University of Zurich[UZH]

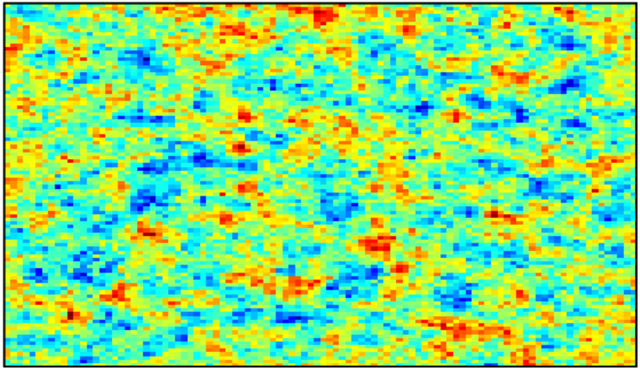
Identification of Dominant Features in Spatial Data

Research in progress talk

Roman Flury & Prof. Dr. Reinhard Furrer

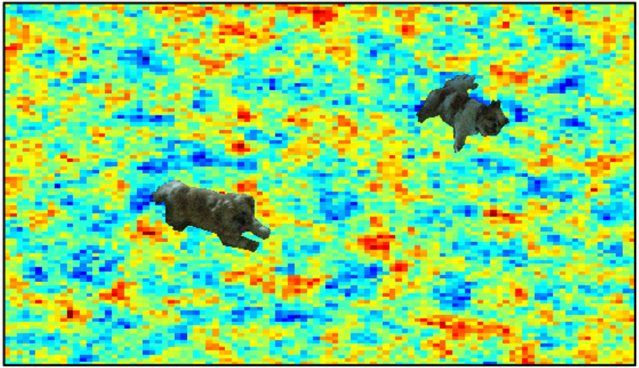
Zurich, 20th of October 2020

Spatial Data



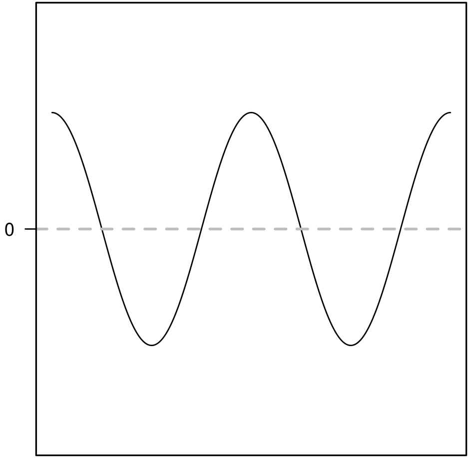
Spatial data.

(Dominant) Features?



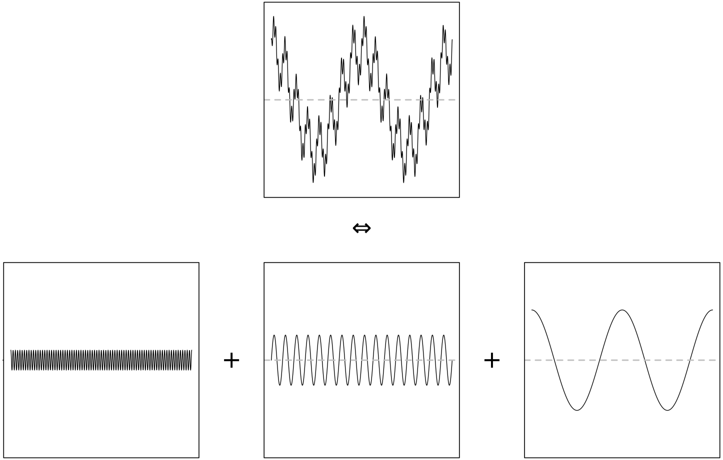
Dogs in spatial data.

(Dominant) Features



Time-series/transect data.

(Dominant) Features

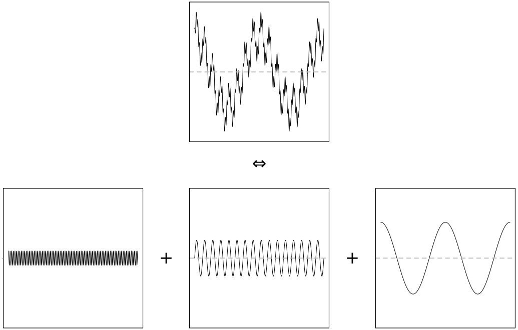


Time-series/transect data.

(Dominant) Features

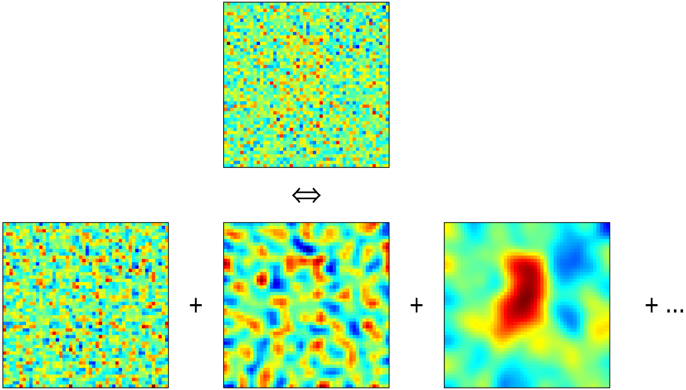
Simplifications

- time-series/transect data
- independent process
- regular/harmonic process
- no observational noise



Time-series/transect data.

Dominant Features



Multiresolution decomposition of spatial data.

Observational Noise

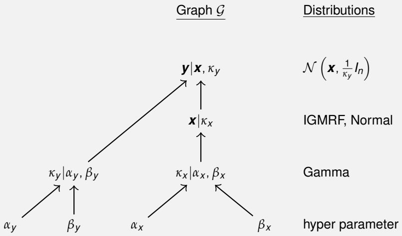
let

- $\boldsymbol{y} \in \mathbb{R}^n$ be the observed spatial data, where n is the number of locations
- $\boldsymbol{x} \in \mathbb{R}^n$ be the true underlying data
- $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ observational/measurement noise which follows a $\mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_n\right)$ distribution

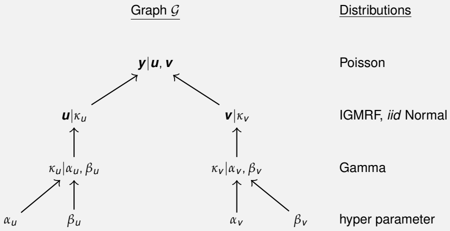
then we assume

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}$$

Bayesian Model



Normal-Gamma model, used in Flury et al. [2020]



Besag–York–Mollié model, used in Flury and Furrer [2020]

Spatial Multiresolution Decomposition

Introduced by Holmström et al. [2011]:

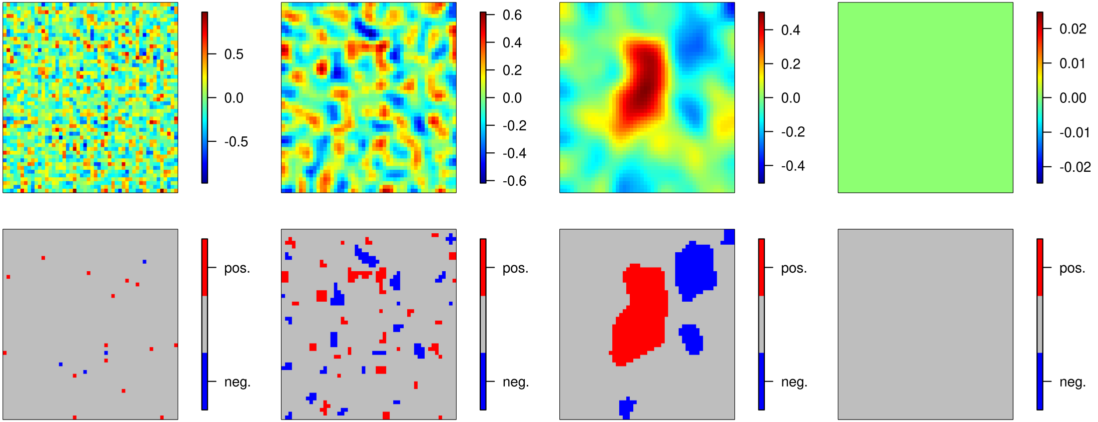
- smooth the spatial data $\boldsymbol{x}$ on different scales using the smoother $\boldsymbol{S}_\lambda = (\boldsymbol{I}_n + \lambda \boldsymbol{Q})^{-1}$
  - $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is the precision matrix modeling dependencies between locations of $\boldsymbol{x}$
  - $\lambda \in \mathbb{R}^+$ the smoothing scale, such that $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_L = \infty$
- decompose $\boldsymbol{x}$ as sum of consecutive differences of smooths

$$\boldsymbol{x} = \boldsymbol{S}_{\lambda_1}\boldsymbol{x} - \boldsymbol{S}_{\lambda_2}\boldsymbol{x} + \boldsymbol{S}_{\lambda_2}\boldsymbol{x} - + \ldots - \boldsymbol{S}_{\lambda_L}\boldsymbol{x} + \boldsymbol{S}_{\lambda_L}\boldsymbol{x}$$

$$\boldsymbol{x} = \sum_{i=1}^{L-1} \left(\boldsymbol{S}_{\lambda_i}\boldsymbol{x} - \boldsymbol{S}_{\lambda_{i+1}}\boldsymbol{x} \right) + \boldsymbol{S}_{\lambda_L}\boldsymbol{x}$$

- optimal smoothing scales are chosen, such that the differences between smooths is maximal

Dominant Features



Complete multiresolution decomposition of spatial data.

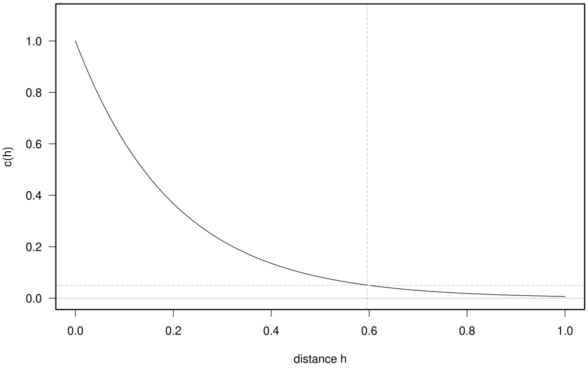
Scale-Dependent Feature Width-Extents

- assume an intrinsic stationary spatial process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$, where $\mathcal{D} \subset \mathbb{R}^2$, i.e.

$$\mathbb{E}(Z(\boldsymbol{s_1})) \equiv \mu,$$
$$\text{Cov}(Z(\boldsymbol{s_1}),\ Z(\boldsymbol{s_2})) = \text{c}(\boldsymbol{s}_1 - \boldsymbol{s}_2; \boldsymbol{\theta})$$

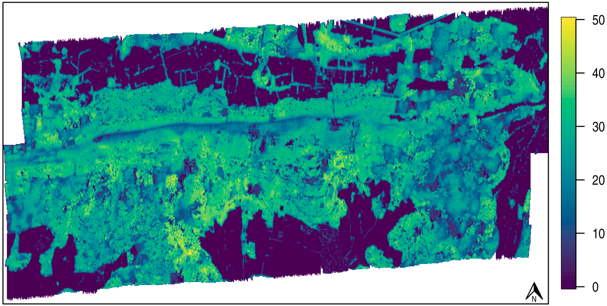
  for all locations $\boldsymbol{s_1}, \boldsymbol{s_2} \in \mathcal{D}$ and $\boldsymbol{\theta}$ a vector of fix parameters such that $c(\cdot)$ a positive (semi-) definite function
- assume a suitable covariance function (Matérn)
- estimate $\hat{\theta}_{\text{ML}}$ for e.g. $Z \sim \mathcal{N}_n(\mu\mathbf{1}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$
- assess the scale-dependent feature width-extents with the spatial data-driven **effective-range** parameter

Covariance Functions



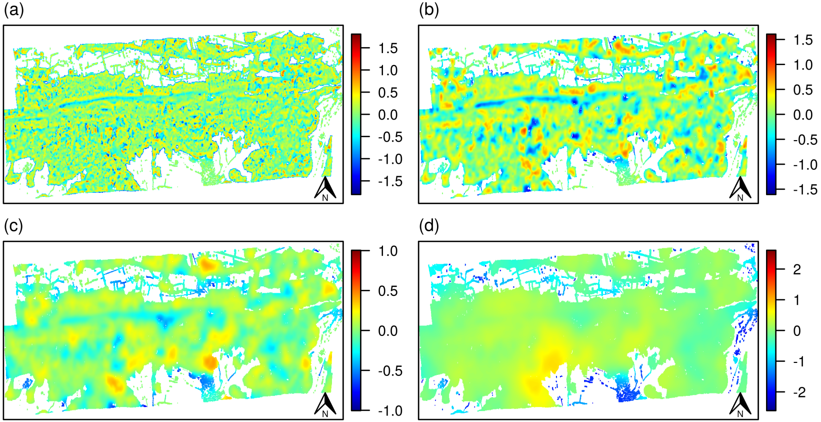
Matérn covariance function $c(h; \hat{\theta}_{\text{ML}})$.

Identification of Dominant Features in Gridded Spatial Data



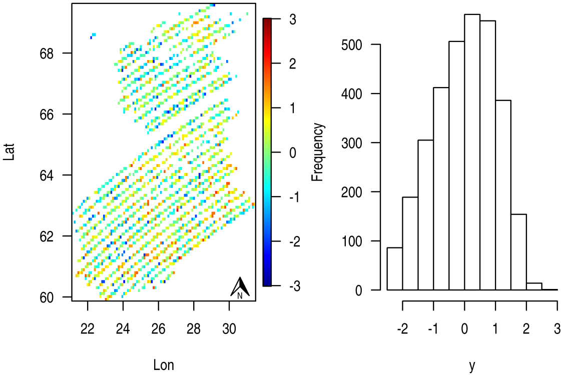
Canopy heights of the mountain Laegeren forest.

Identification of Dominant Features in Gridded Spatial Data



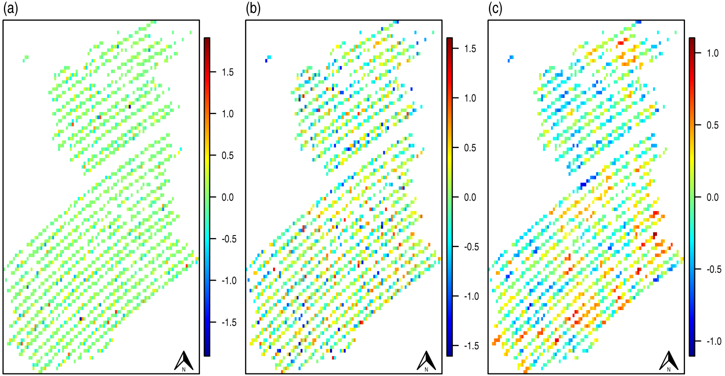
Scale-dependent details for the morphological trait *canopy heights* of the mountain Laegeren forest.

Identification of Dominant Features in **Non**-Gridded Spatial Data



Basal areas of pines across Finland.

Identification of Dominant Features in **Non**-Gridded Spatial Data



Scale-dependent details for basal areas of pines across Finland.

Discussion

- is it possible to (statistically) validate that the scales/dominant features are really there?
- are identified dominant features consistent when using approximation techniques like covariance tapering?

References

R. Flury and R. Furrer. Multiresolution decomposition of areal count data, 2020.

R. Flury, F. Gerber, B. Schmid, and R. Furrer. Identification of dominant features in spatial data, 2020.

L. Holmström, L. Pasanen, R. Furrer, and S. R. Sain. Scale space multiresolution analysis of random signals. *Computational Statistics & Data Analysis*, 55(10):2840–2855, 2011. ISSN 0167-9473. doi: http://doi.org/10.1016/j.csda.2011.04.011.