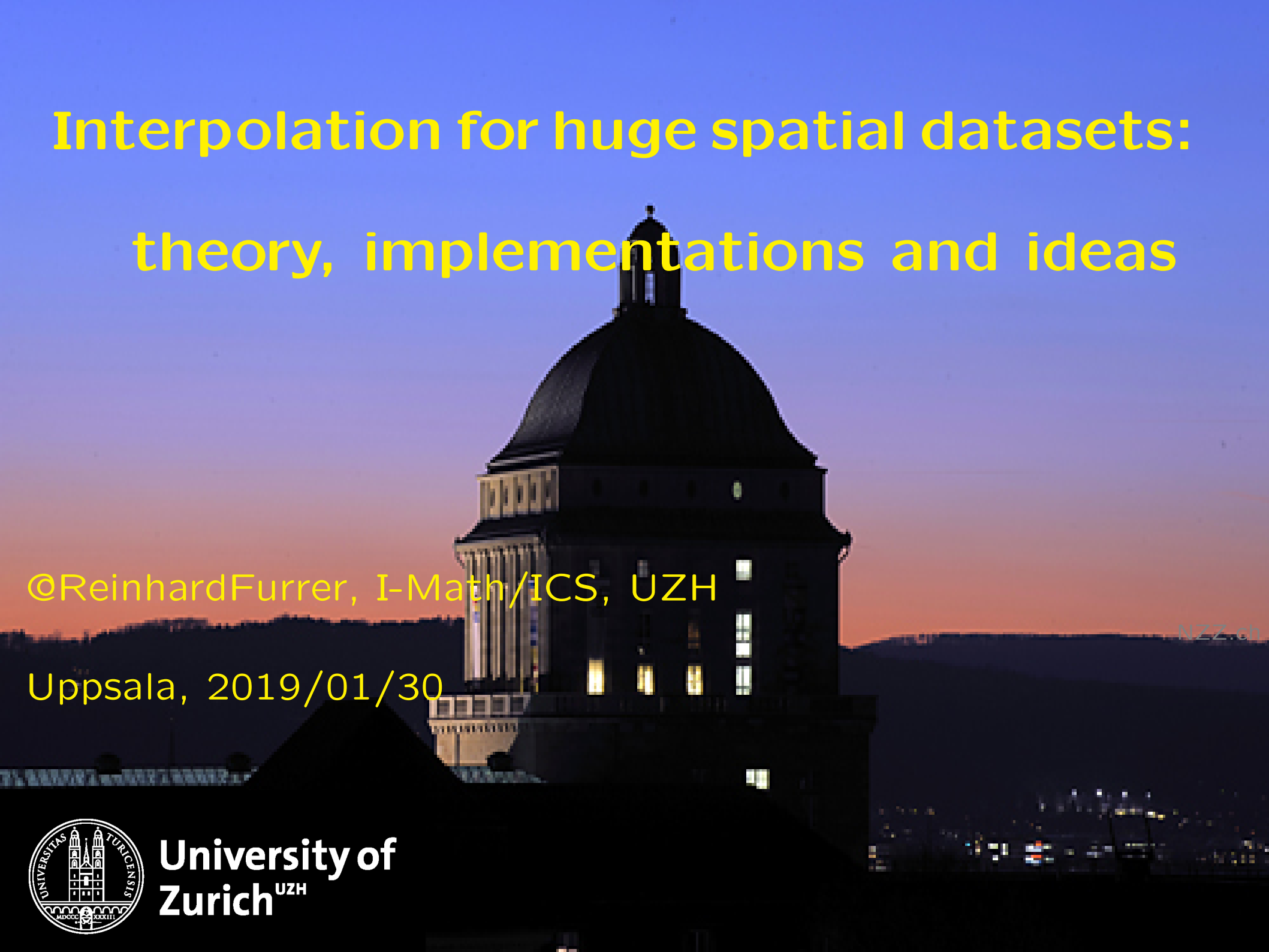
Interpolation for huge spatial datasets: theory, implementations and ideas

@ReinhardFurrer, I-Math/ICS, UZH

Uppsala, 2019/01/30



University of Zurich[UZH]

Joint work

- Emilio Porcu
- Florian Gerber
- Francois Bachoc

and with contributions of several others

University of Zurich[UZH]

Outlook

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

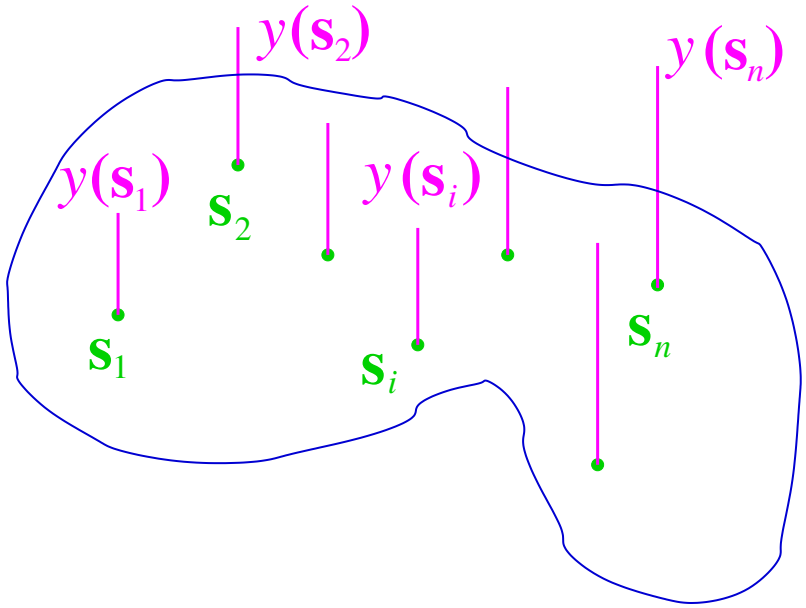
$\rightsquigarrow$ Questions and $\ll$Fast-Forward$\gg$ appreciated!

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

Spatial statistics: prediction

Observations: $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$

Source: wikipedia.org

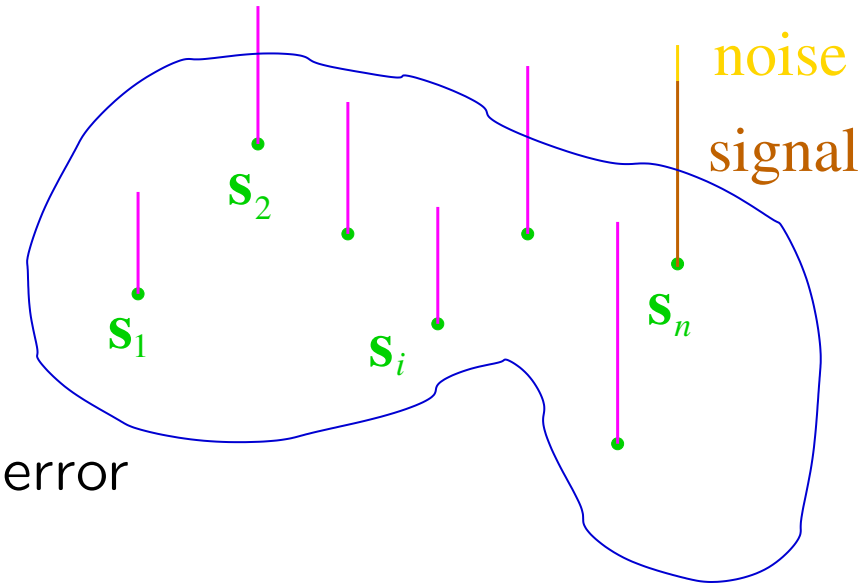


First law of geography (Waldo Tobler):

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

Spatial statistics: prediction

Predict the quantity of interest at an arbitrary location.



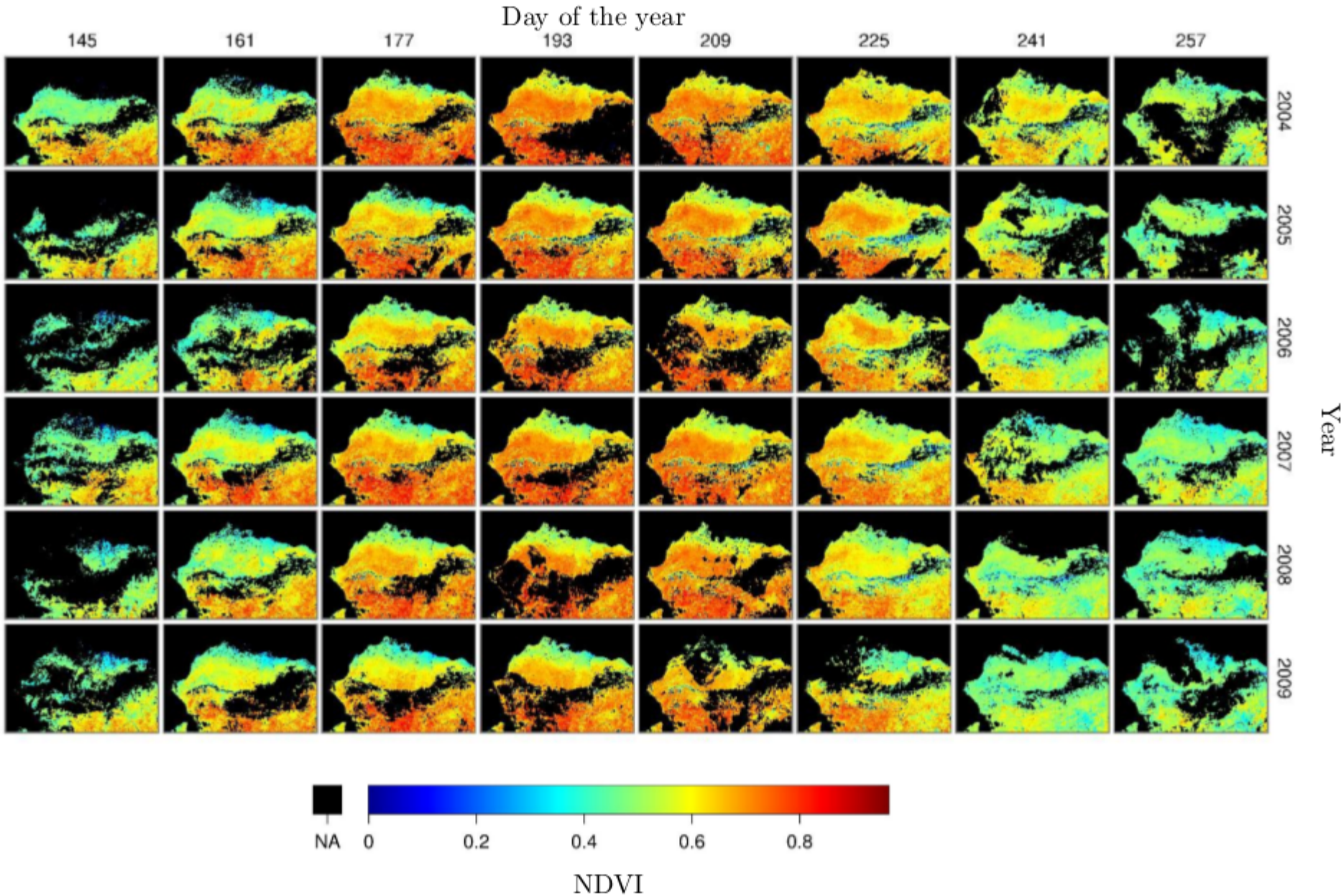
Why?

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

How?

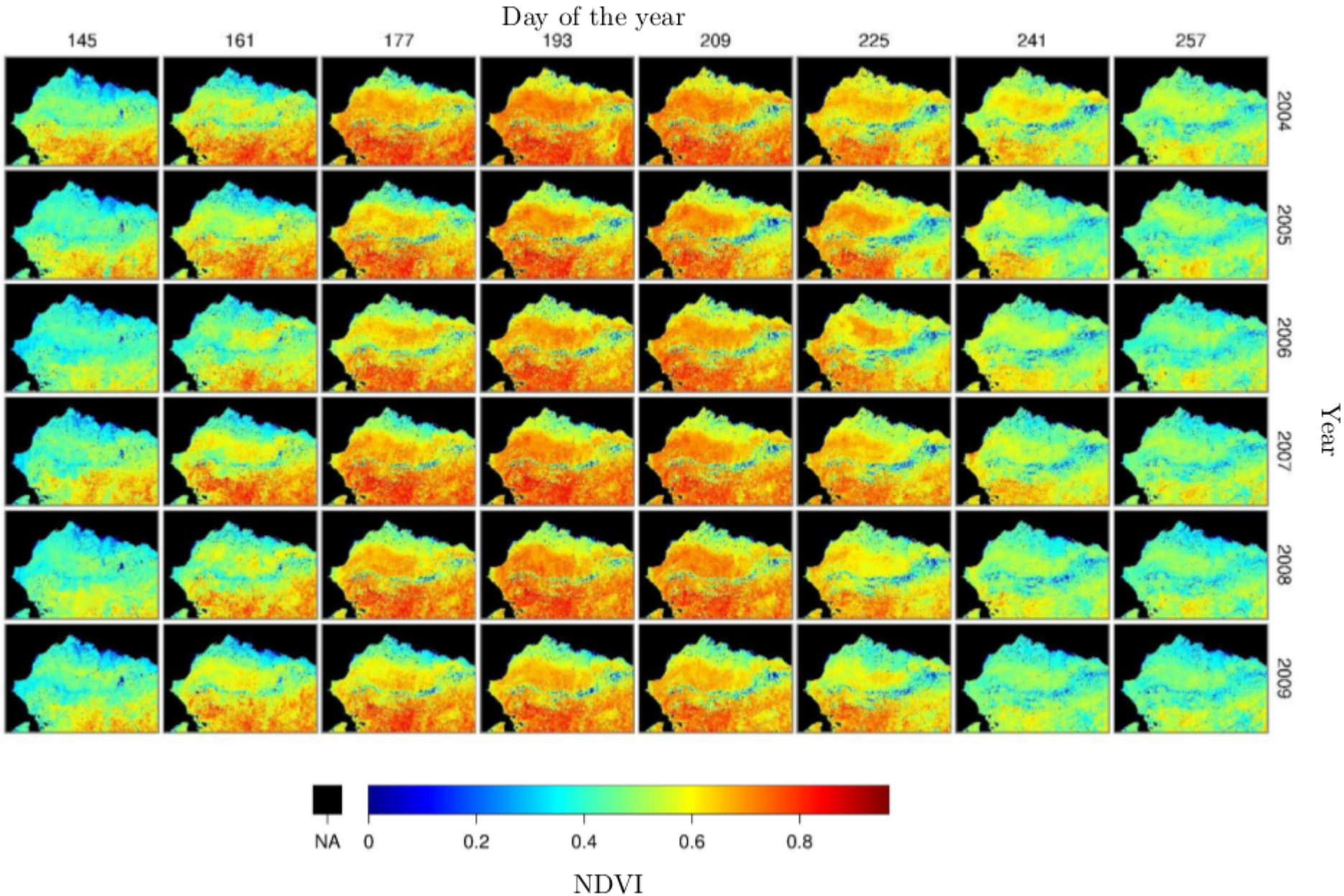
- By eye
- Linear interpolation
- The correct way . . .

Visual example



Gerber et al (2018), TGRS

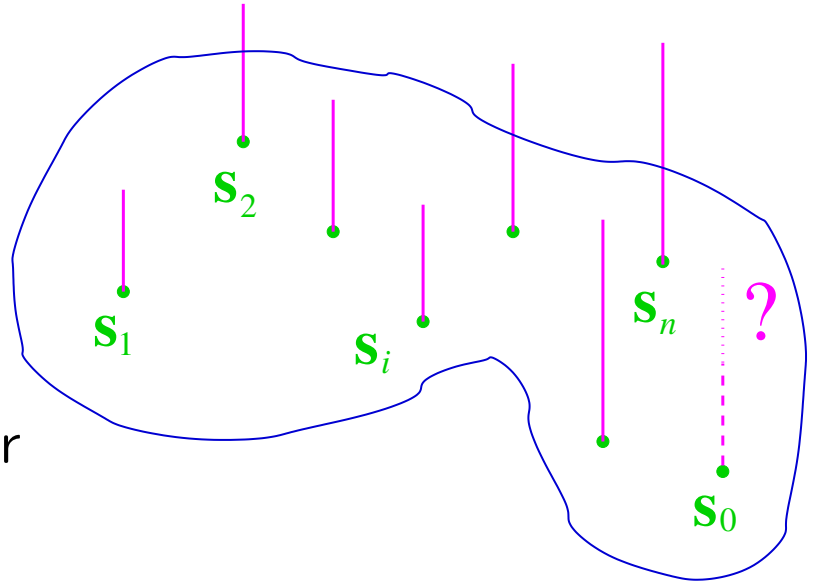
Visual example



Gerber et al (2018), TGRS

Spatial statistics: prediction

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



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

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

$\rightsquigarrow$ Best Linear Unbiased Predictor:

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

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

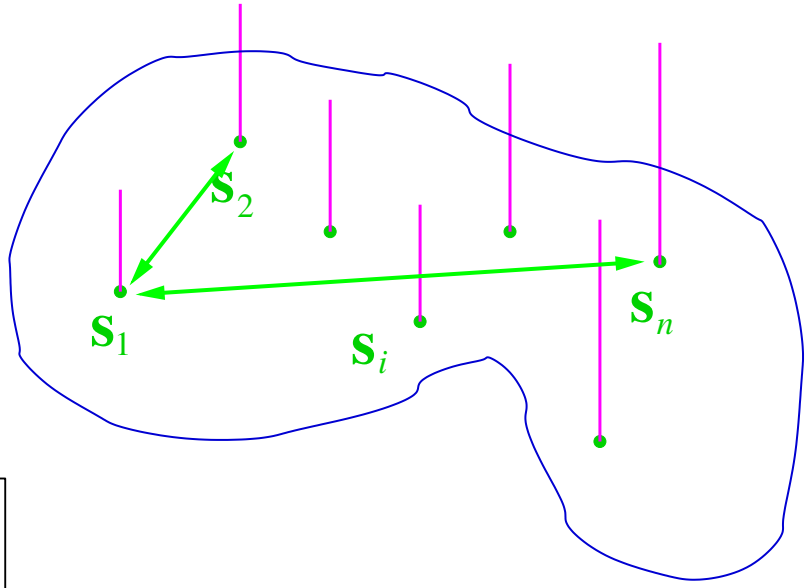
(one spatial process, no trend, known covariance structure;
otherwise *almost* the same)

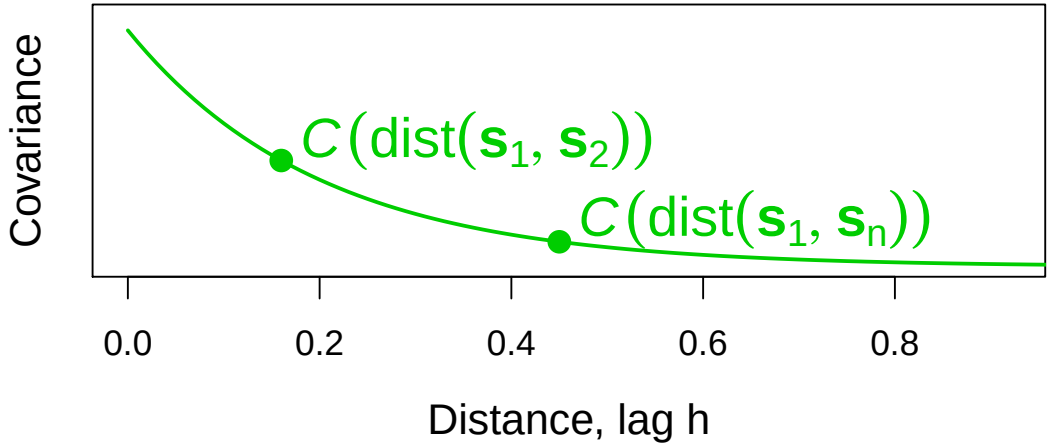
University of Zurich[UZH]

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 matrix Σ contains elements $C\Big(\text{dist}(\mathbf{s}_i, \mathbf{s}_j)\Big)$.

Issues of basic, classical kriging

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

- "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 . . .

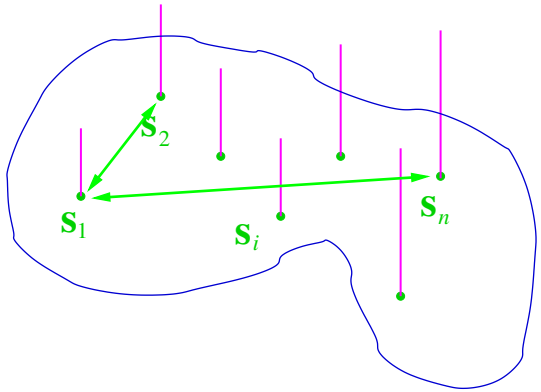
Heaton et al. JABES

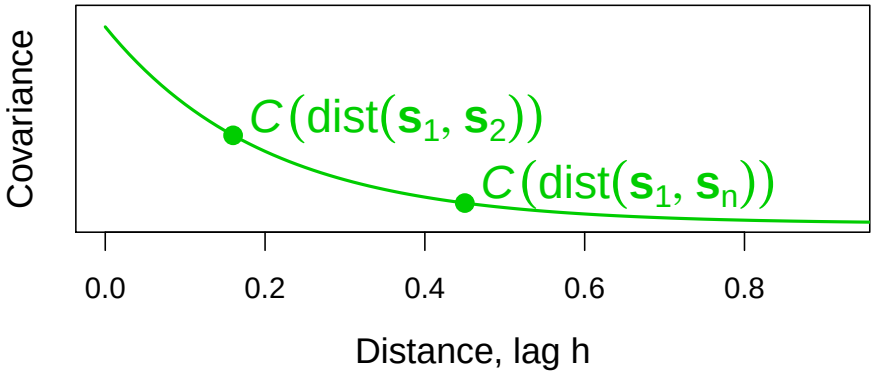
Outlook

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

Spatial modeling

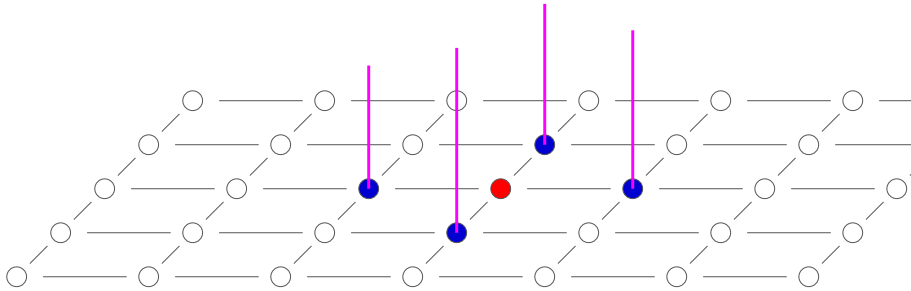
Geostatistical model (GRF):





Covariance matrix: $\boldsymbol{\Sigma}$

Lattice model (GMRF):



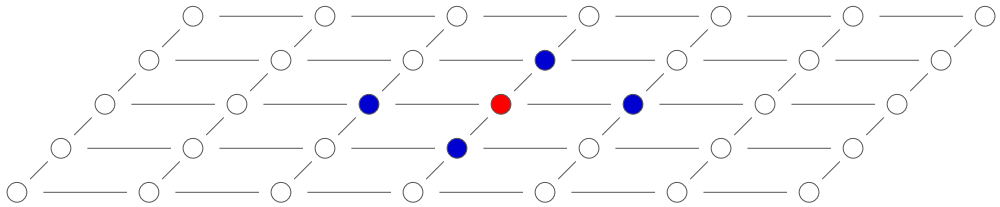
$$\mathrm{E}(Z_i|z_{-i}) = \beta \sum_{j \text{ neighbor of } i} z_j$$

$$\mathrm{Var}(Z_i|z_{-i}) = \tau^2$$

Gaussianity and regularity conditions:

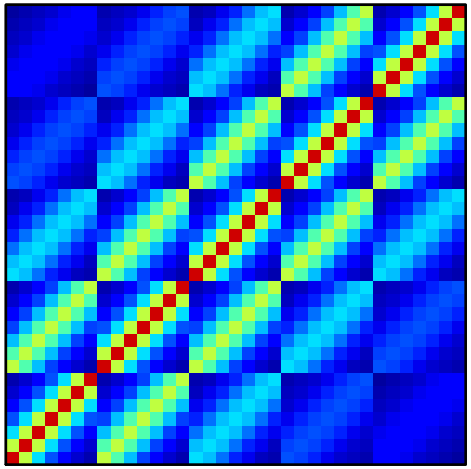
$$\boldsymbol{\Sigma} = \tau^2(\mathbf{I} - \mathbf{B})^{-1}$$

Spatial modeling

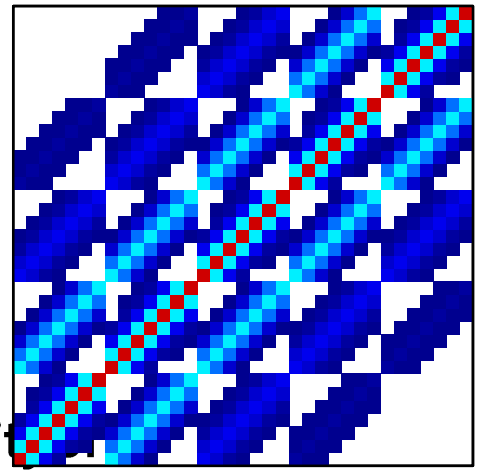


Geostatistical model (GRF):

Σ

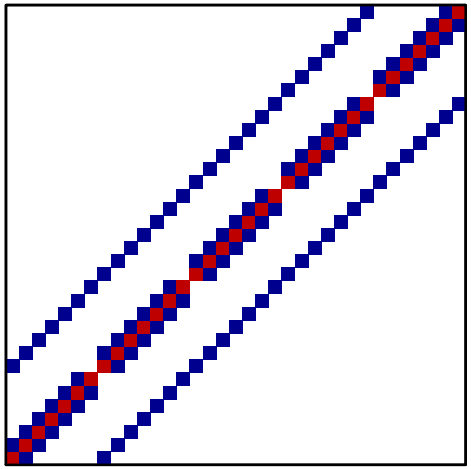


Σ_{app}

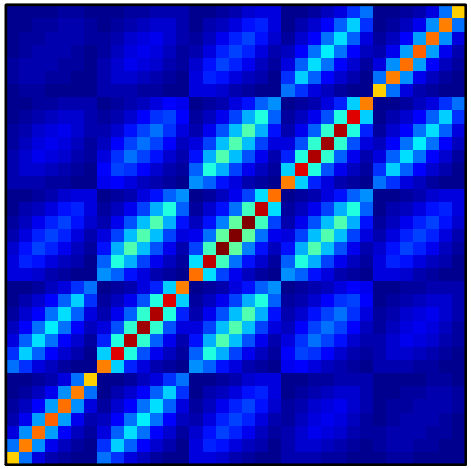


Lattice model (GMRF):

Σ^{-1}



Σ

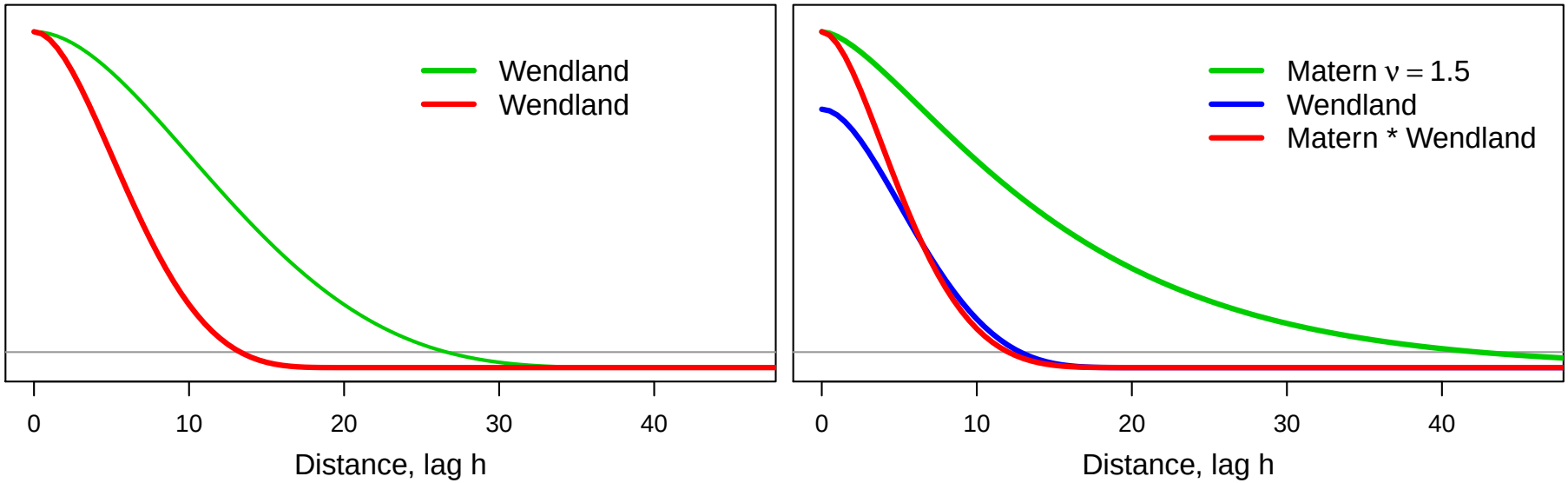


Sparseness

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: 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.

$$\frac{\mathrm{MSPE}(\mathbf{s}_0, C_0 C_\rho)}{\mathrm{MSPE}(\mathbf{s}_0, C_0)} = \frac{\mathrm{MSPE}_{\mathrm{taper}}}{\mathrm{MSPE}_{\mathrm{true}}} \to 1$$

$$\frac{\varrho(\mathbf{s}_0, C_0 C_\rho)}{\mathrm{MSPE}(\mathbf{s}_0, C_0)} = \frac{\mathrm{MSPE}_{\mathrm{naive}}}{\mathrm{MSPE}_{\mathrm{true}}} \to 1 \qquad \varrho(\mathbf{s}_0, C) = C(0) - \mathbf{c}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{c}$$

University of Zurich[UZH]

Tapering: asymptotic optimality

For an isotropic and stationary process with covariance $C_0(h)$,

find $C_1(h)$, such that kriging with $C_1(h)$

is asymptotically optimal.

$$\frac{\text{MSPE}(\mathbf{s}_0, C_0 C_\rho)}{\text{MSPE}(\mathbf{s}_0, C_0)} = \frac{\text{MSPE}_{\text{taper}}}{\text{MSPE}_{\text{true}}} \to 1$$

$$\frac{\varrho(\mathbf{s}_0, C_0 C_\rho)}{\text{MSPE}(\mathbf{s}_0, C_0)} = \frac{\text{MSPE}_{\text{naive}}}{\text{MSPE}_{\text{true}}} \to 1 \qquad \varrho(\mathbf{s}_0, C) = C(0) - \mathbf{c}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{c}$$

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
Bevilacqua et al (2019) AoS

Covariance functions

Matérn:

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

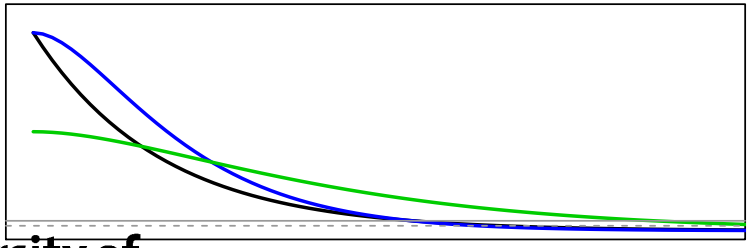
$$\mathcal{M}_{\nu,\alpha,\sigma^2}(r) = \sigma^2 \mathcal{M}_\nu(r/\alpha) \qquad \alpha > 0 \quad \sigma^2 > 0$$

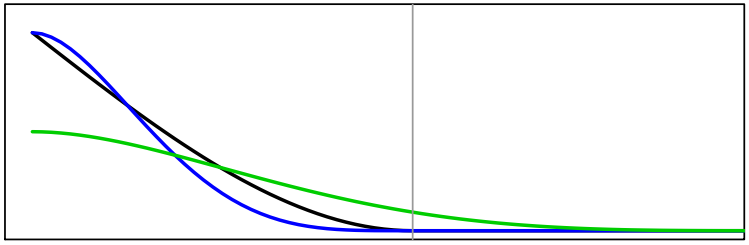
Generalized Wendland:

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

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

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





Outlook

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

Equivalence of Gaussian measures

$\mathbb{P}_0$, $\mathbb{P}_1$ two probability measures on a measureable space $\{\Omega, \mathcal{S}\}$.

Definition. $\mathbb{P}_0$, $\mathbb{P}_1$ are equivalent if $\mathbb{P}_0(A) = 1$, for any $A \in \mathcal{S}$ implies $\mathbb{P}_1(A) = 1$ and vice versa.

- Restrict A to the σ-algebra generated by $\{Z(\mathbf{s}), \mathbf{s} \in D\}$

  Here: '$\equiv$' equivalent on the paths of $\{Z(\mathbf{s}),\ \mathbf{s} \in D\}$.

- For Gaussian measures:

  equivalence or orthogonality

  first two moments discriminate: $\mathbb{P}(\mu, C)$

- 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 $C_0^{\mathcal{F}}$, $C_1^{\mathcal{F}}$.

$D \subset \mathbb{R}^d$ any bounded infinite set, $\mathbf{s}_0 \in D$.

Theorem. (Stein 2004)

If $\exists a > 0$, $\quad 0 < b_\ell \leq C_0^{\mathcal{F}}(z)\, z^a \leq b_u < \infty, \qquad z \to \infty$

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

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

University of Zurich[UZH]

Fourier transforms

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

$$\mathcal{W}_{\mu,\kappa}^{\mathcal{F}}(z) = \text{cte} \cdot {}_1F_2\Big(\lambda; \lambda + \frac{\mu}{2}, \lambda + \frac{\mu}{2} + \frac{1}{2}; -\frac{z^2}{4}\Big), \quad \lambda = (d+1)/2 + \kappa$$

$$\asymp z^{-2\lambda} \qquad \text{for} \qquad z \to \infty,$$

Equivalence of Gaussian measures

Theorem. (Zhang 2004)

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

$$\mathbb{P}(\mathcal{M}_{\nu,\alpha_0,\sigma_0^2}) \equiv \mathbb{P}(\mathcal{M}_{\nu,\alpha_1,\sigma_1^2}) \quad \Longleftrightarrow \quad \sigma_0^2 \big/ \alpha_0^{2\nu} = \sigma_1^2 \big/ \alpha_1^{2\nu}$$

- $d \leq 3$: not all parameters consistently estimable microergodic parameter $\sigma^2 \big/ \alpha^{2\nu}$
- $d \geq 5$ all fine (Anderes 2010)
- $d = 4$ open ...

Equivalence of Gaussian measures

Theorem. (BFFP 2019)

Let $\mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta_0,\sigma_0^2})$, $\mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta_1,\sigma_1^2})$.
For a given $\kappa \geq 0$, let $\mu > d + 1/2 + \kappa$:

$$\mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta_0,\sigma_0^2}) \equiv \mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta_1,\sigma_1^2}) \quad \Longleftrightarrow \quad \sigma_0^2 \big/ \beta_0^{2\kappa+1} = \sigma_1^2 \big/ \beta_1^{2\kappa+1}$$

Theorem. (BFFP 2019)

Let $\mathbb{P}(\mathcal{M}_{\nu,\alpha,\sigma_0^2})$ and $\mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta,\sigma_1^2})$, $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, $\kappa > 0$.
If $\nu = \kappa + 1/2$, $\mu > d + 1/2 + \kappa$:

$$\mathbb{P}(\mathcal{M}_{\nu,\alpha,\sigma_0^2}) \equiv \mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta,\sigma_1^2}) \quad \Longleftrightarrow \quad \sigma_0^2 \big/ \alpha^{2\nu} = \mathrm{cte}_{\nu,\kappa,\mu} \cdot \sigma_1^2 \big/ \beta^{2\kappa+1}$$

Outlook

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

Prediction: asymptotic optimality

Prediction of $Z(\mathbf{s}_0)$ using "BLUP" under misspecification $\widehat{Z}_n(\mu, \kappa, \beta)$.

Theorem. (BFFP 2019)

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

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta,\sigma^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)\Big)}{\mathrm{Var}_{\mu,\kappa,\beta,\sigma^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta) - Z(\boldsymbol{s}_0)\Big)} \longrightarrow 1$$

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

$$\frac{\mathrm{Var}_{\nu,\alpha,\sigma^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)\Big)}{\mathrm{Var}_{\nu,\alpha,\sigma^2}\Big(\widehat{Z}_n(\nu,\alpha) - Z(\boldsymbol{s}_0)\Big)} \longrightarrow 1$$

Prediction: presumed MSPE

Prediction of $Z(\mathbf{s}_0)$ using "BLUP" under misspecification $\widehat{Z}_n(\mu, \kappa, \beta)$.

Theorem. (BFFP 2019)

$\mathbb{P}(\mathcal{W}_{\mu,\kappa,\beta,\sigma^2})$: if $\sigma_0^2 \big/ \beta_0^{2\kappa+1} = \sigma_1^2 \big/ \beta_1^{2\kappa+1}$

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_1,\sigma_1^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)\Big)}{\mathrm{Var}_{\mu,\kappa,\beta,\sigma^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)\Big)} \longrightarrow 1$$

$\mathbb{P}(\mathcal{M}_{\nu,\alpha,\sigma^2})$: if $\nu = \kappa + 1/2,\ \sigma^2 \big/ \alpha^{2\nu} = \mathrm{cte} \cdot \sigma_1^2 \big/ \beta_1^{2\kappa+1}$

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_1,\sigma_1^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)\Big)}{\mathrm{Var}_{\nu,\alpha,\sigma^2}\Big(\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)\Big)} \longrightarrow 1$$

Estimation: tapering

Likelihood:

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

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

- Dimension plays crucial role Kaufman et al. (2008) JASA
- Two-taper approach cannot be used in practice

Estimation: Matérn

Let $\mathbb{P}(\mathcal{M}_{\nu,\alpha_0,\sigma_0^2})$, with certain constrains on the parameter space.

Theorem. (Shaby Kaufmann 2013)

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

1. $\hat{\sigma}_n^2 \big/ \hat{\alpha}_n^{2\nu} \xrightarrow{a.s.} \sigma_0^2 \big/ \alpha_0^{2\nu}$

2. $\sqrt{n}(\hat{\sigma}_n^2/\hat{\alpha}_n^{2\nu} - \sigma_0^2/\alpha_0^{2\nu}) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, 2(\sigma_0^2/\alpha_0^{2\nu})^2\Big)$

- Also holds for $\hat{\sigma}_n^2(\alpha)$ with misspecified α

Estimation: generalized Wendland

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

Theorem. (BFFP 2019)

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

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

2. $\sqrt{n}(\widehat{\sigma}_n^2(\beta)\big/\beta^{2\kappa+1} - \sigma_0^2\big/\beta_0^{2\kappa+1}) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, 2(\sigma_0^2/\beta_0^{2\kappa+1})^2\Big)$

University of Zurich[UZH]

Estimation: generalized Wendland

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

Theorem. (BFFP 2019)

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

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

2. $\sqrt{n}(\widehat{\sigma}_n^2(\widehat{\beta}_n) \big/ \widehat{\beta}_n^{2\kappa+1} - \sigma_0^2 \big/ \beta_0^{2\kappa+1}) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, 2(\sigma_0^2/\beta_0^{2\kappa+1})^2\Big)$

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 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

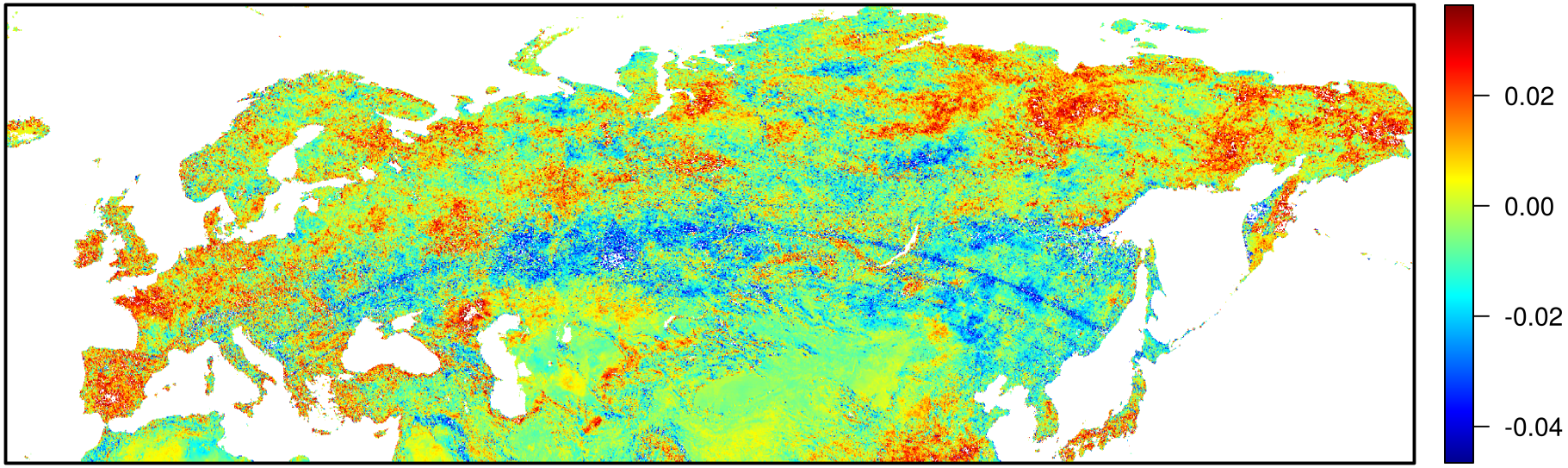
- R objects have at most 2^{31} elements (almost)
- R does not 'have' 64-bit integers: stored as doubles
- 64-bit exploitation consists of type conversions between front-end R and pre-compiled code

Gerber, Mösinger, Furrer (2017) CaGeo
Gerber, Mösinger, Furrer (2018) SoftwareX

Illustration

- residual field from AVHRR NDVI_{3g} product:

  $\mathbf{y} = \mathbf{y}_{2000-2009} - \mathbf{y}_{1990-1999}$, $n = 769,940$ observations



- Nonstationary Gaussian random field (zero mean)

Illustration: nonstationary covariance

$$\Sigma(\kappa, \boldsymbol{\beta}, \tau) = \kappa\, \mathbf{D}(\boldsymbol{\beta}) \big((1-\tau)\mathbf{I} + \tau \mathbf{R}\big)\, \mathbf{D}(\boldsymbol{\beta})$$

- $\kappa > 0$: scaling parameter
- $\mathbf{D}(\boldsymbol{\beta}) = \mathrm{diag}(\exp(\mathbf{X}\boldsymbol{\beta}))$: controls strength via covariates
- $\tau \in [0, 1]$: "no spatial correlation" vs "spatial correlation"
- $\mathbf{I}$: identity matrix
- $\mathbf{R}$: stationary correlation matrix:
  - compactly supported covariance
  - range 50km, sparsity 0.2%

Illustration: optimization

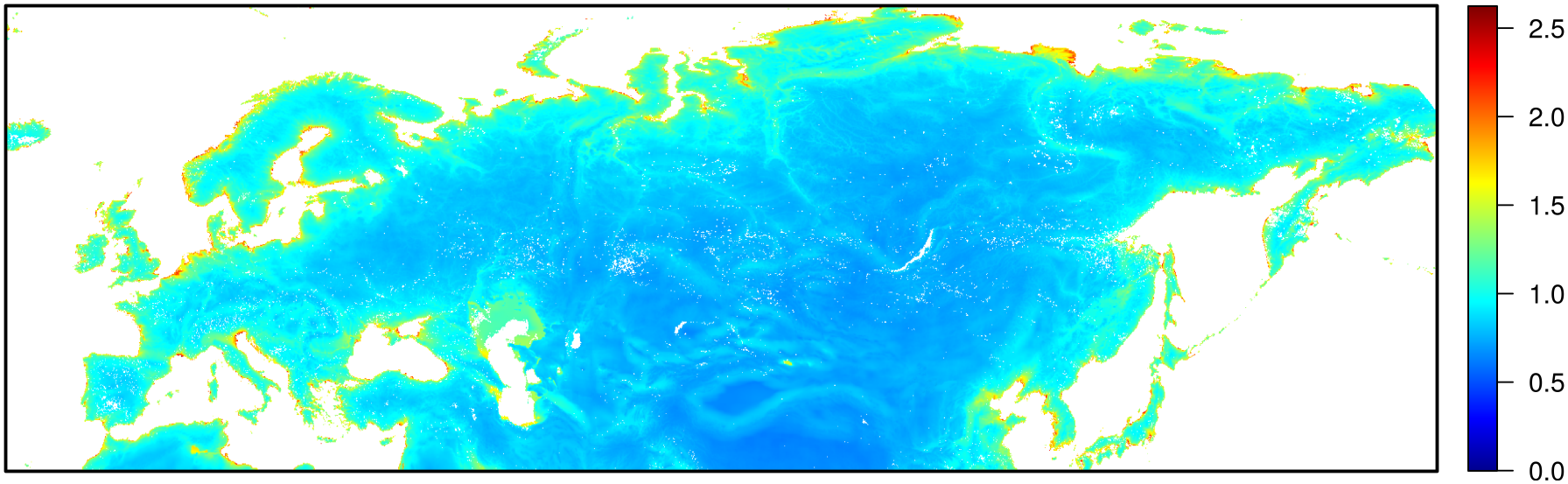
- Fast is relative . . . `optim` suboptimal

Task	Function	Time	Sparsity	
Distances	`h <- nearest.dist(...)`	23min	1.4 Gb	0.2‰
Covariances	`T <- cov.Wend(h, ...)`	2min	1.4 Gb	0.2‰
Cholesky	`chol(T, ...)`	29min	8.5 Gb	1.9‰ (64-bit)

- optimization strategy:
  - (iterative) grid search over τ
    exploit multicore architecture
  - for given τ use quasi-Newton optimizer to optimize $\kappa, \boldsymbol{\beta}$

Illustration: nonstationary covariance

- with covariates "distance to nearest coast" and "elevation"
- $\mathrm{diag}(\widehat{\Sigma})$:

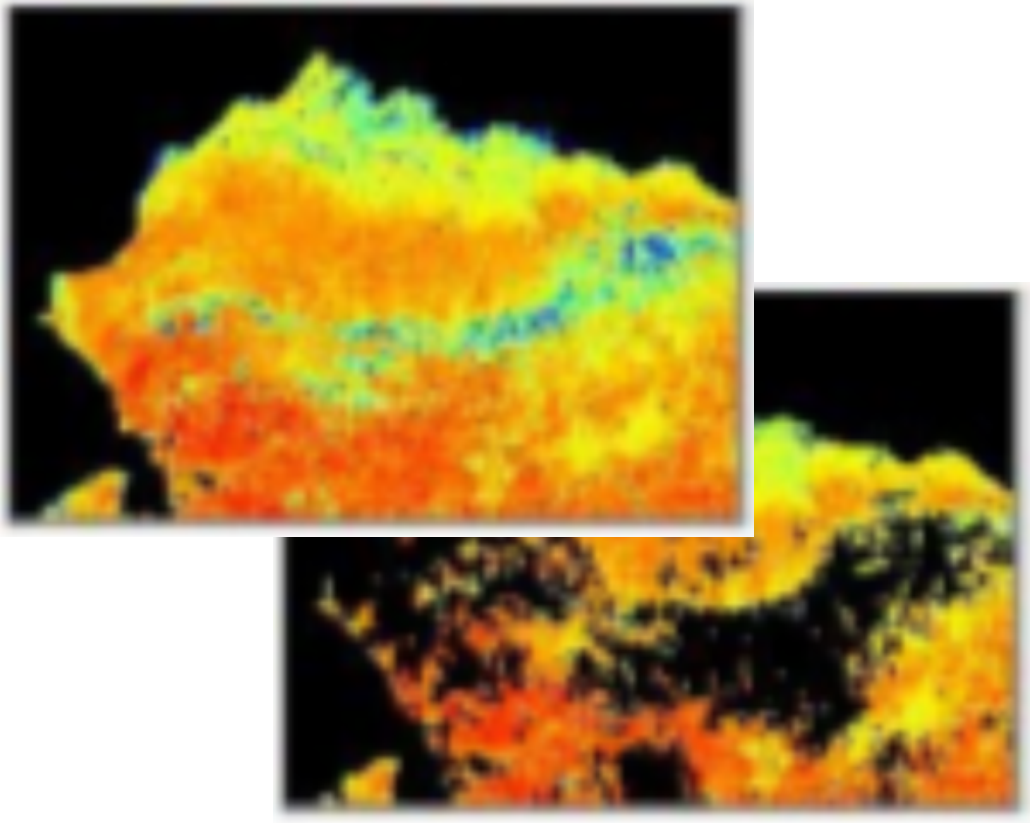


- BIC improvement compared to nonstationary model

. . . 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

URPP Global Change and Biodiversity

FNSNF Swiss National Science Foundation 143282, 144973, 175529

University of Zurich[UZH]