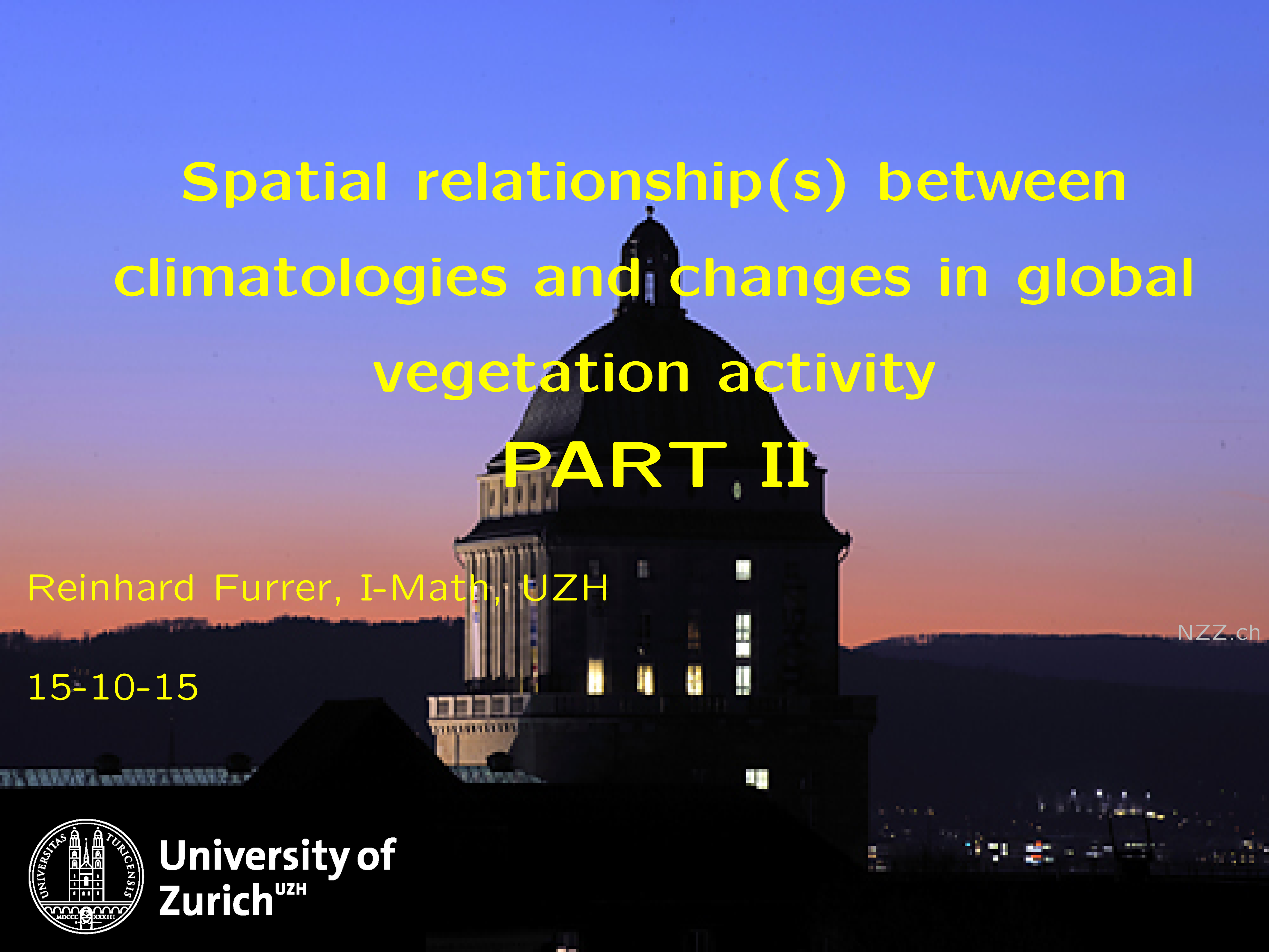


Spatial relationship(s) between climatologies and changes in global vegetation activity

PART II

Reinhard Furrer, I-Math, UZH

15-10-15

University of Zurich[UZH]

Big picture: spatial statistics

Source: wikipedia.org

First law of geography
(Waldo Tobler):

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

University of Zurich[UZH]

Big picture: spatial statistics

Source: wikipedia.org

First law of geography
(Waldo Tobler):

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

Exploit correlation for prediction (smoothing, inter-, extrapolation)

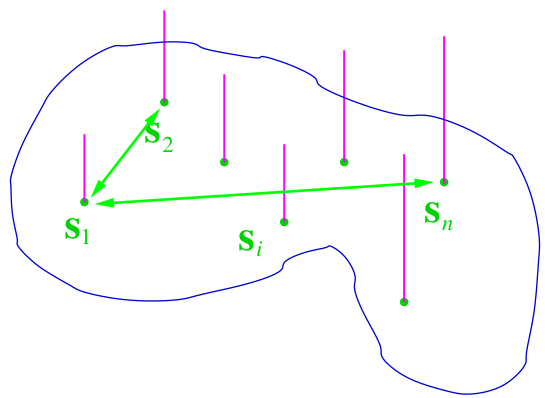
"Spatial" nature encoded in the covariance / precision matrix

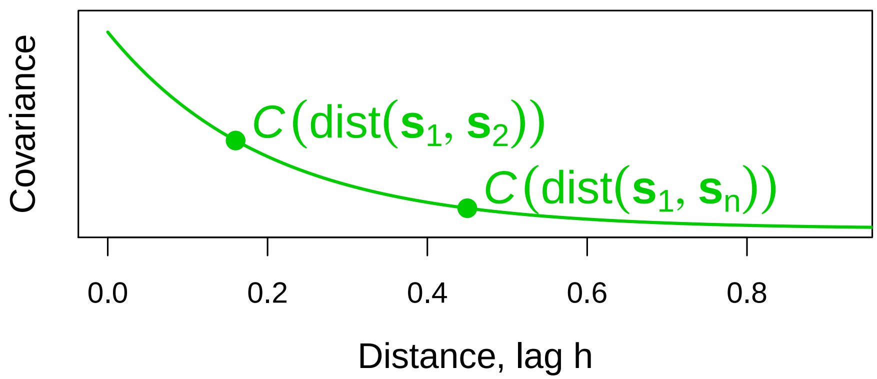
$\rightsquigarrow$ Cholesky factorization

University of Zurich[UZH]

Spatial modeling

Geostatistical model (GRF):



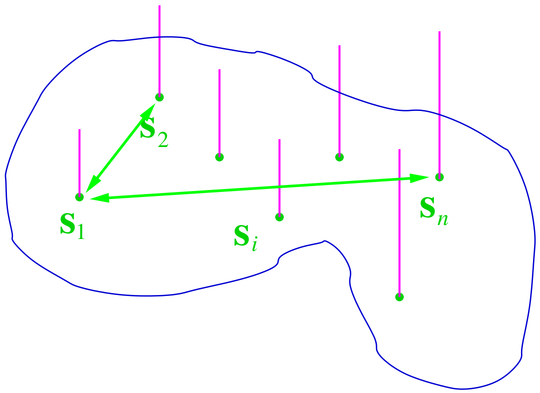


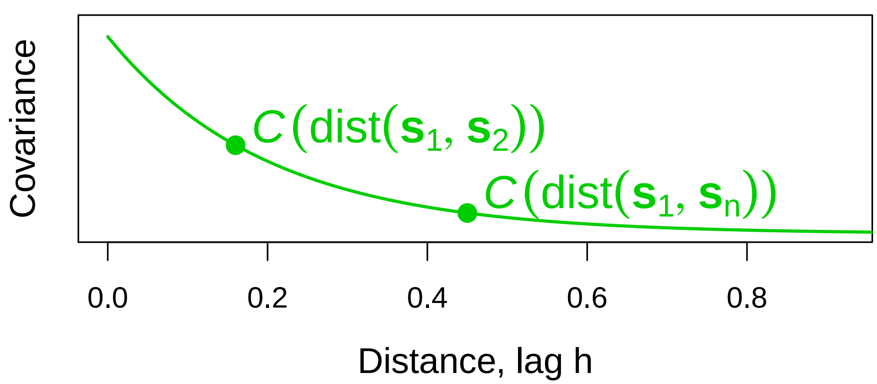
Covariance matrix: Σ

University of Zurich[UZH]

Spatial modeling

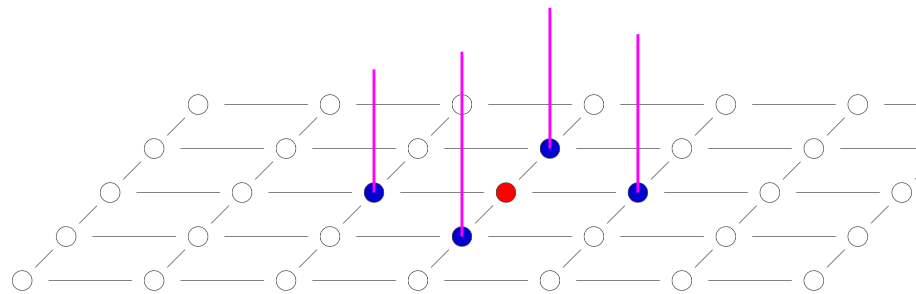
Geostatistical model (GRF):





Covariance matrix: $\mathbf{\Sigma}$

Lattice model (GMRF):



$$\text{E}[y_i|y_{-i}] = \beta \sum_{j \text{ neighbor of } i} y_j$$

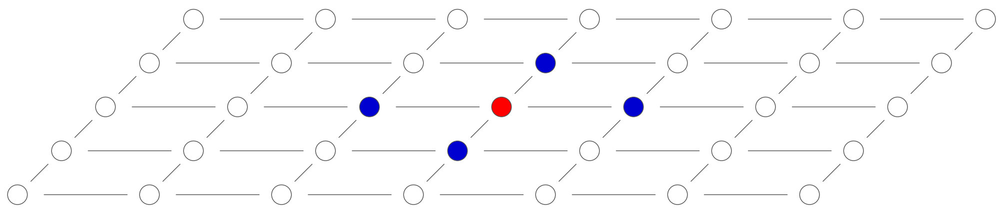
$$\text{Var}[y_i|y_{-i}] = \tau^2$$

Gaussianity and regularity conditions:

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

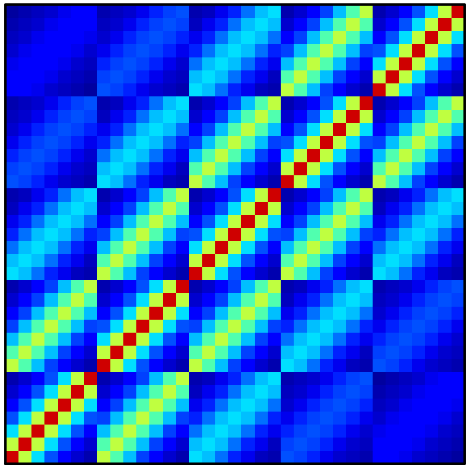
University of Zurich[UZH]

Spatial modeling



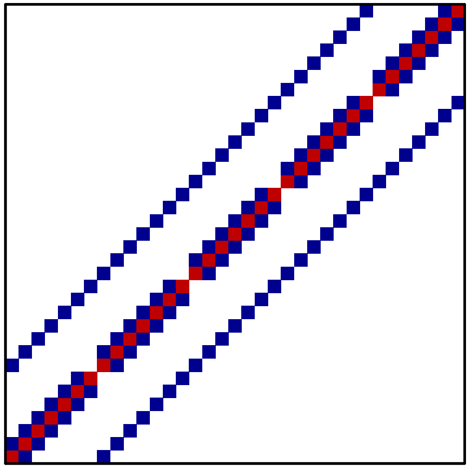
Geostatistical model (GRF):

Σ

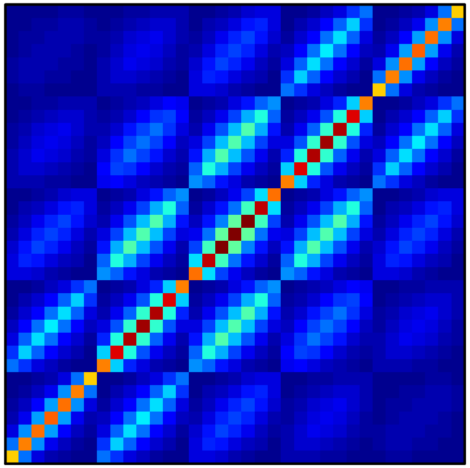


Lattice model (GMRF):

Σ^{-1}



Σ



University of Zurich[UZH]

Spatial model

Spatial, additive mixed effects model for measurements:

$$
\begin{aligned}
\text{data} &= \text{signal} + \text{noise} \\
&= \text{fixed effects} + \text{trend}(\mathbf{s}) + \text{spatial term}(\mathbf{s}) + \text{error}
\end{aligned}
$$

$$
Y(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta} + Z(\mathbf{s}) + \varepsilon(\mathbf{s}) \qquad \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d,\ d \geq 1
$$

with

$\mathbf{X}\boldsymbol{\beta}$: fixed effects and trend

$Z(\mathbf{s})$: zero mean spatial Gaussian process

$\varepsilon(\mathbf{s})$: iid noise, orthogonal to $Z(\mathbf{s})$

Spatial model

Spatial, additive mixed effects model for measurements:

$$\begin{aligned} \text{data} &= \text{signal} + \text{noise} \\ &= \text{fixed effects} + \text{trend}(\mathbf{s}) + \text{spatial term}(\mathbf{s}) + \text{error} \end{aligned}$$

$$Y(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta} + Z(\mathbf{s}) + \varepsilon(\mathbf{s}) \qquad \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d,\ d \geq 1$$

with

$\mathbf{X}\boldsymbol{\beta}$: fixed effects and trend
coefficients $\boldsymbol{\beta}$

$Z(\mathbf{s})$: zero mean spatial Gaussian process
parameters $\boldsymbol{\theta}_Z$ describing the covariance function

$\varepsilon(\mathbf{s})$: i.i.d. noise, orthogonal to $Z(\mathbf{s})$
variance σ^2

Backfitting

Spatial, additive mixed effects model for measurements:

$$
\begin{aligned}
\text{data} &= \text{signal} + \text{noise} \\
&= \text{fixed effects} + \text{trend}(\mathbf{s}) + \text{spatial term}(\mathbf{s}) + \text{error}
\end{aligned}
$$

$$
Y(\mathbf{s}) = \mathbf{X}\beta + Z(\mathbf{s}) + \varepsilon(\mathbf{s}) \qquad \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d,\ d \geq 1
$$

Extending the 'classical' backfitting approach to dependent data:

```
repeat until convergence
        estimate fixed effects
        estimate parameters
        predict smooth field
```

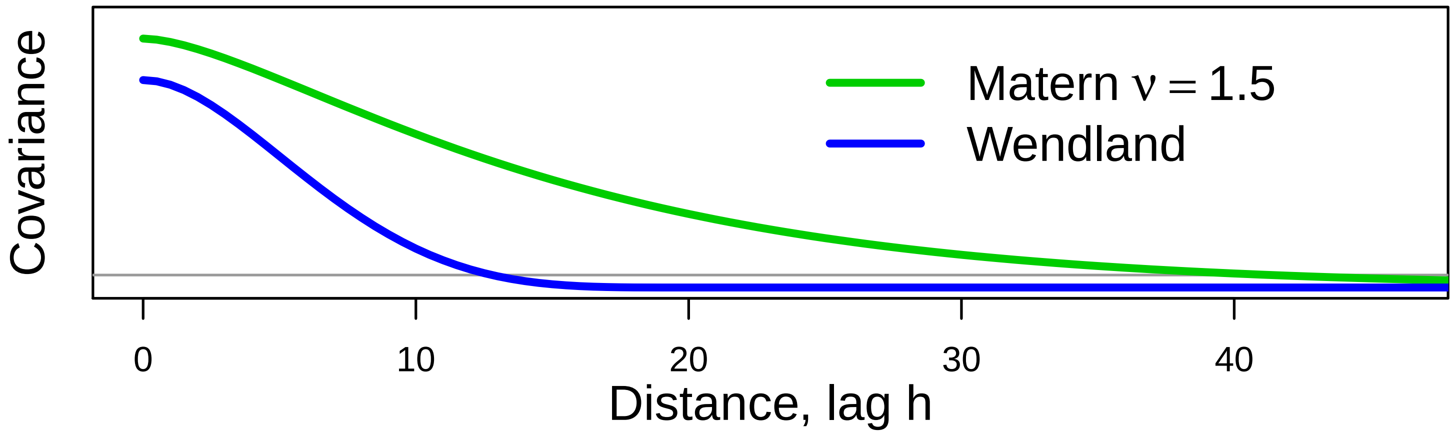
See Furrer, Sain (2009) Heersink, Furrer (2012|3)

Tapering: 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 $\rightsquigarrow$ tapering

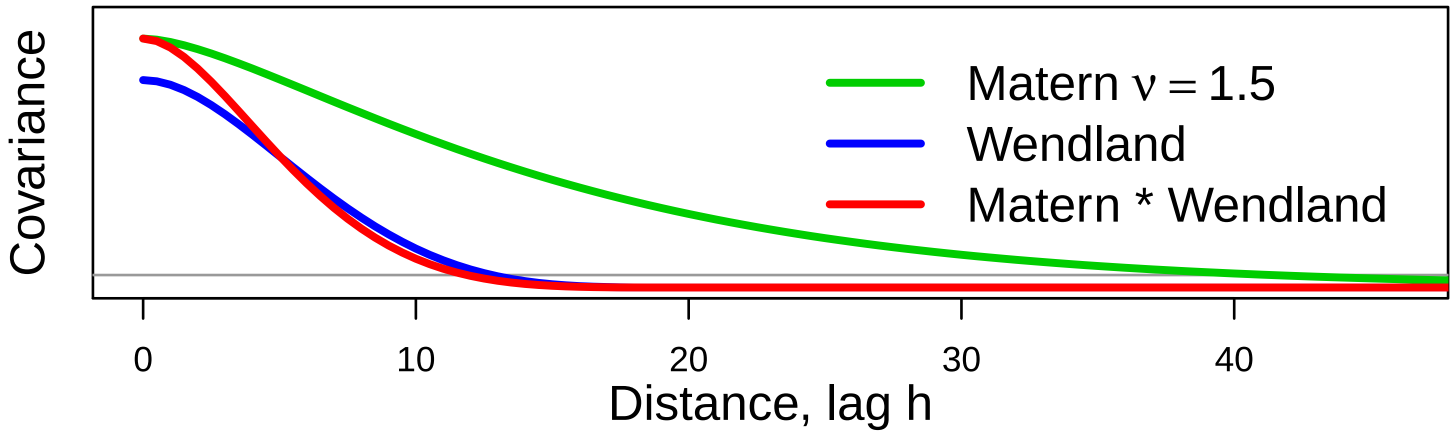


Tapering: sparseness

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 $\rightsquigarrow$ tapering



University of Zurich[UZH]

Tapering: software

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

- an R package for sparse matrix algebra
- tailored for MCMC calculations within G(M)RF
- storage economical and fast
  Fortran based;
  block sparse Cholesky algorithm of Ng and Peyton (1993)
- versatile, intuitive and simple
  transparent; supports (essentially) one sparse matrix format

See Furrer, Sain (2010) JSS; Gerber Furrer (2015) JSS. . .

“All models are wrong, but . . .”

- Iterative approaches
  - \+ Flexible, numerically feasible
  - − Uncertainties
- Maximum likelihood
  - \+ Uncertainties, asymptotics
  - − Numerical issues
- Bayesian hierarchical models
  - \+ Flexible, uncertainties
  - − MCMC
- SPDE models
  - \+ flexible, scalable
  - − interpretability

Quo vadis

(A) Conceptual modeling approach:

1. "i.i.d." $\rightsquigarrow$ space $\rightsquigarrow$ space-time models

(B) Statistical community should provide:

1. Understandable models
2. Flexible models
3. Scalable models
4. Reproducable implementation

Collaboration with:

— Former and present 'Applied Statistics' team

— Steve Sain The Climate Corporation, Doug Nychka NCAR

— Francois Bachoc U Vienna, Juan Du K-State, . . . and many more

URPP Global Change and Biodiversity

Swiss National Science Foundation 129782, 143282, 144973

University of Zurich[UZH]

References (some)

de Jong Schaepman Furrer de Bruin Verburget (2013) Spatial relationship between climatologies and changes in global vegetation activity *GlobChBio* **19** 1953-1964

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

Furrer Bachoc Du (2015) Tapering for multivariate spatial fields *JMVA*, under revision

Furrer Genton Nychka (2006) Covariance tapering for interpolation of large spatial datasets *JGCS* **15** 502-523

Furrer Sain (2009) Spatial model fitting for large datasets with applications to climate and microarray problems *Statistics and Computing* **19** 113-128

Heersink Furrer (2012) Moore-Penrose inverses of quasi-Kronecker structured matrices, *Linear Algebra and its Applications* **436** 561-570

Heersink Furrer (2013) Sequential spatial analysis of large datasets with applications to modern soil compaction roller measurement values *Spatial Statistics* **6** 41-56

Kaufman Schervish Nychka (2008) Covariance tapering for likelihood-based estimation in large spatial data sets *JASA* **103** 1545-1555

Stein (2013) Statistical properties of covariance tapers *JCGS* **22** 866-885