Latent Gauss-Markov models for spatial and spatiotemporal conditional extremes

Thomas Opitz (Biostatistics and Spatial Processes, INRAE, Avignon, France)

joint work with

Emma Simpson (University College London), Jenny Wadsworth (Lancaster University, UK)

ISBA 2022
Montreal

June 30, 2022

Biostatistique BIO∫Π & Processus Spatiaux

INRAE

Bayesian modeling and specificity of extremes

- Extreme-value setting: extensive probabilistic theory but **data are scarce!**
- **Bayesian benefits:**
  - Incorporation of physical constraints and expert opinion
  - Pooling of information, e.g. across space and time:
    - Hierarchical structures for parameters
    - Regularization/smoothing

Today's talk: Focus on **spatial dependence modeling**
(marginal distributions are normalized beforehand)

Reviews of Bayesian modeling and software for extremes:

- *Coles, S. G., & Powell, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments. International Statistical Review/Revue Internationale de Statistique, 119-136.*
- *Stephenson A.G. (2016) Bayesian inference for extreme value modelling. In: Dey DK, Yan J (eds) Extreme Value Modeling and Risk Analysis: Methods and Applications, CRC Press, Boca Raton, FL, pp 257–280*
- *Belzile, L. R., Dutang, C., Northrop, P. J., & Opitz, T. (2022). A modeler's guide to extreme value software. arXiv preprint arXiv:2205.07714.*

Modeling spatially dependent extremes

Spatial extent and duration of extreme episodes often matter
(*e.g.*, heatwaves, storms, precipitation cumulated over a catchment)

⚠ Plain Gaussian processes usually inappropriate for extreme events

- **Extreme-value limit theory** for linearly rescaled **maxima** or **threshold exceedances**
  $\rightarrow$ Max-stable processes, Peaks-over-threshold stable processes

- **Strong assumptions on dependence stability** in such classical limit models

  ⚠ Dependence strength does not change with the level of extremeness:

  $$t \times \Pr(\boldsymbol{X}^\star \in tE) = \Pr(\boldsymbol{X}^\star \in E) \quad \text{for extreme events } E \text{ and any } t > 0$$

Conditional extremes framework [Heffernan and Tawn, 2004]

Principle: Condition the process $X(s)$ on exceedance $X(s_0) > u$ at fixed location s_0

$\Rightarrow$ **Higher flexibility** with respect to extremal dependence structures

$\Rightarrow$ Recent **spatial and spatiotemporal extensions**
[Simpson and Wadsworth, 2020, Wadsworth and Tawn, 2022]

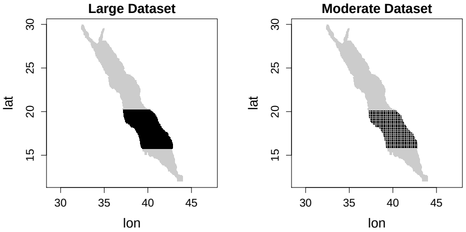
The "big-n problem" in spatial modeling

Spatial: $n =$ number of observation locations
Spatiotemporal: $n =$ number of spatial locations $\times$ number of time steps

- Likelihoods require costly computations involving **large variance-covariance matrices**
- **Low-rank representations** of matrices can alleviate the problem
- **Latent Gaussian fields**: number of latent variables $\ll$ number of observations
- **Markov structures** in latent Gaussian fields $\Rightarrow$ **Sparse precision matrices**
- **Bayesian INLA–SPDE framework** leverages latent Gauss–Markov fields

---

Example: OSTIA reanalysis data
$\rightarrow$ Red Sea Surface Temperatures
$\rightarrow$ 16703 pixels



Scope of today's talk

- Framework of **spatial conditional extremes**:
  certain **nonstationary Gaussian processes** flexibly capture extremal dependence

- **INLA-SPDE framework:**
  flexible Bayesian modeling **with many observation locations**

- Application to **marine heatwaves in the Red Sea**

1. Modeling spatial conditional extremes
2. High-dimensional Bayesian spatial conditional extremes
3. Application to Red Sea surface temperatures

The general spatial conditional extremes framework

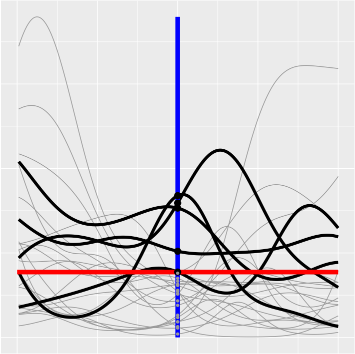
Spatial conditional extremes [Wadsworth and Tawn, 2022]

$\Rightarrow$ A flexible modeling framework beyond asymptotic dependence stability

- Observations $X_t(s)$ at regular time steps $t = 1, 2, \dots$
- Process is modeled conditional on threshold exceedance at fixed location s_0 (i.e. an extreme event occurs at time t if $X_t(s_0) > u$)

Illustration in $\mathbb{R}^1$:

- Conditioning location s_0 (blue)
- Threshold u (red)
- Extreme episodes (black)



Asymptotic formulation of conditional extremes

[⚠ We drop the time replicate index t to simplify notations!]

Assume a stationary spatial process $\{X(s)\}$ with **exponential upper tails** :

$$\exp(x) \times \Pr(X(s) > x) \to 1, \quad x \to \infty$$

$\Rightarrow$ In practice, we **pretransform to standard Laplace distribution**

Asymptotic assumption

Given a threshold $u \to \infty$, we assume that there exist **deterministic normalizing functions** $a_{s-s_0}(X(s_0))$ and $b_{s-s_0}(X(s_0))$ such that

$$\left(\frac{X(s) - a_{s-s_0}(X(s_0))}{b_{s-s_0}(X(s_0))}, X(s_0) - u \right) \Bigg| \, (X(s_0) > u) \xrightarrow{d} (Z^0(s), E), \quad u \to \infty$$

- $E \sim \mathrm{Exp}(1)$ at the conditioning location s_0
- **Residual stochastic process** $Z^0(s)$: independent of E, with constraint $Z^0(s_0) = 0$
- For coherence, we need $a_0(x) = x$ and $a_{s-s_0}(x) \leq x$
- No dependence stability if we can define $a_{s-s_0}(x) < x$ for $s \neq s_0$

Modeling assumption in practice [Wadsworth and Tawn, 2022]

$$X(s) \mid (X(s_0) = x) \quad \stackrel{d}{=} \quad a_{s-s_0}(x) + b_{s-s_0}(x)\, Z^0(s), \quad x > u$$

[Wadsworth and Tawn, 2022] develop parametric models:

- Residual process $Z^0(s)$ based on parametric stationary Gaussian process $Z(s)$, *e.g.*

  $$Z^0(s) = Z(s) - Z(s_0)$$

  with the flexible **Mátern covariance**

- Parametric forms of mean function $a_{s-s_0}(x)$ and scaling function $b_{s-s_0}(x)$

$\Rightarrow$ Frequentist full likelihood inference is tractable with up to hundreds of locations

⚠ **We want to handle data with thousands of locations!**

Modeling extensions to the Wadsworth–Tawn models

New extensions:

- **Semi-parametric conditional mean** $a_{s-s_0}(x)$ using basis expansions
- **Higher-dimensional inference:** work with latent Gaussian processes to allow for

  number of latent Gaussian variables $\ll$ number of observations

- **Bayesian inference** using **Penalized Complexity priors** for hyperparameters (variances, correlations) [Simpson et al., 2017]

Implementation:

- **Stochastic Partial Differential Equation** for latent Gauss–Markov process $Z(s)$
- **Integrated Nested Laplace Approximation** for Bayesian estimation

The Stochastic Partial Differential Equation approach

SPDE approach $\Rightarrow$ computationally convenient representations of Matérn covariance

Theoretical result [Whittle, 1954]:

Gaussian process $Z(s)$ with **Matérn covariance** (regularity ν) is stationary solution to

$$\left(\kappa^2 - \Delta\right)^{\zeta/2} Z(s) = B(s), \quad s \in \mathbb{R}^D, \quad \zeta = \nu + D/2$$

with

- Laplacian operator $\Delta y = \sum_{j=1}^{D} \partial^2 y / \partial^2 x_j$
- Gaussian white noise process $B(s)$

From SPDEs to high-dimensional spatial statistics

Approximate Gauss–Markov solution [Lindgren et al., 2011]

- Triangulation of space using a mesh with knots $\tilde{s}_\ell$, $\ell = 1, \ldots, m$
- Finite-element basis representation $Z(s) = \sum_{\ell=1}^{m} Z(\tilde{s}_\ell)\psi_\ell(s)$
- Solve SPDE on mesh to obtain

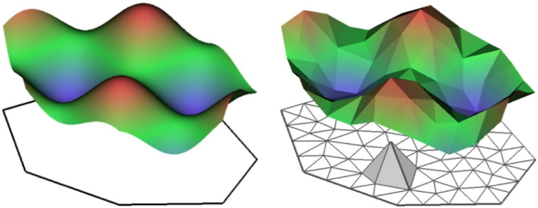
$$(Z(\tilde{s}_1), \ldots, Z(\tilde{s}_m)) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1})$$

with **sparse precision matrix** $\boldsymbol{Q}$

Spatial interpolation (towards any set of locations $s_1, \ldots, s_d$):

$$(Z(s_1), \ldots, Z(s_d))^T = \boldsymbol{A}(Z(\tilde{s}_1), \ldots, Z(\tilde{s}_m))^T$$

with **observation matrix** $\boldsymbol{A}$ having entries $a_{i,k} = \psi_k(s_i)$



“INLA models” are Bayesian hierarchical models

Integrated Nested Laplace Approximation [Rue et al., 2009]:
allows estimating **latent Gaussian models** through **Laplace approximations**

General Bayesian hierarchical structure

Given a generic observation vector $\boldsymbol{V}$, we assume:

$$\begin{aligned} \boldsymbol{\theta} &\sim \pi(\cdot) && \text{hyperparameters} \\ \boldsymbol{W} \mid \boldsymbol{\theta} &\sim \mathcal{N}_m(\mathbf{0}, Q(\boldsymbol{\theta})^{-1}) && \text{latent Gaussian components} \\ V_i \mid \boldsymbol{W}, \boldsymbol{\theta} &\overset{\text{ind.}}{\sim} \pi(\cdot \mid \eta_i(\boldsymbol{W}), \boldsymbol{\theta}) && \text{likelihood of observations} \end{aligned}$$

- Laplace approximation is deterministic and is used to
  - integrate out $\boldsymbol{W}$ in the posterior $\pi(\boldsymbol{\theta} \mid \boldsymbol{V})$, and
  - integrate out $\boldsymbol{\eta}_{-j}$ in the posterior $\pi(\eta_j, \boldsymbol{\theta} \mid \boldsymbol{V})$
- Numerical integration to integrate out small number of hyperparameters $\boldsymbol{\theta}$

$\Rightarrow$ **Accurate deterministic approximations** for all **univariate posteriors**

Spatial conditional extremes in INLA

Considered model structure ($\beta \in [0,1]$ controls residual variance depending on x):

$$X(s_i) \mid [X(s_0) = x] = x \times \alpha(\|s_i - s_0\|) + \gamma(\|s_i - s_0\|) + x^\beta Z^0(s_i)$$

Model formulation in INLA

$$V_i \mid \eta_i \sim \mathcal{N}(\eta_i, \sigma^2) \quad \text{with} \quad \eta_i = x \cdot \alpha(s_i - s_0) + \gamma(s_i - s_0) + x^\beta Z^0(s_i)$$

- Residual process $Z^0(s) = Z(s) - Z(s_0) \to$ SPDE in 2D for $Z(s)$
- Spline functions α and $\gamma \to$ SPDE in 1D with boundary constraints at $s_i - s_0 = 0$
- Small i.i.d. noise with variance σ^2 to "match" η_i and V_i

Observation matrix $A \in \mathbb{R}^{n\times m}$ with $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{W}) = A\boldsymbol{W}$

- $\boldsymbol{W} = (\boldsymbol{W}_\alpha^\top, \boldsymbol{W}_\gamma^\top, \boldsymbol{W}_Z^\top)^\top \in \mathbb{R}^{m_\alpha} \times \mathbb{R}^{m_\gamma} \times \mathbb{R}^{m_Z}$, with $m_\alpha + m_\gamma + m_Z = m$
- A is the concatenation of $A_\alpha \in \mathbb{R}^{d\times m_\alpha}$, $A_\gamma \in \mathbb{R}^{d\times m_\gamma}$, and $A_S^0 \in \mathbb{R}^{d\times m_Z}$
- To model $Z^0(s) = Z(s) - Z(s_0)$, we set $A_S^0 = A_{s_1,\dots,s_d} - \begin{pmatrix} A_{s_0} \\ \vdots \\ A_{s_0} \end{pmatrix} \in \mathbb{R}^{d\times m_Z}$

Recap of the full algorithm

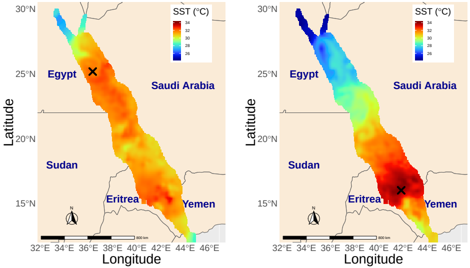
1. Estimate univariate model and normalize margins to standard Laplace $X_t(s)$
2. Fix a conditioning site s_0 and a high threshold u
3. Extract extreme episodes occuring at times $\tilde{t}$ with $X_{\tilde{t}}(s_0) > u$
4. Fit spatial INLA-SPDE Gaussian model to replicates of extreme episodes

1. Modeling spatial conditional extremes
2. High-dimensional Bayesian spatial conditional extremes
3. Application to Red Sea surface temperatures

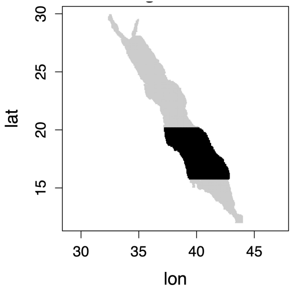
Red Sea surface temperature hotspots

- The Red Sea is a rich ecosystem with many endemic species and coral reefs
  ⚠ Climate change impacts (*e.g.*, coral bleaching due to marine heatwaves)
- **Dataset:** OSTIA gridded daily SST data (1985–2015, 16,703 pixels)
- Focus on southern Red Sea with many coral reefs (6239 grid cells)
- Conditioning location s_0 far from the coast
- Using the 95%-quantile as threshold, we extract **141 extreme episodes**

Examples of extreme episodes



Dataset



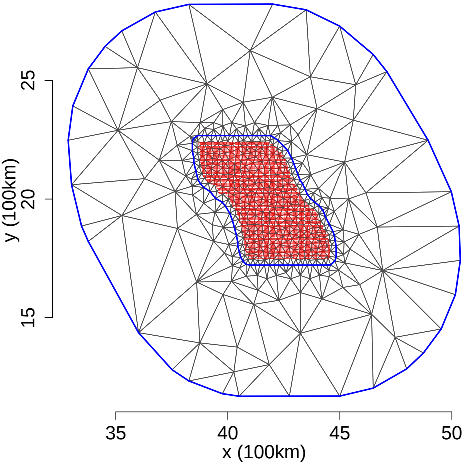
Model variants of conditional extremes

Various submodels of the most complex model are considered:

$$
\begin{aligned}
&(1) \quad x \cdot \alpha(\|s - s_0\|) + Z^0(s) \\
&(2) \quad x + \gamma(\|s - s_0\|) + Z^0(s) \\
&(3) \quad x \cdot \alpha(\|s - s_0\|) + \gamma(\|s - s_0\|) + Z^0(s) \\
&(4) \quad x \cdot \alpha(\|s - s_0\|) + \gamma(\|s - s_0\|) + x^\beta \, Z^0(s)
\end{aligned}
$$

Finite-element mesh for Gauss–Markov $Z(s)$

- Mesh extension beyond the study area to mitigate SPDE boundary conditions
- 541 mesh nodes ($= 8.7\%$ of number of observation pixels n)
- Total of $541 \times 141 = 76281$ **latent Gaussian variables** for residual process Z^0



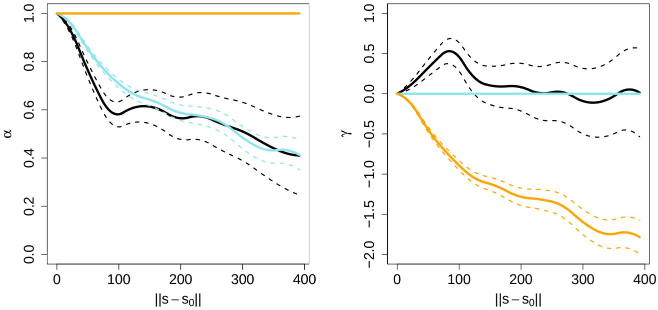
Estimation results

- INLA computation times between 20 and 110 minutes
- WAIC selects most complex models 3 and 4 (approximately equivalent)

Parameter (Model 3)	Posterior mean	95% credible interval
Noise variance σ^2	0.0107	(0.0106,0.0107)
Standard deviation of $Z(s)$	1.56	(1.50,1.62)
Correlation range of $Z(s)$ in km	428	(410, 447)

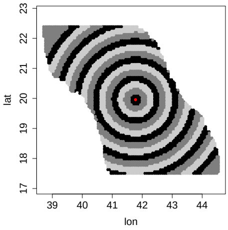
Estimated splines in the conditional mean: α (left), γ (right)

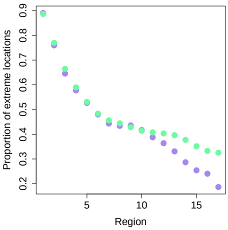
$\rightarrow$ Black curves for Model 3



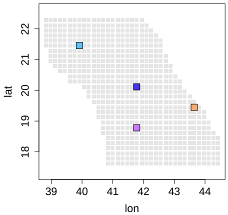
Validation of key properties

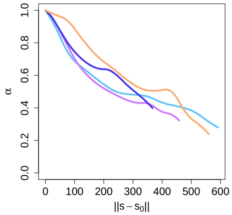
Joint exceedance probabilities: Model 3 (green) vs. Empirical (purple)





Spatial stationarity (when shifting s_0): function α estimated for different s_0





Take-home messages

- Flexible models to assess **extent of extreme-event episodes**
- **Latent variable approach** to decouple latent and observed vector dimensions
- **INLA-SPDE** for fast Bayesian estimation at high spatial resolution
- Extensions to **extreme space-time episodes** spanning $T > 1$ consecutive days:
  - Conditioning location (s_0, t_0) at first day t_0 of episode
  - Tensor-product spline functions $\alpha(\mathrm{dist}, \mathrm{day})$ and $\gamma(\mathrm{dist}, \mathrm{day})$ for $\mathrm{day} \in 1, \ldots, T$
  - $Z(s, t)$ has first-order autoregressive structure (AR1) in time

$\Rightarrow$ Tensor products and AR1 preserve sparse precision matrices $\boldsymbol{Q}$!

Simpson, Opitz, Wadsworth, 2022+, in revision for Extremes
arXiv: `https://arxiv.org/abs/2011.04486`
Code (Github): `https://github.com/essimpson/INLA-conditional-extremes`

— Thank you for your attention —

Heffernan, J. E. and Tawn, J. A. (2004).
A conditional approach for multivariate extreme values (with discussion).
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(3):497–546.

Lindgren, F., Rue, H., and Lindström, J. (2011).
An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(4):423–498.

Rue, H., Martino, S., and Chopin, N. (2009).
Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.
Journal of the Royal Statistical Society (Series B), 71(2):319–392.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., and Others (2017).
Penalising model component complexity: A principled, practical approach to constructing priors.
Statistical Science, 32(1):1–28.

Simpson, E. S. and Wadsworth, J. L. (2020).
Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures.
arXiv preprint arXiv:2002.04362.

Wadsworth, J. L. and Tawn, J. (2022+).
Higher-dimensional spatial extremes via single-site conditioning.
Spatial Statistics.

Whittle, P. (1954).
On stationary processes in the plane.
Biometrika, 41(3/4):434–449.