Asymptotic properties of truncated-ML estimators based on covariance approximations

@ReinhardFurrer, I-Math/ICS, UZH

Comstat 2022, Bologna, 2022/08/26

University of Zurich[UZH]

Technical details in https://arxiv.org/pdf/2112.12317

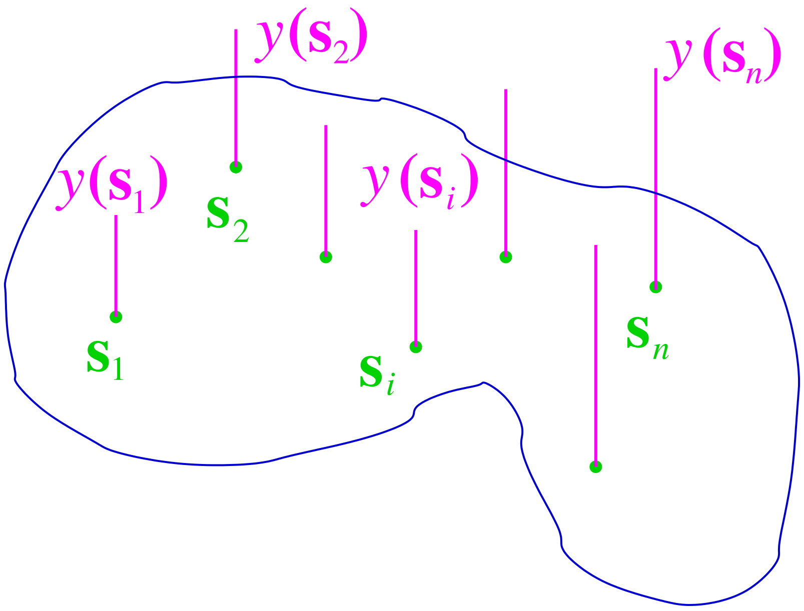
Joint work

- Michael Hediger
- Francois Bachoc
- Emilio Porcu

FNSNF 175529
SWISS NATIONAL SCIENCE FOUNDATION

University of Zurich[UZH]

Spatial statistics



$$Y(\mathbf{s}) = \mathbf{x}^{\mathsf{T}}(\mathbf{s})\boldsymbol{\beta} + \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s})$$

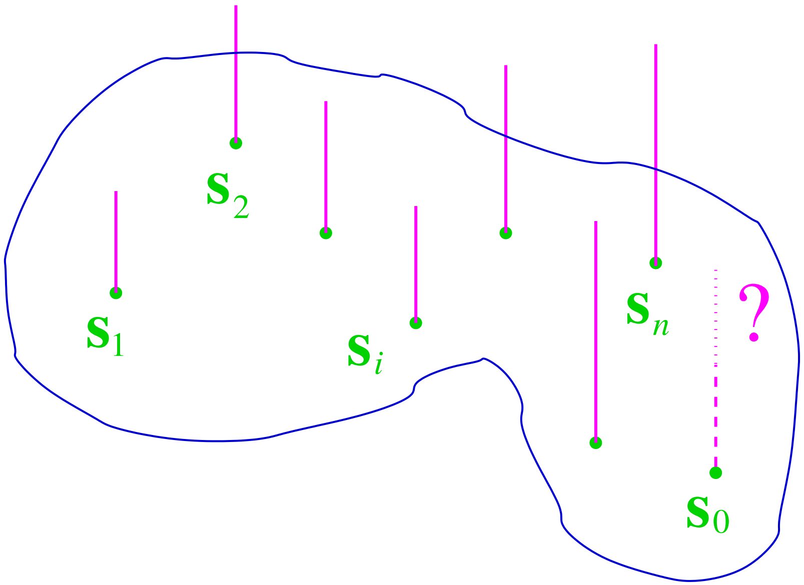
with

- $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d$
- $\mathbf{x}^{\mathsf{T}}(\mathbf{s})\boldsymbol{\beta}$, $\mu(\mathbf{s})$ large scale/mean structure
- $Z(\mathbf{s})$, $\varepsilon(\mathbf{s})$ Gaussian, independent

University of Zurich[UZH]

Spatial statistics: prediction

Predict $Z(\mathbf{s}_0)$ given $y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)$



$$Y(\mathbf{s}) = \mathbf{x}^\top(\mathbf{s})\beta + \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s})$$

⇝ Best Linear Unbiased Predictor:

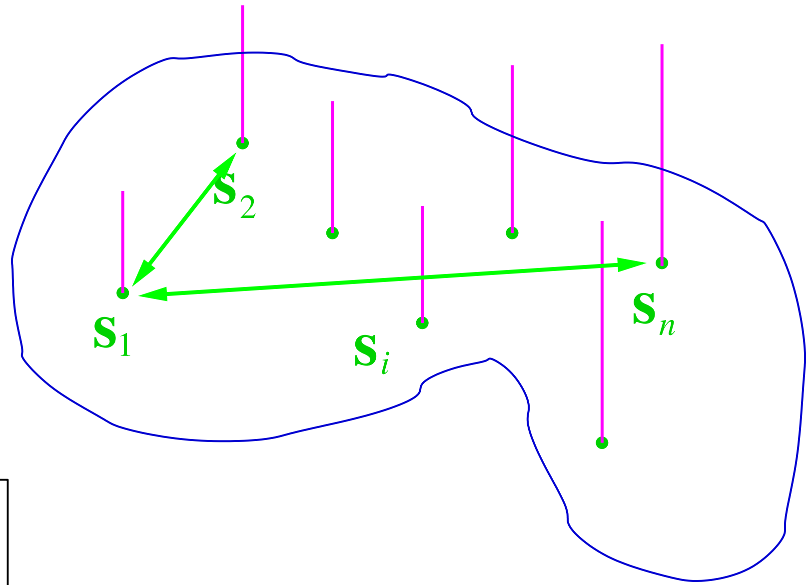
$$\text{BLUP} = \text{Cov}\Big[Z(\mathbf{s}_{\text{predict}}), Y(\mathbf{s}_{\text{obs}})\Big]\, \text{Var}\Big[Y(\mathbf{s}_{\text{obs}})\Big]^{-1} \text{obs}$$

$$\widehat{Z}(\mathbf{s}_0) = \mathbf{c}^\top \boldsymbol{\Sigma}^{-1}\, \mathbf{y}$$

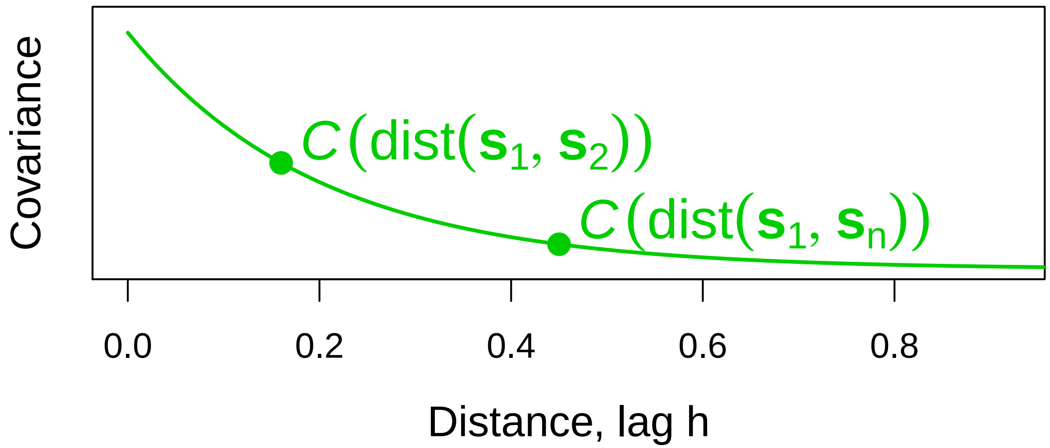
(one spatial process, no trend, known covariance structure; otherwise *almost* the same)

Spatial statistics: prediction

Predict $Z(\mathbf{s}_0)$ given $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$



$$Y(\mathbf{s}) = \mathbf{x}^\top(\mathbf{s})\beta + \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s})$$



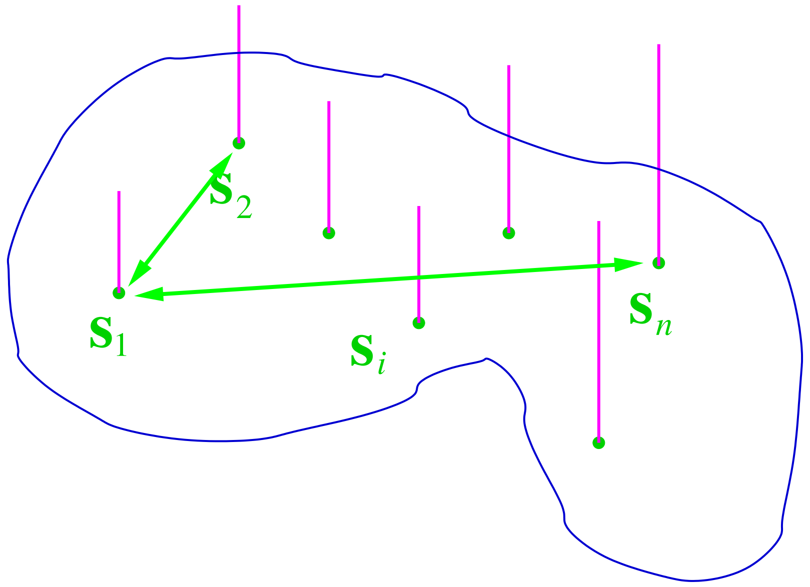
Covariance function $C\big(\text{dist}(\mathbf{s}_i, \mathbf{s}_j)\big)$ parameterized by $\boldsymbol{\theta}$

... induces covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

University of Zurich[UZH]

Spatial statistics: estimation

$$Y(\mathbf{s}) = \mathbf{x}^{\mathsf{T}}(\mathbf{s})\beta + \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s})$$



Estimation of $\boldsymbol{\theta}$ (and β) when

- n large or huge
- based on the likelihood
- using sparse covariance matrices

University of Zurich[UZH]

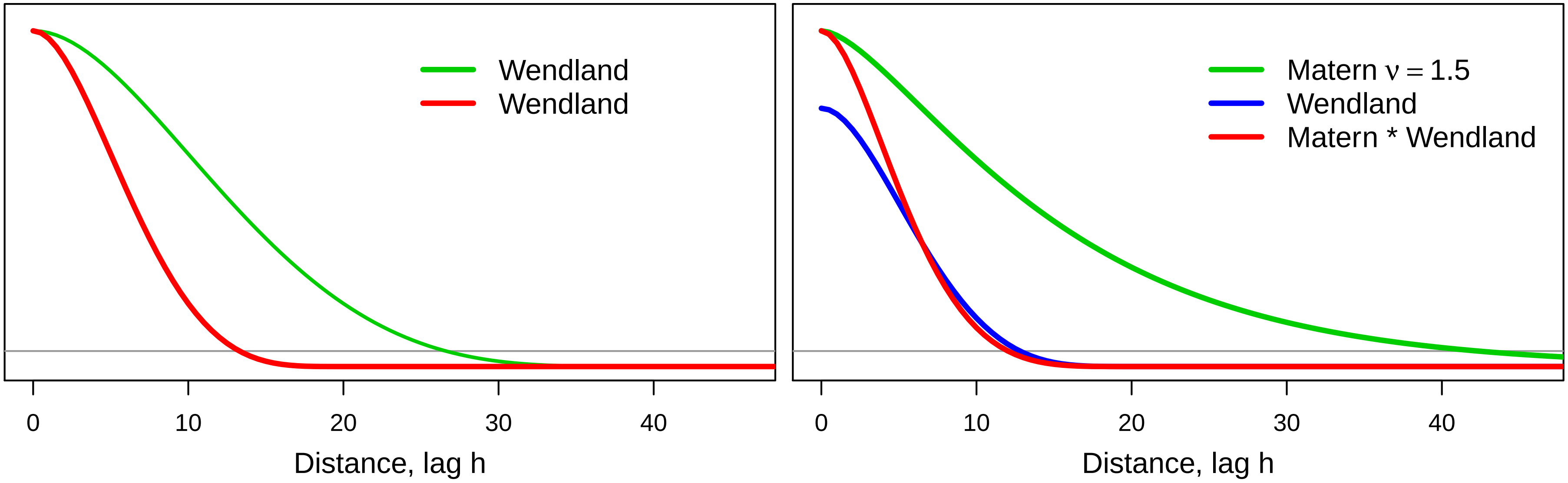
Approaches for large spatial datasets

- Sparse Covariance methods:
  - Covariance Tapering — Furrer
  - Spatial Partitioning — Heaton
- Sparse Precision methods:
  - Lattice Kriging — Nychka
  - Multiresolution Approximations — Katzfuss
  - Stochastic Partial Differential Equations — Lindgren
  - Periodic Embedding — Guinness
  - Nearest Neighbor Processes — Datta
- Low rank approximation:
  - Fixed Rank Kriging — Zammit-Mangion
  - Predictive Processes — Finley
- Algorithmic approaches:
  - Gapfill — Gerber
  - Local Approximate Gaussian Processes — Gramacy
  - Metakriging — Guhaniyogi

Sparse covariance matrix

Using sparse covariance matrix for greater computational efficiency.

- Covariance function has a compact support
  Typically support too large $\rightsquigarrow$ direct misspecification
- Compact support is (artificially) imposed $\rightsquigarrow$ tapering



"Object of desire"

What are constraints on the deliberate mis-specification to conclude

$$\mathsf{MSPE}\left(\widetilde{\boldsymbol{\theta}}_n\right)/\mathsf{MSPE}\left(\widehat{\boldsymbol{\theta}}_n\right) \xrightarrow{n\to\infty} 1$$

$$\widetilde{\boldsymbol{\theta}}_n \xrightarrow[n\to\infty]{\mathbb{P}} \boldsymbol{\theta}_0$$

$$n^{1/2}\left(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\mathsf{d}} \mathcal{N}_p\big(\mathbf{0}, \boldsymbol{\Lambda}^{-1}\big)$$

(A) in a infill-domain asymptotic framework?

- Use Gaussian equivalent measures, handle spectral densities:

tapering (prediction)	FGN 2006
direct misspecification	BFFP 2019
multivariate direct misspecification	BPBFF 2022

(B) in a increasing-domain asymptotic framework?

- Bound smallest and largest eigenvalues of $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$:

multivariate tapering	FBD 2016
transformed Gaussian prcesses	BBFK 2022

“Object of desire”

What are constraints on the deliberate mis-specification to conclude

$$\text{MSPE}\left(\widetilde{\boldsymbol{\theta}}_n\right)/\text{MSPE}\left(\widehat{\boldsymbol{\theta}}_n\right) \xrightarrow{n\to\infty} 1$$

$$\widetilde{\boldsymbol{\theta}}_n \xrightarrow[n\to\infty]{\mathbb{P}} \boldsymbol{\theta}_0$$

$$n^{1/2}\left(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}_p\Big(\mathbf{0}, \boldsymbol{\Lambda}^{-1}\Big)$$

Covariance is approximated with an “arbitrary” function

not-necessarily positive definite!

Increasing-domain asymptotic framework

Consistency & asymptotic normality

Perturbed grid increasing-domain asymptotic framework

Assumptions $c_{\boldsymbol{\theta}}$:

(1) compact support

(2) partial derivatives exist and are bounded

(3) Fourier inversion exists, local and asymptotic identifiability

Assumptions on approximations $\left\{(\widetilde{c}_{m,\boldsymbol{\theta}})_m\right\}$:

(1), (2) as above

(4) $\widetilde{c}_{m,\boldsymbol{\theta}} \xrightarrow{m\to\infty} c_{\boldsymbol{\theta}}$ uniformly on $\boldsymbol{\Theta}$, $\mathbb{R}^d$, (same for its derivatives)

Theorem. Let $\left(\widetilde{\boldsymbol{\theta}}_n\right)_{n\in\mathbb{N}}$ be a sequence of truncated-ML estimators for $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, based on $\left\{\left(\widetilde{c}_{m,\boldsymbol{\theta}}\right)_{m\in\mathbb{N}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$. Under assumptions above:

$$\widetilde{\boldsymbol{\theta}}_n \xrightarrow[n\to\infty]{\mathbb{P}} \boldsymbol{\theta}_0 \qquad\qquad n^{1/2}\left(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$$

University of Zurich[UZH]

Idea of proof I

Let $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$ based on $c_{\boldsymbol{\theta}}$ and $\widetilde{c}_{m,\boldsymbol{\theta}}$, respectively.
Let $m = r(n) \xrightarrow{n\to\infty} \infty$ (no specific rate).

The largest and smallest eigenvalues are bounded

$$\sup_{n\in\mathbb{N}} \sup_{\mathbf{s}_{(n)}\in\mathcal{G}_n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) \right\|_2 < \infty$$

$$\inf_{n\geq N} \inf_{\mathbf{s}_{(n)}\in\mathcal{G}_n} \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \lambda_n\Big(\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\theta})\Big) > 0$$

Further, the matrices are asymptotically equivalent

$$\sup_{\mathbf{s}_{(n)}\in\mathcal{G}_n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \boldsymbol{\Sigma}(\boldsymbol{\theta}) - \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) \right\|_2 \xrightarrow{n\to\infty} 0$$

Results depend on compact support of $c_{\boldsymbol{\theta}}$ and $\widetilde{c}_{m,\boldsymbol{\theta}}$!

Idea of proof II

Classical Likelihood:

$$\propto \frac{1}{n} \log \det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \frac{1}{n} \mathbf{z}^{\mathsf{T}} (\boldsymbol{\Sigma}(\boldsymbol{\theta}))^{-1} \mathbf{z}$$

For $c_{\boldsymbol{\theta}}$, define the random variable

$$\ell_n(\theta) = \frac{1}{n} \log \Big(1 + \big(\det(\boldsymbol{\Sigma}_n(\boldsymbol{\theta})) - 1 \big) \mathbb{I}_{\{\det(\boldsymbol{\Sigma}_n(\boldsymbol{\theta}))>0\}} \Big) + \frac{1}{n} \mathbf{Z}_n^{\mathsf{T}} \boldsymbol{\Sigma}_n(\boldsymbol{\theta})^{+} \mathbf{Z}_n$$

Similarly, for $\widetilde{c}_{m,\boldsymbol{\theta}}$ define $\widetilde{\ell}_n(\theta)$.

Then:

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \ell_n(\boldsymbol{\theta}) - \widetilde{\ell}_n(\boldsymbol{\theta}) \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0$$

University of Zurich[UZH]

Examples: Generalized Wendland

Matérn:

$$\mathcal{M}_\nu(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} r^\nu \mathcal{K}_\nu(r) \qquad \nu > 0$$

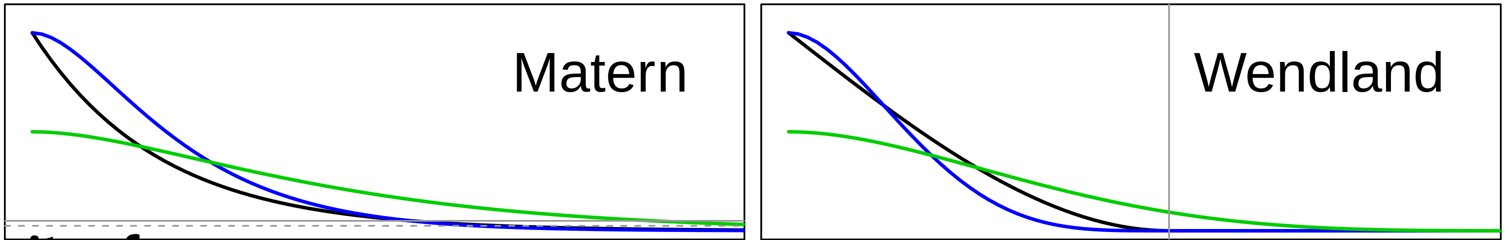
$$\mathcal{M}_{\nu,\alpha,\sigma^2}(r) = \sigma^2 \mathcal{M}_\nu(r/\alpha) \qquad \alpha > 0 \quad \sigma^2 > 0$$

Generalized Wendland:

$$\mathcal{W}_{\mu,\kappa}(r) = \frac{1}{B(2\kappa, \mu+1)} \int_r^\infty u(u^2 - r^2)^{\kappa-1}(1-u)_+^\mu \, \mathrm{d}u$$

$$\kappa > 0 \quad \mu \geq (d+1)/2 + \kappa$$

$$\mathcal{W}_{\mu,\kappa,\beta,\sigma^2}(r) = \sigma^2 \mathcal{W}_\mu(r/\beta) \qquad \beta > 0 \quad \sigma^2 > 0$$



Examples:

Let $\phi(r) = \mathcal{W}_{\mu,\kappa,\beta,\sigma^2}(r)$. Results hold with

Truncation:

$$\widetilde{\phi}(r) = \phi(r)\mathbb{I}_{\{r<\delta_m\}} \qquad \delta_m \xrightarrow{m\to\infty} \beta$$

Linear interpolation:

Linearly interpolating $\phi(r)$ between knots
$0 = k_0 < k_1 < \cdots < k_m = \beta$

Vanishing shift:

$$\widetilde{\phi}(r) = \phi(r) + \delta_m\mathbb{I}_{\{r=0\}} \qquad \delta_m \xrightarrow{m\to\infty} 0$$

Trimmed Bernstein polynomials . . .

Examples: in practice

Let $\phi(r) = \mathcal{W}_{\mu,\kappa,\beta,\sigma^2}(r)$. Results hold with

Truncation:

unstable estimates: *optim()* struggles

Linear interpolation:

speed-up of factor $>$ 50 on covariance construction

Vanishing shift:

difficult to determine optimal shift

Non-compact support:

asymptotically minimizes Kullback–Leibler divergence (tapered misspecified and true distributions)

University of Zurich[UZH]

Extensions/outlook

- Tailor optimization routine
- Determine best-practice settings
- Single precision implementation
- ...

References (some, alphabetical)

Bachoc Porcu Bevilacqua Furrer Faouzi (2022) Asymptotically Equivalent Prediction in Multivariate Geostatistics *Bernoulli*

Bevilacqua Faouzi Furrer Porcu (2019) Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics *AoS* **47** 828–856

Furrer Bachoc Du (2016) Asymptotic Properties of Multivariate Tapering for Estimation and Prediction *JMVA* **149** 177–191

Furrer Genton Nychka (2006) Covariance Tapering for Interpolation of Large Spatial Datasets *JCGS* **15** 502–523

Furrer Hediger (2022) Asymptotic analysis of ML-covariance parameter estimators based on covariance approximations arxiv: 2112.12317

Furrer Sain (2010) spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields *JSS* **36** 1–25

Gerber Moesinger Furrer (2017) Extending R Packages to Support 64-bit Compiled Code: An Illustration with spam64 and GIMMS NDVI3g Data *Comput Geosci* **104** 109–119

Heaton et al (2019) A Case Study Competition among Methods for Analyzing Large Spatial Data *JABES* **24** 398–425

More at: www.math.uzh.ch/furrer/research/publications.shtml

Appendix

Implementation: software

Software to exploit the sparse structure **spam64** for R:

- an R package for 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
Gerber, Mösinger, Furrer (2018) SoftwareX

Implementation: `spam64`

- choose 32-bit or 64-bit compiled code accordingly
- exploit *DUP=FALSE* use

- *spam64* is ‘`sed`‘ed out of *spam*
- *.Fortran* ⟶ *dotCall64::.C64*

University of Zurich[UZH]