A Fixed Rank Prediction Algorithm for Massive Spatial Data with Application to Ocean Color

Dr. Roberto Rivera

[1]Department of Business
University of Puerto Rico - Mayaguez

November 30, 2010

Outline

- Spatial model, assumptions and some implementation challenges.
- Summary of Fixed Rank Kriging by Cressie and Johannesson (2008).
- Propose algorithm to estimate FRK parameters σ^2 and K.
- Apply FRK to ocean color to predict missing observations.
- Future paths of research are discussed

Outline

- Spatial model, assumptions and some implementation challenges.
- Summary of Fixed Rank Kriging by Cressie and Johannesson (2008).
- Propose algorithm to estimate FRK parameters σ^2 and K.
- Apply FRK to ocean color to predict missing observations.
- Future paths of research are discussed

Outline

- Spatial model, assumptions and some implementation challenges.
- Summary of Fixed Rank Kriging by Cressie and Johannesson (2008).
- Propose algorithm to estimate FRK parameters σ^2 and K.
- Apply FRK to ocean color to predict missing observations.
- Future paths of research are discussed

Outline

- Spatial model, assumptions and some implementation challenges.
- Summary of Fixed Rank Kriging by Cressie and Johannesson (2008).
- Propose algorithm to estimate FRK parameters σ^2 and K.
- Apply FRK to ocean color to predict missing observations.
- Future paths of research are discussed

Outline

- Spatial model, assumptions and some implementation challenges.
- Summary of Fixed Rank Kriging by Cressie and Johannesson (2008).
- Propose algorithm to estimate FRK parameters σ^2 and K.
- Apply FRK to ocean color to predict missing observations.
- Future paths of research are discussed

Spatial model

$$\begin{aligned} Y(\mathbf{s}) &= \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + W(\mathbf{s}) + \epsilon(\mathbf{s}) \\ &= H(\mathbf{s}) + \epsilon(\mathbf{s}) \end{aligned}$$

- $H(\mathbf{s})$ is L^2 continuous process in space ($\mathbf{s} \in \mathcal{D} \subset \Re^d$) .
- Observations are taken at n locations, $\mathbf{s}_1, ..., \mathbf{s}_n$.
- Assumed that measurement error is present and $\epsilon(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.
- The goal is to predict $H(\mathbf{s}_o)$ for any $\mathbf{s}_o$.
- Sometimes $H(\mathbf{s})$ is modelled deterministically using regression models or semiparametric models (e.g. See Paciorek (2007)).
- Other times $H(\mathbf{s})$ is modelled stochastically with component $W(\mathbf{s})$, that is independent of $\epsilon(\mathbf{s})$ for all $\mathbf{s}$ (e.g. see Cressie, (1993)).

Spatial Prediction of $H(\mathbf{s}_o)$ (Universal Kriging)

$$\hat{H}(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\hat{\boldsymbol{\beta}} + \mathbf{c}_W(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - X\hat{\boldsymbol{\beta}})$$

- where $\mathbf{Y} = (Y(\mathbf{s}_1), ..., Y(\mathbf{s}_n))'$, $\mathbf{x}(\mathbf{s}_o) = (x_1(\mathbf{s}_o), ..., x_p(\mathbf{s}_o))'$
- $Var(\mathbf{Y}) = \Sigma_Y$, $\mathbf{c}_W(\mathbf{s}_o) = Cov(W(\mathbf{s}_o), \mathbf{W})$, $\mathbf{W} = (W(\mathbf{s}_1), ..., W(\mathbf{s}_n))'$
- $\hat{H}(\mathbf{s}_0)$ has prediction variance,

$$\begin{aligned}
\sigma_k^2(\mathbf{s}_0) &= E(H(\mathbf{s}_o) - \hat{H}(\mathbf{s}_o))^2 \\
&= C_W(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}_W(\mathbf{s}_0)'\Sigma_Y^{-1}\mathbf{c}_W(\mathbf{s}_0) \\
&+ (\mathbf{x}(\mathbf{s}_0) - X'\Sigma_Y^{-1}\mathbf{c}_W(\mathbf{s}_0))'(X'\Sigma_Y^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_0) - X'\Sigma_Y^{-1}\mathbf{c}_W(\mathbf{s}_0))
\end{aligned}$$

- Kriging is the Best Linear Unbiased Predictor (BLUP)
- Provided spatial covariance is known.

Spatial Prediction of $H(\mathbf{s}_o)$ (Universal Kriging)

$$\hat{H}(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\hat{\boldsymbol{\beta}} + \mathbf{c}_W(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - X\hat{\boldsymbol{\beta}})$$

- where $\mathbf{Y} = (Y(\mathbf{s}_1), ..., Y(\mathbf{s}_n))'$, $\mathbf{x}(\mathbf{s}_o) = (x_1(\mathbf{s}_o), ..., x_p(\mathbf{s}_o))'$
- $Var(\mathbf{Y}) = \Sigma_Y$, $\mathbf{c}_W(\mathbf{s}_o) = Cov(W(\mathbf{s}_o), \mathbf{W})$, $\mathbf{W} = (W(\mathbf{s}_1), ..., W(\mathbf{s}_n))'$
- $\hat{H}(\mathbf{s}_0)$ has prediction variance,

$$\begin{aligned}
\sigma_k^2(\mathbf{s}_0) &= E(H(\mathbf{s}_o) - \hat{H}(\mathbf{s}_o))^2 \\
&= C_W(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}_W(\mathbf{s}_0)'\Sigma_Y^{-1}\mathbf{c}_W(\mathbf{s}_0) \\
&+ (\mathbf{x}(\mathbf{s}_0) - X'\Sigma_Y^{-1}\mathbf{c}_W(\mathbf{s}_0))'(X'\Sigma_Y^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_0) - X'\Sigma_Y^{-1}\mathbf{c}_W(\mathbf{s}_0))
\end{aligned}$$

- Kriging is the Best Linear Unbiased Predictor (BLUP)
- Provided spatial covariance is known.

Spatial Prediction in practice

- Spatial covariance is seldom known. Optimality properties fall apart.
- Stationarity typically assumed to estimate spatial association.
  - Parametric, exponential example,

$$Cov(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = \begin{cases} \nu^2 exp(-\frac{||\mathbf{h}||}{\rho}), & \text{if } ||\mathbf{h}|| > 0, \\ \nu^2 + \sigma^2 & \text{if } ||\mathbf{h}|| = 0 \end{cases}$$

  where $\boldsymbol{\theta} = (\nu^2, \rho, \sigma^2)'$.

- Σ_Y^{-1} for $\hat{H}(\mathbf{s}_o)$ is $O(n^3)$, making its implementation difficult for data with over a few thousand observations.
- One way to perform spatial prediction when the spatial covariance is unknown is to plug-in estimate $\hat{\theta}$ of $\boldsymbol{\theta}$.

Spatial Prediction in practice

- Spatial covariance is seldom known. Optimality properties fall apart.
- Stationarity typically assumed to estimate spatial association.
  - Parametric, exponential example,

$$Cov(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = \begin{cases} \nu^2 exp(-\frac{||\mathbf{h}||}{\rho}), & \text{if } ||\mathbf{h}|| > 0, \\ \nu^2 + \sigma^2 & \text{if } ||\mathbf{h}|| = 0 \end{cases}$$

where $\boldsymbol{\theta} = (\nu^2, \rho, \sigma^2)'$.

- Σ_Y^{-1} for $\hat{H}(\mathbf{s}_o)$ is $O(n^3)$, making its implementation difficult for data with over a few thousand observations.
- One way to perform spatial prediction when the spatial covariance is unknown is to plug-in estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$.

Spatial Prediction in practice

- Spatial covariance is seldom known. Optimality properties fall apart.
- Stationarity typically assumed to estimate spatial association.
  - Parametric, exponential example,

$$Cov(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = \begin{cases} \nu^2 exp(-\frac{||\mathbf{h}||}{\rho}), & \text{if } ||\mathbf{h}|| > 0, \\ \nu^2 + \sigma^2 & \text{if } ||\mathbf{h}|| = 0 \end{cases}$$

  where $\boldsymbol{\theta} = (\nu^2, \rho, \sigma^2)'$.

- Σ_Y^{-1} for $\hat{H}(\mathbf{s}_o)$ is $O(n^3)$, making its implementation difficult for data with over a few thousand observations.
- One way to perform spatial prediction when the spatial covariance is unknown is to plug-in estimate $\hat{\theta}$ of θ.

Spatial Prediction in practice

- Spatial covariance is seldom known. Optimality properties fall apart.
- Stationarity typically assumed to estimate spatial association.
  - Parametric, exponential example,

$$Cov(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = \begin{cases} \nu^2 exp(-\frac{||\mathbf{h}||}{\rho}), & \text{if } ||\mathbf{h}|| > 0, \\ \nu^2 + \sigma^2 & \text{if } ||\mathbf{h}|| = 0 \end{cases}$$

where $\boldsymbol{\theta} = (\nu^2, \rho, \sigma^2)'$.

- Σ_Y^{-1} for $\hat{H}(\mathbf{s}_o)$ is $O(n^3)$, making its implementation difficult for data with over a few thousand observations.
- One way to perform spatial prediction when the spatial covariance is unknown is to plug-in estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$.

Spatial prediction for massive datasets

- Recall that kriging requires $O(n^3)$ computations for Σ_Y^{-1}.
- Spatial Stationarity is often assumed.
- Recently a method called Fixed Rank Kriging (FRK) has been proposed to obtain prediction using massive datasets. Cressie and Johannesson (2008).
- FRK is much faster than kriging and does not assume stationarity.
- Recall that our model is,

$$\begin{aligned} Y(\mathbf{s}_i) &= \mu(\mathbf{s}_i) + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \\ &= H(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \end{aligned}$$

Spatial prediction for massive datasets

- Recall that kriging requires $O(n^3)$ computations for Σ_Y^{-1}.
- Spatial Stationarity is often assumed.
- Recently a method called Fixed Rank Kriging (FRK) has been proposed to obtain prediction using massive datasets. Cressie and Johannesson (2008).
- FRK is much faster than kriging and does not assume stationarity.
- Recall that our model is,

$$\begin{aligned} Y(\mathbf{s}_i) &= \mu(\mathbf{s}_i) + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \\ &= H(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \end{aligned}$$

Spatial prediction for massive datasets

- Recall that kriging requires $O(n^3)$ computations for Σ_Y^{-1}.
- Spatial Stationarity is often assumed.
- Recently a method called Fixed Rank Kriging (FRK) has been proposed to obtain prediction using massive datasets. Cressie and Johannesson (2008).
- FRK is much faster than kriging and does not assume stationarity.
- Recall that our model is,

$$\begin{aligned} Y(\mathbf{s}_i) &= \mu(\mathbf{s}_i) + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \\ &= H(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \end{aligned}$$

Spatial prediction for massive datasets

- Recall that kriging requires $O(n^3)$ computations for Σ_Y^{-1}.
- Spatial Stationarity is often assumed.
- Recently a method called Fixed Rank Kriging (FRK) has been proposed to obtain prediction using massive datasets. Cressie and Johannesson (2008).
- FRK is much faster than kriging and does not assume stationarity.
- Recall that our model is,

$$
\begin{aligned}
Y(\mathbf{s}_i) &= \mu(\mathbf{s}_i) + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \\
&= H(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)
\end{aligned}
$$

Spatial prediction for massive datasets (contd.)

- If a fix number of basis functions $r << n$ are chosen such that $\mathbf{Z}(\mathbf{s}) \equiv (Z_1(\mathbf{s}), ..., Z_r(\mathbf{s}))'$ are the basis functions,

$$C(W(\mathbf{s}_i), W(\mathbf{s}_j)) \approx \mathbf{Z}(\mathbf{s}_i)^{'} K \mathbf{Z}(\mathbf{s}_j)$$

- Then the data covariance matrix will be,

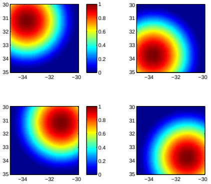
$$Var(\mathbf{Y}) \equiv \Sigma_Y = ZKZ' + \sigma^2 V.$$

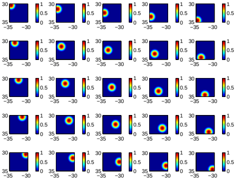
Basis functions

- Examples: Wavelets, thin plate regression splines and eigenvectors
- Multiscale bisquare basis functions

$$Z_{j(l)}(\mathbf{s}) = \begin{cases} \left(1 - (\|\mathbf{s} - \mathbf{v}_{j(l)}\|/r_l)^2\right)^2 & , \text{if } \|\mathbf{s} - \mathbf{v}_{j(l)}\| \leq r_l \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{v}_{j(l)}$ is the j^{th} centroid location at gridding resolution l.





Inversion of Σ_Y^{-1}

- The benefit of representing the covariance matrix in terms of fixed basis functions is that we can use the Sherman-Morrison-Woodbury equation (Golub and Van Loan, (1996)),

$$\Sigma_Y^{-1} = (\sigma^2 V)^{-1} - (\sigma^2 V)^{-1} Z \left(K^{-1} + Z'(\sigma^2 V)^{-1} Z \right)^{-1} Z'(\sigma^2 V)^{-1}$$

- Most importantly notice that now we can obtain Σ_Y^{-1} by only inverting K and diagonal V.

The FRK equations

- Based on the basis function representation and Σ_Y^{-1},

$$\hat{H}(\mathbf{s}_o) = \mathbf{x}(\mathbf{s}_o)'\hat{\boldsymbol{\beta}} + \mathbf{Z}(\mathbf{s}_o)'KZ'\Sigma_Y^{-1}(\mathbf{Y} - X\hat{\boldsymbol{\beta}})$$

  and the prediction variance is,

$$\begin{aligned}\sigma^2_{FRK}(\mathbf{s}_o) &= \mathbf{Z}(\mathbf{s}_o)'K\mathbf{Z}(\mathbf{s}_o) - \mathbf{Z}(\mathbf{s}_o)'KZ'\Sigma_Y^{-1}ZK\mathbf{Z}(\mathbf{s}_o) + (\mathbf{x}(\mathbf{s}_o)' - X'\Sigma_Y^{-1}ZK\mathbf{Z}(\mathbf{s}_o))' \\ &\quad (X'\Sigma_Y^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_o)' - X'\Sigma_Y^{-1}ZK\mathbf{Z}(\mathbf{s}_o))\end{aligned}$$

- σ^2 and the $r \times r$ matrix K need to be estimated

Parameter estimation

- A covariance matrix that is a function of K and σ^2 is fitted to $\widehat{\Sigma}_M$ by minimizing the squared Frobenius norm,

$$\|\widehat{\Sigma}_M - \bar{\Sigma}_M(K, \sigma^2)\|_F^2 = \|\widehat{\Sigma}_M - \bar{Z} K \bar{Z}' - \sigma^2 \bar{V}\|_F^2$$

- This gives a constrained least squares estimate of σ^2 while the estimator of K is,

$$\hat{K}(\hat{\sigma}^2) = R^{-1} Q' \left(\widehat{\Sigma}_M - \hat{\sigma}^2 \bar{V} \right) Q (R^{-1})'.$$

- $\hat{K}(\hat{\sigma}^2)$ needs to be positive definite:
  - To define a proper covariance matrix.
  - To be invertible.
  - To ensure positive prediction variance estimates

Parameter estimation

- A covariance matrix that is a function of K and σ^2 is fitted to $\widehat{\Sigma}_M$ by minimizing the squared Frobenius norm,

$$\|\widehat{\Sigma}_M - \bar{\Sigma}_M(K, \sigma^2)\|_F^2 = \|\widehat{\Sigma}_M - \bar{Z} K \bar{Z}' - \sigma^2 \bar{V}\|_F^2$$

- This gives a constrained least squares estimate of σ^2 while the estimator of K is,

$$\hat{K}(\hat{\sigma}^2) = R^{-1} Q' \left(\widehat{\Sigma}_M - \hat{\sigma}^2 \bar{V} \right) Q (R^{-1})'.$$

- $\hat{K}(\hat{\sigma}^2)$ needs to be positive definite:
  - To define a proper covariance matrix.
  - To be invertible.
  - To ensure positive prediction variance estimates

Parameter estimation

- A covariance matrix that is a function of K and σ^2 is fitted to $\widehat{\Sigma}_M$ by minimizing the squared Frobenius norm,

$$\|\widehat{\Sigma}_M - \bar{\Sigma}_M(K, \sigma^2)\|_F^2 = \|\widehat{\Sigma}_M - \bar{Z}K\bar{Z}' - \sigma^2\bar{V}\|_F^2$$

- This gives a constrained least squares estimate of σ^2 while the estimator of K is,

$$\hat{K}(\hat{\sigma}^2) = R^{-1}Q'\left(\widehat{\Sigma}_M - \hat{\sigma}^2\bar{V}\right)Q(R^{-1})'.$$

- $\hat{K}(\hat{\sigma}^2)$ needs to be positive definite:
  - To define a proper covariance matrix.
  - To be invertible.
  - To ensure positive prediction variance estimates

Parameter estimation

- A covariance matrix that is a function of K and σ^2 is fitted to $\widehat{\Sigma}_M$ by minimizing the squared Frobenius norm,

$$\|\widehat{\Sigma}_M - \bar{\Sigma}_M(K, \sigma^2)\|_F^2 = \|\widehat{\Sigma}_M - \bar{Z}K\bar{Z}' - \sigma^2\bar{V}\|_F^2$$

- This gives a constrained least squares estimate of σ^2 while the estimator of K is,

$$\hat{K}(\hat{\sigma}^2) = R^{-1}Q'\left(\widehat{\Sigma}_M - \hat{\sigma}^2\bar{V}\right)Q(R^{-1})'.$$

- $\hat{K}(\hat{\sigma}^2)$ needs to be positive definite:
  - To define a proper covariance matrix.
  - To be invertible.
  - To ensure positive prediction variance estimates

When is $F = C - bD$ positive definite?

- We state a general result that helps us reach our goal

Lemma

Define $F = C - bD$ where all matrices F, C, D are $r \times r$ real matrices, $C \succ 0$, and $D \succ 0$, and C and D are symmetric. Furthermore, assume F has distinct eigenvalues and that b is any constant such that $b > 0$. Then,

$$F \succ 0 \Leftrightarrow b < \frac{\mathbf{e}_1' C \mathbf{e}_1}{\mathbf{e}_1' D \mathbf{e}_1}$$

where $\mathbf{e}_1$ is the eigenvector associated with minimum eigenvalue of F, λ_1.

- In short, if b is smaller than the bound given in this lemma, F is p.d.

When is $\hat{K}(\hat{\sigma}^2)$ positive definite?

- Notice that the form of $\hat{K}$ is a special case of $F = C - bD$

Corollary

Assume $\hat{K}$ has distinct eigenvalues, $\lambda_1 < .... < \lambda_r$. Then $\hat{K}$ is positive definite if and only if,

$$\hat{\sigma}^2 < \frac{\mathbf{e}_1' R^{-1} Q' \widehat{\Sigma}_M Q (R^{-1})' \mathbf{e}_1}{\mathbf{e}_1' R^{-1} Q' \bar{V} Q (R^{-1})' \mathbf{e}_1}$$

where $\mathbf{e}_1$ is the $r \times 1$ normalized eigenvector corresponding to the smallest eigenvalue λ_1 of $\hat{K}$.

- The result of this corollary inspires the use of the required positive definiteness of $\hat{K}$ as a linear constraint on $\hat{\sigma}^2$.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, \ldots$ Set $\hat{\sigma}_g^2$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q'\left(\widehat{\Sigma}_M - \hat{\sigma}_g^2\bar{V}\right)Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}_g^2$. Let the upper bound be $\hat{\sigma}_{u,g}^2$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}_g^2$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}_{u,g}^2$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}_g^2$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, ...$ Set $\hat{\sigma}_g^2$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q' \left(\widehat{\Sigma}_M - \hat{\sigma}_g^2 \bar{V} \right) Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}_g^2$. Let the upper bound be $\hat{\sigma}_{u,g}^2$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}_g^2$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}_{u,g}^2$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}_g^2$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, ...$ Set $\hat{\sigma}_g^2$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q'\left(\widehat{\Sigma}_M - \hat{\sigma}_g^2\bar{V}\right)Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}_g^2$. Let the upper bound be $\hat{\sigma}_{u,g}^2$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}_g^2$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}_{u,g}^2$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}_g^2$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, ...$ Set $\hat{\sigma}_g^2$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q' \left(\widehat{\Sigma}_M - \hat{\sigma}_g^2 \bar{V}\right) Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}_g^2$. Let the upper bound be $\hat{\sigma}_{u,g}^2$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}_g^2$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}_{u,g}^2$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}_g^2$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, ...$ Set $\hat{\sigma}_g^2$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q' \left(\widehat{\Sigma}_M - \hat{\sigma}_g^2 \bar{V}\right) Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}_g^2$. Let the upper bound be $\hat{\sigma}_{u,g}^2$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}_g^2$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}_{u,g}^2$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}_g^2$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, \ldots$ Set $\hat{\sigma}_g^2$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q'\left(\widehat{\Sigma}_M - \hat{\sigma}_g^2\bar{V}\right)Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}_g^2$. Let the upper bound be $\hat{\sigma}_{u,g}^2$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}_g^2$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}_{u,g}^2$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}_g^2$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

FRK Spatial dependence estimation algorithm

- We propose the following algorithm to iteratively estimate σ^2 and K,
  1. Calculate Q, R, $\bar{V}$, and $\hat{\Sigma}_M$.
  2. Estimate σ^2 by minimizing the Frobenius norm only subject to a constraint that $\hat{\sigma}^2 > 0$. Start at zero an index of the iteration, $g = 0, 1, ...$ Set $\hat{\sigma}^2_g$ as the result of the initial minimization.
  3. Calculate $\hat{K}_g \equiv \hat{K}(\hat{\sigma}_g)$ using $\hat{K}_g = R^{-1}Q'\left(\widehat{\Sigma}_M - \hat{\sigma}^2_g \bar{V}\right)Q(R^{-1})'$.
  4. Check if $\hat{K}_g \succ 0$. This is so if $\lambda_{min,g} > 0$. If not, calculate an upper bound for $\hat{\sigma}^2_g$. Let the upper bound be $\hat{\sigma}^2_{u,g}$.
  5. Minimize the squared Frobenius norm over $\hat{\sigma}^2_g$ but now subject to both, the greater than zero constraint and to the upper bound $\hat{\sigma}^2_{u,g}$ constraint.
  6. Repeat steps 3-5 above until $\hat{K}_g \succ 0$. Then $\hat{\sigma}^2_g$ will be the 'best' estimator of σ^2 such that $\hat{K}_g$ is positive definite.

The FRK spatial dependence estimation algorithm numerically converges

- As g increases, does the algorithm lead to a solution to the estimation problem?

Theorem

If $\lambda_{min,g}$ is the minimum eigenvalue of $\hat{K}_g$ at iteration g, $\hat{K}_g$ has distinct eigenvalues $\lambda_{min,g}, \ldots, \lambda_{max,g}, \forall g$ and $\sigma^2_{u,g}$ is the upper bound found in Step 4 of the FRK parameter estimation algorithm at iteration g, then $\lambda_{min,g} > \lambda_{min,g-1}$ if and only if $\hat{\sigma}^2_g < \hat{\sigma}^2_{g-1}$.

Ocean color

- Ocean color can measure phytoplankton
- Enable scientists to study biological and biogeochemical properties of the oceans.
- Specifically it is crucial for:
  - the study of organic matter produced by algae and bacteria,
  - the study of the biochemistry of the ocean,
  - the assessment of the role of the ocean in the carbon cycle,
  - and the potential global warming trend
- Yoder and Kennelly (2003), Siegel et al. (2002), Siegel et al. (2005b)

Ocean color

- Ocean color can measure phytoplankton
- Enable scientists to study biological and biogeochemical properties of the oceans.
- Specifically it is crucial for:
  - the study of organic matter produced by algae and bacteria,
  - the study of the biochemistry of the ocean,
  - the assessment of the role of the ocean in the carbon cycle,
  - and the potential global warming trend
- Yoder and Kennelly (2003), Siegel et al. (2002), Siegel et al. (2005b)

Ocean color

- Ocean color can measure phytoplankton
- Enable scientists to study biological and biogeochemical properties of the oceans.
- Specifically it is crucial for:
    - the study of organic matter produced by algae and bacteria,
    - the study of the biochemistry of the ocean,
    - the assessment of the role of the ocean in the carbon cycle,
    - and the potential global warming trend
- Yoder and Kennelly (2003), Siegel et al. (2002), Siegel et al. (2005b)

Ocean color

- Ocean color can measure phytoplankton
- Enable scientists to study biological and biogeochemical properties of the oceans.
- Specifically it is crucial for:
  - the study of organic matter produced by algae and bacteria,
  - the study of the biochemistry of the ocean,
  - the assessment of the role of the ocean in the carbon cycle,
  - and the potential global warming trend
- Yoder and Kennelly (2003), Siegel et al. (2002), Siegel et al. (2005b)

Ocean color

- Ocean color can measure phytoplankton
- Enable scientists to study biological and biogeochemical properties of the oceans.
- Specifically it is crucial for:
  - the study of organic matter produced by algae and bacteria,
  - the study of the biochemistry of the ocean,
  - the assessment of the role of the ocean in the carbon cycle,
  - and the potential global warming trend
- Yoder and Kennelly (2003), Siegel et al. (2002), Siegel et al. (2005b)

Ocean color satellite missions

- With satellite ocean color data, analysis of space and time variability of the processes that regulate ocean color can now be conducted
- Doney et al. (2003) and Fuentes et al. (2000) present studies of the spatial correlation of chlorophyll at the mesoscale.
- Siegel et al. (2005a) analyze the association of Inherent Optical Properties

Satellite ocean color woes

- Datasets are massive
- Ocean processes are generally non-stationary in both space and time.
- Data have large amounts of missing data.
  - Cloud cover
  - Orbital sampling
  - and Sun glint among other things.

Satellite ocean color woes

- Datasets are massive
- Ocean processes are generally non-stationary in both space and time.
- Data have large amounts of missing data.
  - Cloud cover
  - Orbital sampling
  - and Sun glint among other things.

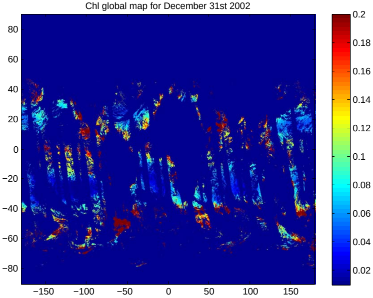
Satellite ocean color woes

- Datasets are massive
- Ocean processes are generally non-stationary in both space and time.
- Data have large amounts of missing data.
  - Cloud cover
  - Orbital sampling
  - and Sun glint among other things.

Satellite ocean color woes

- Datasets are massive
- Ocean processes are generally non-stationary in both space and time.
- Data have large amounts of missing data.
  - Cloud cover
  - Orbital sampling
  - and Sun glint among other things.

Satellite ocean color image



Predicting missing observations

- Several predictors are compared: OLS, AM, Kriging, FRK.
- Campbell (1995) states that CHL_i follows approximately a lognormal distribution. Therefore $Y(\mathbf{s}_i) = log(CHL_i)$

A 'large' region in the North Atlantic

- $n = 3,600$ observations, 15% used as test data.
- AM was fit using thin plate regression splines and the basis function matrix was truncated at 50.
- Kriging was fit according to a Matern covariance function.
- FRK was fit using two scales of variability: with 4, and 25 basis functions.

Model	Mean($\widehat{\textbf{AMSPE}}(\Upsilon_m)$)	CPU time (sec)
OLS	0.0136	0.25
AM	0.0076	7.12
Kriging	0.0048	66.51
FRK	0.0085	5.06

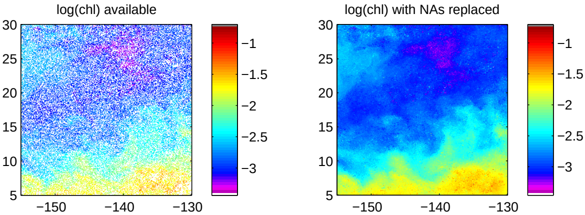
A 'very large' region in the North Pacific

- Almost $n = 90,000$ observations, 50% was used as test data.
- AM was fit using thin plate regression splines and the basis function matrix was truncated at 100.
- FRK was fit using three scales of variability: with 16, 64 and 225 basis functions.

Model	$\widehat{\textbf{AMSPE}}$	CPU time
OLS	0.0516	0.27
AM	0.0169	140.84
FRK	0.0100	167.68

An example of filling missing values using FRK

- Very large region in the North Pacific



Future work

- Assessing the choice of basis functions in FRK predictions.
- A multivariate extension to FRK.
- Implement space-time FRK model to ocean color data.
- Study the spatial variability in ocean color when missing values are imputed using FRK.
- Determining the spatial/temporal distribution of ocean color