Interpolation for huge spatial datasets: theory, implementations and ideas
Joint work

- Emilio Porcu
- Florian Gerber
- Francois Bachoc

and with contributions of several others
Outlook

- Motivation
- Covariance approximations
- Gaussian equivalent measures
- Prediction and estimation
- Implementation and illustration

~ Questions and «Fast-Forward» appreciated!

Slides at http://www.math.uzh.ch/furrer/slides/
Spatial statistics: prediction

Observations: $y(s_1), \ldots, y(s_n)$

First law of geography (Waldo Tobler):

Everything is related to everything else, but near things are more related than distant things.

Source: wikipedia.org
Spatial statistics: prediction

Predict the quantity of interest at an arbitrary location.

Why?
- Fill-in missing data
- Force data onto a regular grid
- Smooth out the measurement error

How?
- By eye
- Linear interpolation
- The correct way . . .
Visual example

Day of the year

NDVI

Source: Gerber et al (2018), TGRS
Visual example

Source: Gerber et al (2018), TGRS
Spatial statistics: prediction

Predict $Z(s_0)$ given $y(s_1), \ldots, y(s_n)$

$$Y(s) = x^T(s)\beta + \mu(s) + Z(s) + \varepsilon(s)$$

Minimize mean squared prediction error (over all linear unbiased predictors)

$\Rightarrow$ Best Linear Unbiased Predictor:

$$\text{BLUP} = \text{Cov}[Z(s_{\text{predict}}), Y(s_{\text{obs}})] \text{Var}[Y(s_{\text{obs}})]^{-1}_{\text{obs}}$$

$$\hat{Z}(s_0) = c^T \Sigma^{-1} y$$

(one spatial process, no trend, known covariance structure; otherwise almost the same)
Spatial statistics: prediction

Predict $Z(s_0)$ given $y(s_1), \ldots, y(s_n)$

$$Y(s) = x^T(s)\beta + \mu(s) + Z(s) + \varepsilon(s)$$

Covariance matrix $\Sigma$ contains elements $C(dist(s_i, s_j))$. 

![Covariance graph](chart.png)
Issues of basic, classical kriging

\[ \text{Cov}(\text{pred}, \text{obs}) \cdot \text{Var}(\text{obs})^{-1} \cdot \text{obs} = c \Sigma^{-1} y \]

- “Simple” spatial interpolation ... ... on paper or in class!

- BUT:
  1. Complex mean structure
  2. Unknown covariance function, unknown parameters
  3. Large spatial fields
  4. Non-stationary covariances
  5. Multivariate/space-time/data on the sphere

Many R packages ... ... many black boxes...

Heaton et al. JABES
Outlook

- Motivation
- Covariance approximations
- Gaussian equivalent measures
- Prediction and estimation
- Implementation and illustration
Spatial modeling

Geostatistical model (GRF):

Lattice model (GMRF):

Covariance matrix: $\Sigma$

$\Sigma = \tau^2(\mathbf{I} - \mathbf{B})^{-1}$
Spatial modeling

Geostatistical model (GRF):

\[ \Sigma \]

\[ \Sigma_{app} \]

Lattice model (GMRF):

\[ \Sigma^{-1} \]

\[ \Sigma \]
Sparseness

Using sparse covariance functions for greater computational efficiency.

Sparseness is guaranteed when

- the covariance function has a compact support
- a compact support is (artificially) imposed \( \Rightarrow \) tapering
Tapering: asymptotic optimality

For an isotropic and stationary process with covariance $C_0(h)$, find a taper $C_\rho(h)$, such that kriging with $C_0(h)C_\rho(h)$ is asymptotically optimal.

$$\frac{\text{MSPE}(s_0, C_0C_\rho)}{\text{MSPE}(s_0, C_0)} = \frac{\text{MSPE}_{\text{taper}}}{\text{MSPE}_{\text{true}}} \to 1$$

$$\frac{\varrho(s_0, C_0C_\rho)}{\text{MSPE}(s_0, C_0)} = \frac{\text{MSPE}_{\text{naive}}}{\text{MSPE}_{\text{true}}} \to 1$$

$$\varrho(s_0, C) = C(0) - c^T \Sigma^{-1} c$$
Tapering: asymptotic optimality

For an isotropic and stationary process with covariance $C_0(h)$, find $C_1(h)$, such that kriging with $C_0(h)C_1(h)$ is asymptotically optimal.

$$\frac{\text{MSPE}(s_0, C_0 C_\rho)}{\text{MSPE}(s_0, C_0)} = \frac{\text{MSPE}_{\text{taper}}}{\text{MSPE}_{\text{true}}} \to 1$$

$$\frac{\varrho(s_0, C_0 C_\rho)}{\text{MSPE}(s_0, C_0)} = \frac{\text{MSPE}_{\text{naive}}}{\text{MSPE}_{\text{true}}} \to 1$$

$\varrho(s_0, C) = C(0) - c^T \Sigma^{-1} c$

Proofs based on infill asymptotics and “misspecified” covariances

Conditions on the tail behaviour of the spectrum of the (tapered) covariance

Furrer, Genton, Nychka (2006) JCGS
Stein (2013) JCGS
Covariance functions

Matérn:

\[ M_\nu(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} r^\nu K_\nu(r) \quad \nu > 0 \]

\[ M_{\nu,\alpha,\sigma^2}(r) = \sigma^2 M_\nu(r/\alpha) \quad \alpha > 0 \quad \sigma^2 > 0 \]

Generalized Wendland:

\[ W_{\mu,\kappa}(r) = \frac{1}{B(2\kappa, \mu + 1)} \int_r^\infty u(u^2 - r^2)^{\kappa-1} (1 - u)^{\mu}_+ \, du \quad \kappa > 0 \quad \mu \geq (d + 1)/2 + \kappa \]

\[ W_{\mu,\kappa,\beta,\sigma^2}(r) = \sigma^2 W_{\mu}(r/\beta) \quad \beta > 0 \quad \sigma^2 > 0 \]
Outlook

- Motivation
- Covariance approximations
- Gaussian equivalent measures
- Prediction and estimation
- Implementation and illustration
Equivalence of Gaussian measures

$P_0$, $P_1$ two probability measures on a measurable space $\{\Omega, \mathcal{S}\}$.

**Definition.** $P_0$, $P_1$ are equivalent if $P_0(A) = 1$, for any $A \in \mathcal{S}$ implies $P_1(A) = 1$ and vice versa.

- Restrict $A$ to the $\sigma$-algebra generated by $\{Z(s), s \in D\}$
  Here: ‘≡’ equivalent on the paths of $\{Z(s), s \in D\}$.

- For Gaussian measures:
  equivalence or orthogonality
  first two moments discriminate: $P(\mu, C)$

- Here: $\mu = 0$ and $D \subset \mathbb{R}^d$, $d = 1, 2, 3$
Equivalence of Gaussian measures

Covariance functions $C_0, C_1$, associated spectral densities $C_0^F, C_1^F$.

$D \subset \mathbb{R}^d$ any bounded infinite set, $s_0 \in D$.

**Theorem.** (Stein 2004)

If $\exists a > 0$, 

$$0 < b_\ell \leq C_0^F(z) z^a \leq b_u < \infty, \quad z \to \infty$$

and if

$$\int_{0<b}^{\infty} z^{d-1} \left( \frac{C_1^F(z) - C_0^F(z)}{C_0^F(z)} \right)^2 \, dz < \infty,$$

then $\mathbb{P}(C_0) \equiv \mathbb{P}(C_1)$. 
Fourier transforms

\[ \mathcal{M}_{\nu}^{\mathcal{F}}(z) = \text{cte} \cdot \frac{1}{(1 + z^2)^{\nu + d/2}}, \quad z \geq 0. \]

\[ \mathcal{W}_{\mu,\kappa}^{\mathcal{F}}(z) = \text{cte} \cdot {\text{1F2}}\left(\lambda; \lambda + \frac{\mu}{2}, \lambda + \frac{\mu}{2} + \frac{1}{2}; -\frac{z^2}{4}\right), \quad \lambda = (d + 1)/2 + \kappa \]

\[ \propto z^{-2\lambda} \quad \text{for} \quad z \to \infty, \]
Equivalence of Gaussian measures

**Theorem.** (Zhang 2004)

Let \( \mathbb{P}(\mathcal{M}_{\nu,\alpha_0,\sigma_0^2}) \), \( \mathbb{P}(\mathcal{M}_{\nu,\alpha_1,\sigma_1^2}) \), \( D \subset \mathbb{R}^d \), \( d = 1, 2, 3 \).

\[ \mathbb{P}(\mathcal{M}_{\nu,\alpha_0,\sigma_0^2}) \equiv \mathbb{P}(\mathcal{M}_{\nu,\alpha_1,\sigma_1^2}) \iff \sigma_0^2 / \alpha_0^{2\nu} = \sigma_1^2 / \alpha_1^{2\nu} \]

- \( d \leq 3 \): not all parameters consistently estimable
- Microergodic parameter \( \sigma^2 / \alpha^{2\nu} \)
- \( d \geq 5 \) all fine (Anderes 2010)
- \( d = 4 \) open . . .
Equivalence of Gaussian measures

**Theorem.** (BFFP 2019)

Let $P(W_{\mu,\kappa,\beta,\sigma_0^2})$, $P(W_{\mu,\kappa,\beta,\sigma_1^2})$.

For a given $\kappa \geq 0$, let $\mu > d + 1/2 + \kappa$:

$$P(W_{\mu,\kappa,\beta,\sigma_0^2}) \equiv P(W_{\mu,\kappa,\beta,\sigma_1^2}) \iff \frac{\sigma_0^2}{\beta_0^{2\kappa+1}} = \frac{\sigma_1^2}{\beta_1^{2\kappa+1}}$$

**Theorem.** (BFFP 2019)

Let $P(M_{\nu,\alpha,\sigma_0^2})$ and $P(W_{\mu,\kappa,\beta,\sigma_1^2})$, $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, $\kappa > 0$.

If $\nu = \kappa + 1/2$, $\mu > d + 1/2 + \kappa$:

$$P(M_{\nu,\alpha,\sigma_0^2}) \equiv P(W_{\mu,\kappa,\beta,\sigma_1^2}) \iff \frac{\sigma_0^2}{\alpha^{2\nu}} = \text{cte}_{\nu,\kappa,\mu} \cdot \frac{\sigma_1^2}{\beta^{2\kappa+1}}$$
Outlook

- Motivation
- Covariance approximations
- Gaussian equivalent measures
- Prediction and estimation
- Implementation and illustration
Prediction: asymptotic optimality

Prediction of $Z(s_0)$ using “BLUP” under misspecification $\hat{Z}_n(\mu, \kappa, \beta)$.

**Theorem.** (BFFP 2019)

$\mathbb{P}(W_{\mu, \kappa, \beta, \sigma^2}):$ if $\mu > (d + 1)/2 + \kappa$ and for any fixed $\beta_1 > 0$

\[
\frac{\text{Var}_{\mu, \kappa, \beta, \sigma^2}(\hat{Z}_n(\mu, \kappa, \beta_1) - Z(s_0))}{\text{Var}_{\mu, \kappa, \beta, \sigma^2}(\hat{Z}_n(\mu, \kappa, \beta) - Z(s_0))} \to 1
\]

$\mathbb{P}(M_{\nu, \alpha, \sigma^2}):$ if $\nu = \kappa + 1/2$ and for any fixed $\beta_1 > 0$

\[
\frac{\text{Var}_{\nu, \alpha, \sigma^2}(\hat{Z}_n(\nu, \alpha_1) - Z(s_0))}{\text{Var}_{\nu, \alpha, \sigma^2}(\hat{Z}_n(\nu, \alpha) - Z(s_0))} \to 1
\]
Prediction: presumed MSPE

Prediction of $Z(s_0)$ using “BLUP” under misspecification $\hat{Z}_n(\mu, \kappa, \beta)$.

**Theorem.** (BFFP 2019)

\[
\mathbb{P}(\mathcal{W}_{\mu, \kappa, \beta, \sigma^2}): \quad \text{if } \frac{\sigma_0^2}{\beta_0^{2\kappa+1}} = \frac{\sigma_1^2}{\beta_1^{2\kappa+1}}
\]

\[
\frac{\text{Var}_{\mu, \kappa, \beta_1, \sigma_1^2}(\hat{Z}_n(\mu, \kappa, \beta_1) - Z(s_0))}{\text{Var}_{\mu, \kappa, \beta, \sigma^2}(\hat{Z}_n(\mu, \kappa, \beta_1) - Z(s_0))} \rightarrow 1
\]

\[
\mathbb{P}(\mathcal{M}_{\nu, \alpha, \sigma^2}): \quad \text{if } \nu = \kappa + 1/2, \quad \frac{\sigma^2}{\alpha^{2\nu}} = \text{cte} \cdot \frac{\sigma_1^2}{\beta_1^{2\kappa+1}}
\]

\[
\frac{\text{Var}_{\mu, \kappa, \beta_1, \sigma_1^2}(\hat{Z}_n(\mu, \kappa, \beta_1) - Z(s_0))}{\text{Var}_{\nu, \alpha, \sigma^2}(\hat{Z}_n(\mu, \kappa, \beta_1) - Z(s_0))} \rightarrow 1
\]
Estimation: tapering

Likelihood:

$$-\frac{1}{2} \log \det(\Sigma(\theta) \circ T) - \frac{1}{2} y^T (\Sigma(\theta) \circ T)^{-1} y$$

$$-\frac{1}{2} \log \det(\Sigma(\theta) \circ T) - \frac{1}{2} y^T ((\Sigma(\theta) \circ T)^{-1} \circ T) y$$

- Dimension plays crucial role Kaufman et al. (2008) JASA

- Two-taper approach cannot be used in practice
Estimation: Matérn

Let $\mathcal{P}(\mathcal{M}_{\nu,\alpha_0,\sigma^2_0})$, with certain constrains on the parameter space.

**Theorem.** (Shaby Kaufmann 2013)

Let $\{\hat{\sigma}^2_n, \hat{\alpha}_n\}$ the ML estimator. Then as $n \to \infty$

1. $\hat{\sigma}^2_n / \hat{\alpha}^{2\nu}_n \xrightarrow{a.s.} \sigma^2_0 / \alpha^{2\nu}_0$

2. $\sqrt{n}(\hat{\sigma}^2_n / \hat{\alpha}^{2\nu}_n - \sigma^2_0 / \alpha^{2\nu}_0) \xrightarrow{D} \mathcal{N}(0, 2(\sigma^2_0 / \alpha^{2\nu}_0)^2)$

- Also holds for $\hat{\sigma}^2_n(\alpha)$ with misspecified $\alpha$
Estimation: generalized Wendland

Let $P(\mathcal{W}_{\mu, \kappa, \beta_0, \sigma_0^2})$, with certain constrains on the parameter space.

**Theorem.** (BFFP 2019)

Let $\hat{\sigma}_n^2(\beta)$ the ML estimator with “known” $\beta$. Then, as $n \to \infty$

1. $\hat{\sigma}_n^2(\beta)/\beta^{2\kappa+1} \xrightarrow{a.s.} \sigma_0^2/\beta_0^{2\kappa+1}$

2. $\sqrt{n}(\hat{\sigma}_n^2(\beta)/\beta^{2\kappa+1} - \sigma_0^2/\beta_0^{2\kappa+1}) \xrightarrow{D} \mathcal{N}(0, 2(\sigma_0^2/\beta_0^{2\kappa+1})^2)$
Estimation: generalized Wendland

Let $P(W_{\nu,\alpha,\sigma^2})$, with certain constrains on the parameter space.

**Theorem.** (BFFP 2019)

Let $\{\hat{\sigma}_n^2(\hat{\beta}_n), \hat{\beta}_n\}$ the ML estimator. Then, as $n \to \infty$

1. $\frac{\hat{\sigma}_n^2(\hat{\beta}_n)}{\hat{\beta}_n^{2\kappa+1}} \xrightarrow{a.s.} \frac{\sigma_0^2}{\beta_0^{2\kappa+1}}$

2. $\sqrt{n}(\frac{\hat{\sigma}_n^2(\hat{\beta}_n)}{\hat{\beta}_n^{2\kappa+1}} - \frac{\sigma_0^2}{\beta_0^{2\kappa+1}}) \xrightarrow{D} \mathcal{N}(0, 2(\frac{\sigma_0^2}{\beta_0^{2\kappa+1}})^2)$
Theory vs. practice

- Tapering is “obsolete”, but:
  - with practicable supports, slow convergence
  - evaluation of covariance function

- Univariate setting:
  Proofs based on infill asymptotics and “misspecified” covariances

- Multivariate setting:
  Proofs based on domain increasing framework
  Weak conditions on the taper (Furrer, Du, Bachoc 2016)

- Asymptotics in practice??

Outlook

- Motivation
- Covariance approximations
- Gaussian equivalent measures
- Prediction and estimation
- Implementation and illustration
Implementation: software

Software to exploit the sparse structure \texttt{spam64} for \texttt{R}:

- an R package for \texttt{sparse} matrix algebra
- storage economical and fast
- versatile, intuitive and simple

See Furrer et al. (2006) JCGS; Furrer, Sain (2010) JSS

- R objects have at most $2^{31}$ elements (almost)
- R does not ‘have’ 64-bit integers: stored as doubles
- 64-bit exploitation consists of type conversions between front-end R and pre-compiled code

Gerber, Mösinger, Furrer (2017) CaGeo
Illustration

- residual field from AVHRR NDVI\textsubscript{3g} product:
  \[ y = y_{2000-2009} - y_{1990-1999}, \quad n = 769,940 \text{ observations} \]

- Nonstationary Gaussian random field (zero mean)
Illustration: nonstationary covariance

\[ \Sigma(\kappa, \beta, \tau) = \kappa D(\beta) \left( (1 - \tau)I + \tau R \right) D(\beta) \]

- \( \kappa > 0 \): scaling parameter
- \( D(\beta) = \text{diag}(\exp(X\beta)) \): controls strength via covariates
- \( \tau \in [0, 1] \): "no spatial correlation" vs "spatial correlation"
- \( I \): identity matrix
- \( R \): stationary correlation matrix:
  - compactly supported covariance
  - range 50km, sparsity 0.2%
Fast is relative … *optim* suboptimal

<table>
<thead>
<tr>
<th>Task</th>
<th>Function</th>
<th>Time</th>
<th>Sparsity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distances</td>
<td><code>h &lt;- nearest.dist(...)</code></td>
<td>23min</td>
<td>1.4 Gb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2%</td>
</tr>
<tr>
<td>Covariances</td>
<td><code>T &lt;- cov.Wend(h, ...)</code></td>
<td>2min</td>
<td>1.4 Gb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2%</td>
</tr>
<tr>
<td>Cholesky</td>
<td><code>chol(T, ...)</code></td>
<td>29min</td>
<td>8.5 Gb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.9%</td>
</tr>
</tbody>
</table>

optimization strategy:

- (iterative) grid search over $\tau$
- exploit multicore architecture
- for given $\tau$ use quasi-Newton optimizer to optimize $\kappa, \beta$
Illustration: nonstationary covariance

- with covariates “distance to nearest coast” and “elevation”
- $\text{diag}(\hat{\Sigma})$:

[BIC improvement compared to nonstationary model]

... but not sufficiently flexible
Theory vs. practice

1. Complex mean structure
2. Unknown covariance function, unknown parameters
3. Large spatial fields
4. Non-stationary covariances
5. Multivariate/space-time/data on the sphere

▶ gapfill (distribution free/algorithmic) ◀
Collaboration with:

– Emilio Porcu
– Florian Gerber
– Francois Bachoc
– Moreno Bevilacqua
– Kaspar Mösinger
– former & present ‘Applied Statistics’ team

... and many more