Nonstationary Spatial Process Models with Spatially Varying Covariance Kernels

Pourquoi faire simple quand on peut faire compliqué ?

Sébastien Coube-Sisqueille[1], **Sudipto Banerjee**[2], **Benoît Liquet**[1,3] **Contact : sebastien.coube@lilo.org**

1 Université de Pau et des Pays de l'Adour 2 University of California, Los Angeles 3 Macquarie University.

Spatial process models ?

The data is decomposed as

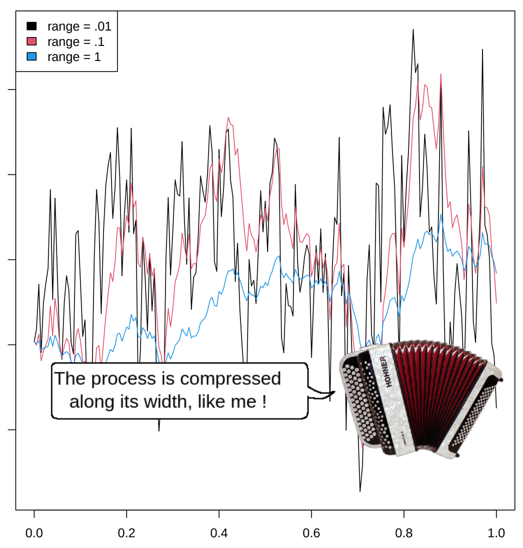
$$y(s) = X(s)\beta + w(s) + \epsilon(s)$$

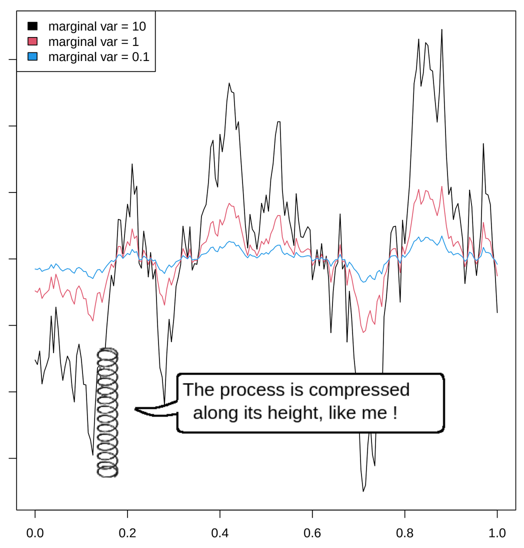
- s : **spatial coordinate**
- $y(\cdot)$: numeric interest variable
- $X(\cdot)$: matrix of covariates
- β : vector of regression coefficients
- $\epsilon(\cdot)$: independent Gaussian noise
- $w(\cdot)$: Gaussian noise with **spatial correlation**

Save for $w(\cdot)$ and s, we have a **good old linear model**

$w(\cdot)$ has a prior with **covariance parameters** θ :

$$(w(s_1), \ldots, w(s_n)) \overset{a\ priori}{\sim} \mathcal{N}(0, \Sigma(\theta, s_1, \ldots, s_n))$$





An **approximation** is used for the prior of $w(\cdot)$: the **Nearest Neighbor Gaussian Process**.

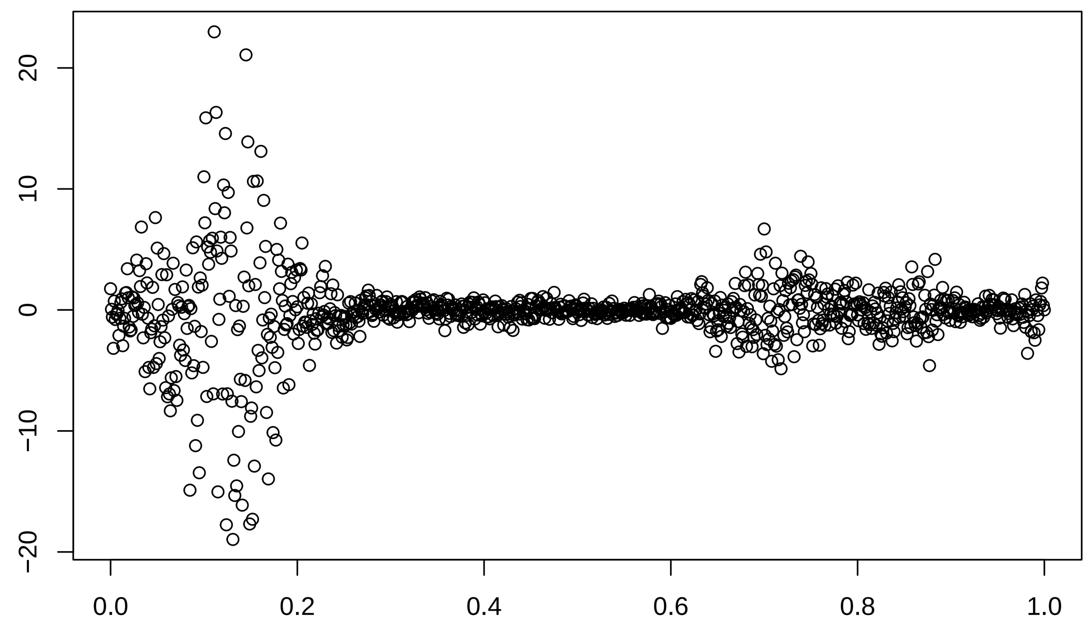
$$\begin{aligned} p(w(s_1), \ldots, w(s_n)) &\leftarrow O(n^3) \\ &= p(w(s_1))\Pi_{i=2}^{n} p(w(s_i)|w(s_1), \ldots, w(s_{i-1})) \\ &\approx p(w(s_1))\Pi_{i=2}^{n} p(w(s_i)|w(nn(s_i))), \end{aligned}$$

- $nn(s_i)$ are the m nerarest neighbors of s_i among $s_1, \ldots, s_{i-1}$, usually 10 or 15
- $p(\cdot)$ is the prior density

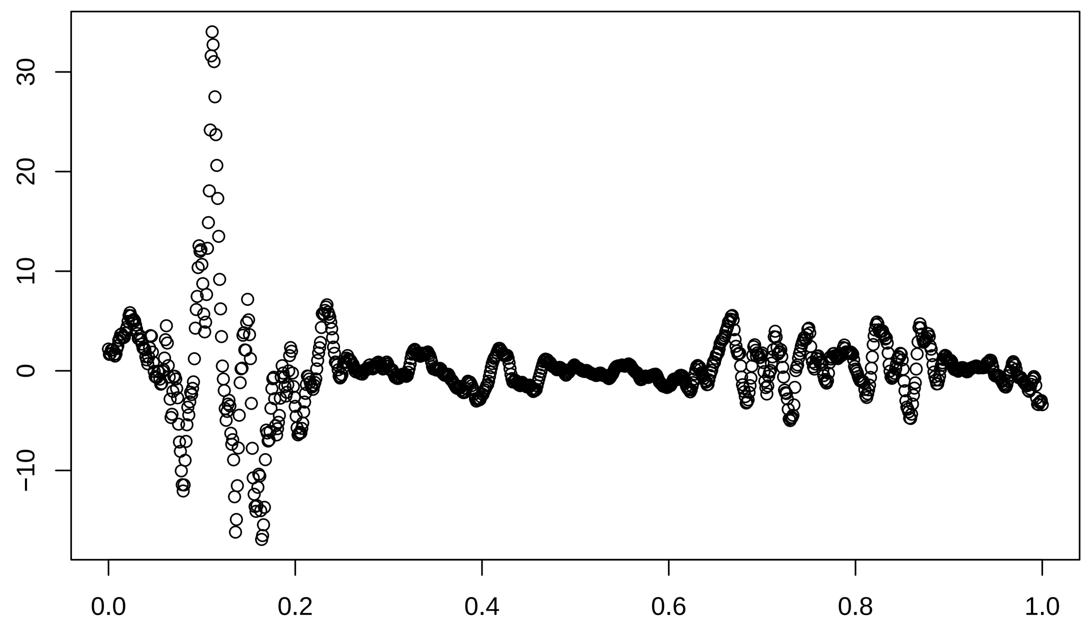
Cost is now $O(nm^3)$!

Nonstationarity ? That's disgusting ! Where ?

In the variance of the noise $\epsilon(\cdot)$

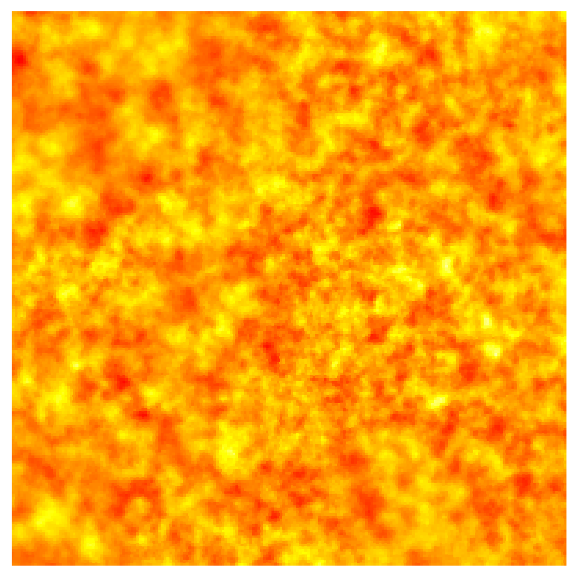


In the marginal variance of $w(\cdot)$

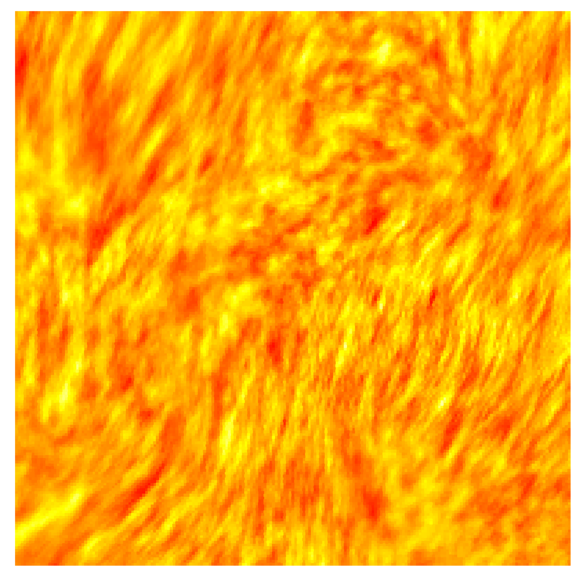


In the latent process range using Paciorek's covariance

With varying range...



... and anisotropy as well



Modeling

The **hierarchical model** has several layers

Covariance parameters (few!)
↓
Gaussian field — Covariates
↓ ↙
Covariance parameters (field)
↓
Gaussian field — Covariates
↘ ↙
Observations

The simple models are states of the complex models

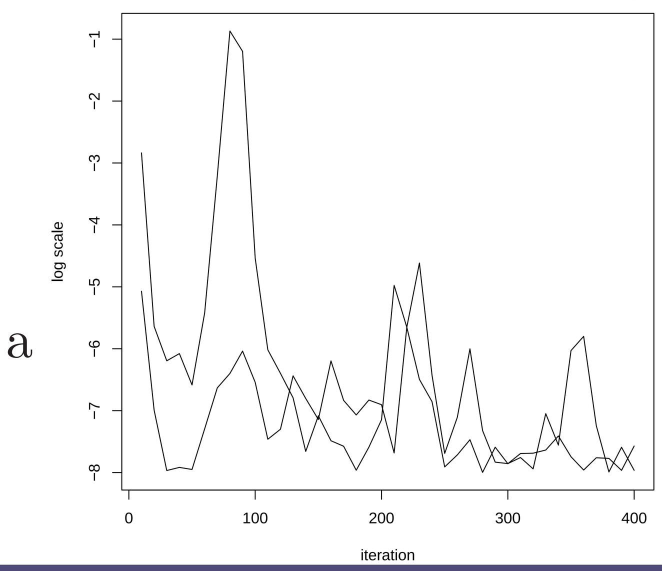
What model do I chose ?

- Heteroskedastic noise: always
- Range or GP variance: pick one
- Anisotropy: if applicable

What if my model is too complex ?

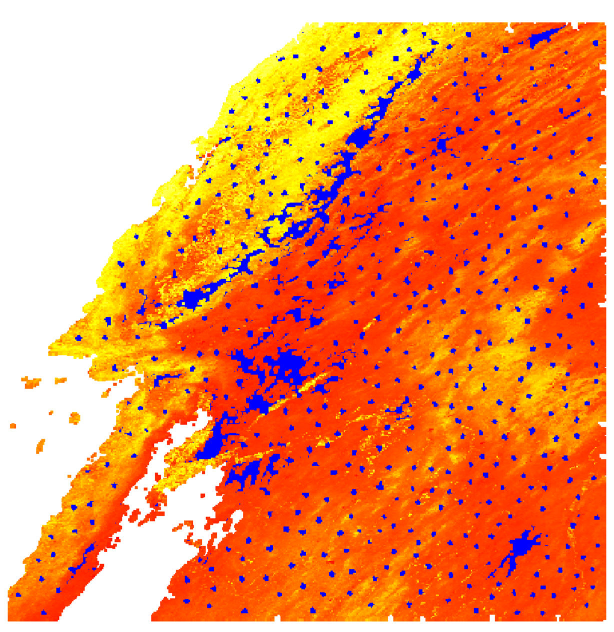
- The model can degenerate towards a simpler model
- Monitored using MCMC samples

MCMC samples dropping because of over-modeling

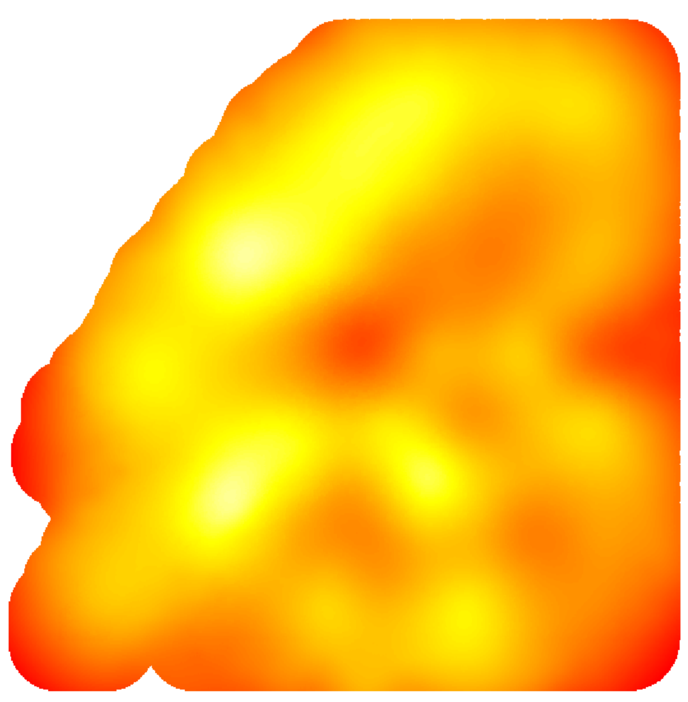


Real data analysis: NDVI data set (10^6+ observations)

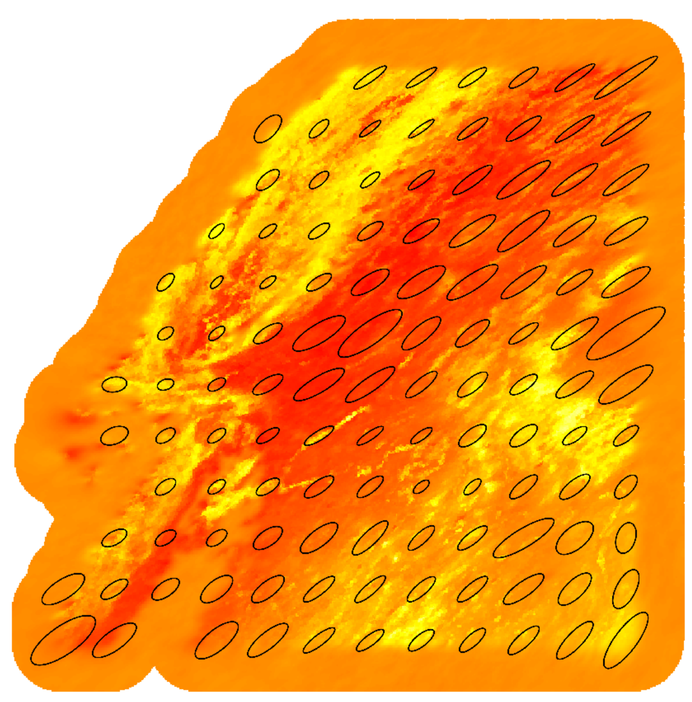
Observed (gaps in blue) :



Noise log var :



GP mean + range ellipses :



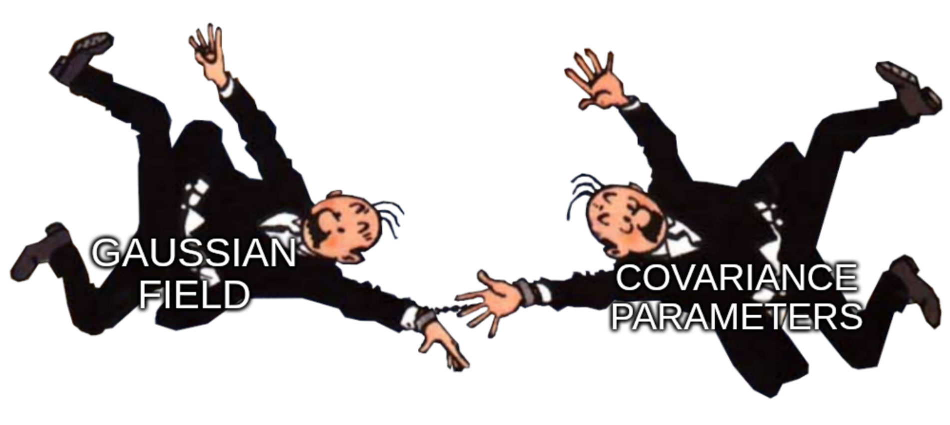
Comparison using empirical log point-wise predictive density

Model	Aniso	Range	Noise	Train	LOO	Lump	Time
NNGP	y	y	y	0.12	-0.06	-0.35	26 h 8
NNGP	y	n	y	0.11	-0.08	-0.39	9 h 3
NNGP	n	n	y	0.11	-0.11	-0.52	3 h 42
NNGP	n	n	n	-0.04	-0.26	-0.57	3 h 7
NNGP	n	y	y	0.12	-0.1	-0.51	7 h 29
Local	n	y	y	-0.42	-0.37	-1.04	0 h 2
INLA	n	n	n	-0.04	-0.27	-0.58	0 h 7
INLA	n	n	y				(crashed)
INLA	n	y	y				(crashed)

Computation

Challenges:

- High dimension of covariance parameters
- Coupling of GP and covariance parameters



Algorithm:

- MCMC loop like a Gibbs sampler
- **Interweaving** for coupling
- **Hybrid Monte-Carlo** for high dimension

Novel stuff:

- Nontrivial (understatement) gradients
- Nested interweaving for "lasagna" model
- Proof-of-concept HMC-within-interweaving
- Cost robustness wrt parameter dimension

Resources

- Preprint on ArXiV 2203.11873
- **Package available** on Github: SebastienCoube/Nonstat-NNGP
- **Vignette** on Github too !
  - Good for non-specialists
  - Lots of sample code
  - Intuitive presentation of the model
  - Gallery of real data analysis
- The horse's mouth. Contact Sébastien!