol{\eta}|K \sim N_r(\boldsymbol{0}, K)
\end{aligned}$$

So we have trend parameter $\boldsymbol{\beta}$, Measurement-Error Variance σ_{ϵ}^2, Spatial-Dependence Parameters σ_{ξ}^2 and K to be estimated from the data.

4.1.2 Basis function

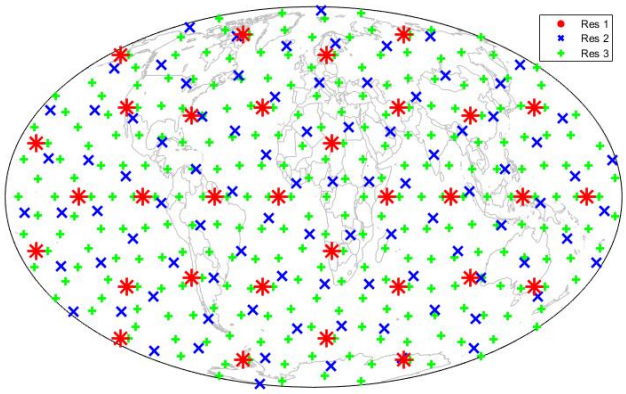

Figure 1: CO2 data

In this example, we chose r = 396 bisquare functions of 3 different resolutions. The 32 basis functions of resolution 1 have a great-arc radius of 6241km, the 92 functions of resolution 2 have a great-arc radius of 3491km, and the 292 functions of resolution 3 have a great-arc radius of 2047km. The locations of the basis-function centers are shown in Figure 1.

4.1.3 Parameter Estimation

First, from expert knowledge Michalak, A. (2010) personal communication, define the covariates to be $\boldsymbol{x} = (1, \textbf{latitude})$, we estimate the $\boldsymbol{\beta}$ by least square estimator which is given by $\hat{\boldsymbol{\beta}} = (X^{'}X)^{-1}X^{'}Z$.

Then for the Measurement-Error Variance σ_ϵ^2 can either be specified from experiments with the measurement instrument or as line's intercept [Kang et al., 2010].

For the Spatial-Dependence Parameters σ_ξ^2 and K, we can either use binned MM estimation from (3.1) or ML estimation via the EM algorithm given in [Katzfuss and Cressie, 2011].

4.1.4 Kriging(Prediction)

By plugging in the estimator into (2.17) and (2.18), we have kriging (prediction) is given by figure 2. The predictions using EM estimations is close to the true data. Also the standard error using EM estimates, is significantly small than the estimator using MM estimates. As the author suggested that the reason is because the EM estimator of $\boldsymbol{K}$ is closer to the empirical covariance structure.

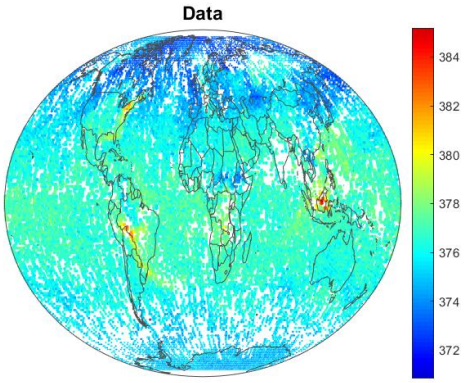

Figure 2: CO2 data

4.2 Real data

In this section, we present how to apply **FRK** method we developed earlier to meuse data. The meuse data set provided by package **sp** is a data set comprising of four heavy metals measured in the top soil in a flood plain along the river Meuse, along with a handful of covariates. The process governing heavy metal distribution seems that polluted sediment is carried by the river, and mostly deposited close to the river bank, and areas with low elevation [Pebesma, 2019].

Package ***FRK*** construct an SRE model on a discretised domain, where the discrete element is known as a basic areal unit (BAU). BAUs essentially allow one to easily combine multiple observations with different supports. Further, the consideration of a discrete element allows one to distinguish between measurement error and fine-scale variation at the resolution of the discrete element which leads to better uncertainty quantification [Zammit-Mangion and Cressie, 2017].

As we can see from the figure 4, the prediction standard error using FRK is very low.

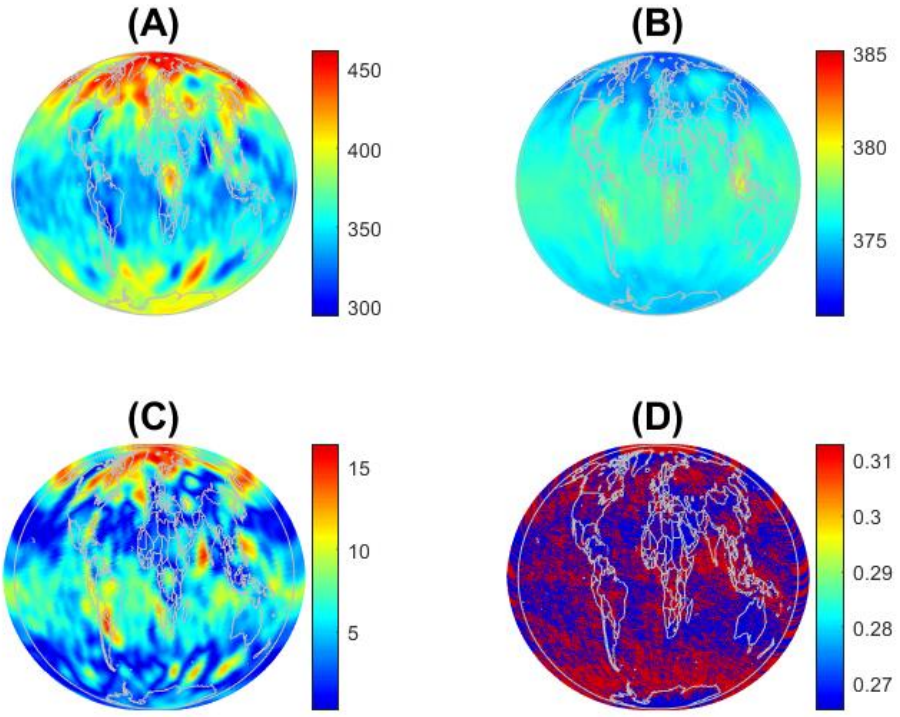

Figure 3: Comparison between predictions and standard error using MM estimates and EM estimates. (A) FRK predictions using MM estimates. (B) FRK predictions using EM estimates. (C) FRK standard errors using MM estimates. (D) FRK standard errors using EM estimates.

5 Conclusion

In summary, we study the fixed rank kriging method for large data set in [Cressie and Johannesson, 2008]. Specifically, fixed rank kriging is kriging within the class of non-stationary covariance functions. It relies on computational simplifications when n is very large, for obtaining the spatial best linear unbiased predictor and its mean-squared prediction error for a hidden spatial process. A method based on minimizing a weighted Frobenius norm yields best estimators of the covariance function parameters, which are then substituted into the fixed rank kriging equations. We applied the method to both synthetic and real data. The result is pretty "good" in terms of prediction standard error. While in real scenarios, we need lots of rules of thumb comes to play in model selection and basis function. So it needs further study to investigate whether the proposed model is truly better than the traditional model.

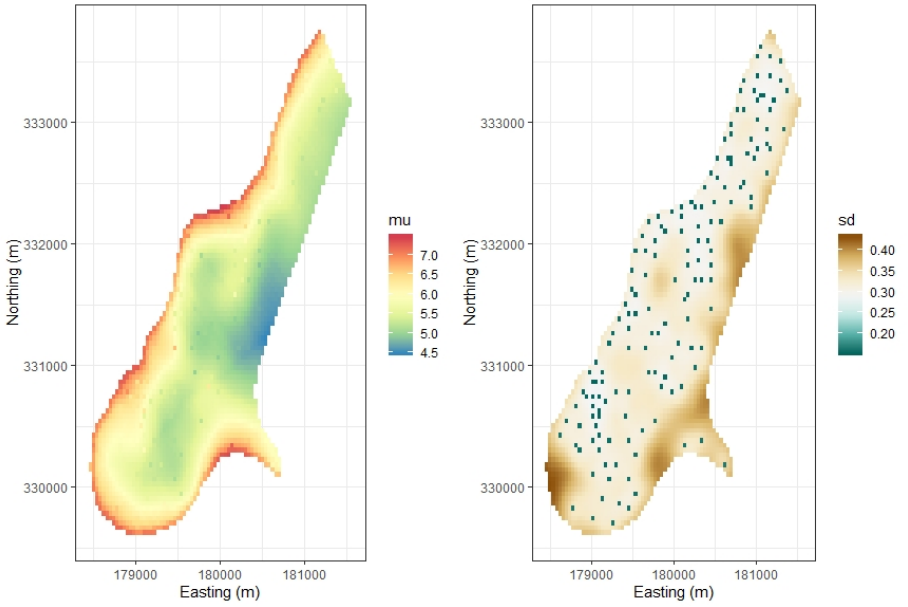

Figure 4: FRK for meuse data. (Left) FRK predictions (Right) FRK standard errors

References

[Cressie, 1992] Cressie, N. (1992). Statistics for spatial data. *Terra Nova*, 4(5):613–617.

[Cressie and Johannesson, 2008] Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

[Cressie and Johannesson, 2006] Cressie, N. A. and Johannesson, G. (2006). Spatial prediction for massive datasets.

[Henderson and Searle, 1981] Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60.

[Kang et al., 2010] Kang, E. L., Cressie, N., and Shi, T. (2010). Using temporal variability to improve spatial mapping with application to satellite data. *Canadian Journal of Statistics*, 38(2):271–289.

[Katzfuss and Cressie, 2011] Katzfuss, M. and Cressie, N. (2011). Tutorial on fixed rank kriging (frk) of co2 data.

[Matheron, 1963] Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

[Pebesma, 2019] Pebesma, E. (2019). The meuse data set: a brief tutorial for the gstat r package.

[Zammit-Mangion and Cressie, 2017] Zammit-Mangion, A. and Cressie, N. (2017). Frk: An r package for spatial and spatio-temporal prediction with large datasets. *arXiv preprint arXiv:1705.08105*.