Spatial joint modeling of extremes and angles

Gaspard Tamagny, Mathieu Ribatet

Data Science pour les risques côtiers, Roscoff

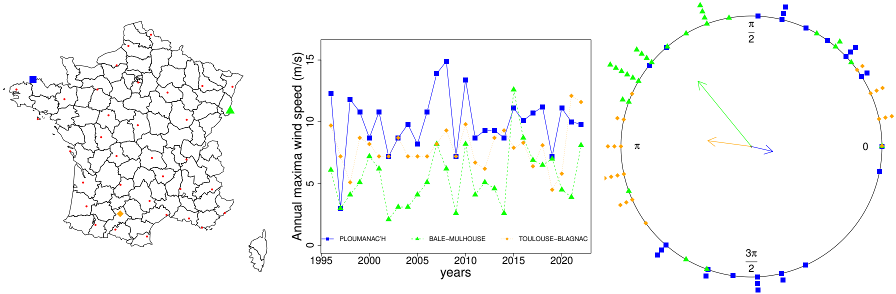


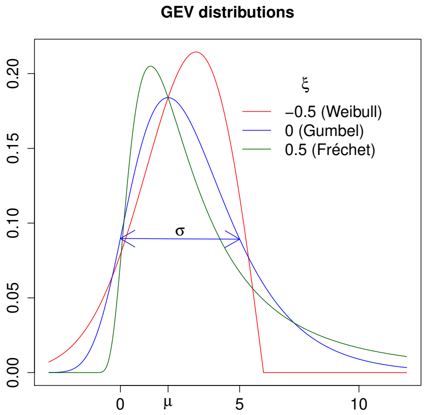
Figure 1: Wind gusts at 10m in France (Météo-France). From left to right: study region and locations of the $k = 36$ weather stations; times series of $n = 27$ years for three observed stations; empirical distribution of wind direction for the observed stations.

⚠Angular data are probably shifted.

⚠Dealing with angular data needs specific statistical precautions (e.g. angular mean, angular dispersion...).

Our model: how to deal with spatial extremes ?

- Block-maxima approach : we only consider the maximum of the data over a fixed period of time
- This maximum con be modeled by a **Generalized Extreme Value distribution (GEV)** with three parameters : location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$.



Our model: how to deal with spatial extremes ?

- **Max-stable process:** the distribution of the extreme process η as a random variable in $\mathcal{C}(\mathcal{X})$ (de Haan and Fereira (2006)).
- **Conditional independence model**: conditionally on the GEV parameters, $\eta(s)$ and $\eta(s')$ are independent (Cooley et al. (2007)).

$$
\begin{aligned}
\eta(s) \mid \{\mu(s), \sigma(s), \xi(s)\} &\overset{\text{ind}}{\sim} \text{GEV}\{\mu(s), \sigma(s), \xi(s)\}, \qquad s \in \mathcal{X} \\
\mu(\cdot) &\sim \text{GP}(m_\mu, \gamma_\mu) \\
\sigma(\cdot) &\sim \text{GP}(m_\sigma, \gamma_\sigma) \\
\xi(\cdot) &\sim \text{GP}(m_\xi, \gamma_\xi)
\end{aligned}
$$

Advantage: simple univariate likelihood for inference.
Drawback: inability to estimate areal quantities;

Our model: how to deal with spatial angles ?

Our model: how to deal with spatial angles ?

- Projection method: $\theta \sim \text{proj}_{\mathbb{S}^1}(Z)$ with Z random variable in $\mathbb{R}^2$.
- Explicit univariate distribution.
- Simple spatial counterpart if Z is a spatial process.

Our model: how to deal with spatial angles ?

The **projected gaussian process** (Gelfand and Wang (2014)) advantages: low number of parameter and high flexibility.

$$Z(\cdot) = \begin{pmatrix} Z_1(\cdot) & Z_2(\cdot) \end{pmatrix}^\top \sim \mathrm{GP}_2(m_\theta, \gamma_\theta)$$
$$\theta(\cdot) = \arctan^* \frac{Z_2(\cdot)}{Z_1(\cdot)}$$

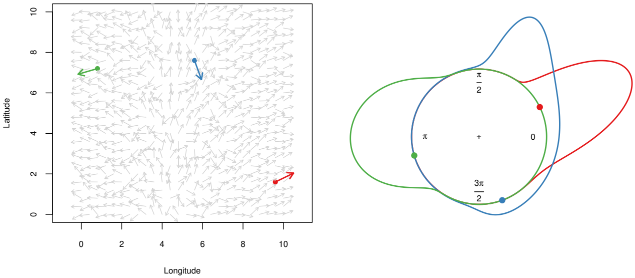


Figure 2: Left: one realisation of a projected Gaussian process. Right: Marginal distributions at the three highlighted locations.

Our model: how to deal with spatial angles ?

- Data augmentation: radial latent processes R, such that

$$Z_1(s) = R(s)\cos\theta(s) \text{ and } Z_2(s) = R(s)\sin\theta(s).$$

- Gaussian likelihood:

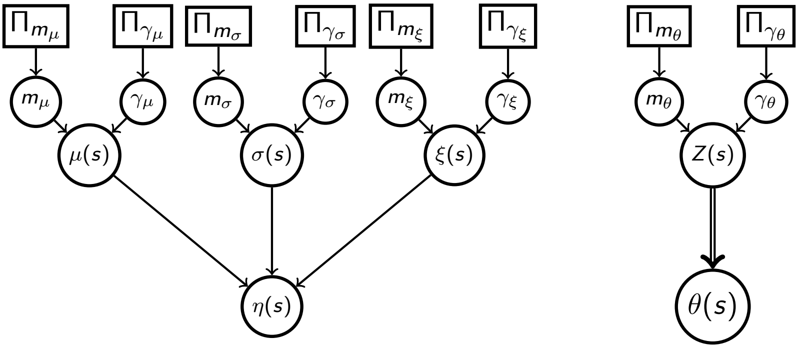
$$f_{\boldsymbol{s}}(\boldsymbol{r}, \boldsymbol{t}) = \frac{\prod_{i=1}^{k} r_i}{(2\pi)^{k/2}\sqrt{|\gamma_\theta(\boldsymbol{s})|}} \exp\left(-\frac{\{\boldsymbol{r}^\top \boldsymbol{u} - m_\theta(\boldsymbol{s})\}^\top \gamma_\theta(\boldsymbol{s})^{-1}(\boldsymbol{r}^\top \boldsymbol{u} - m_\theta(\boldsymbol{s})\}}{2}\right)$$

with

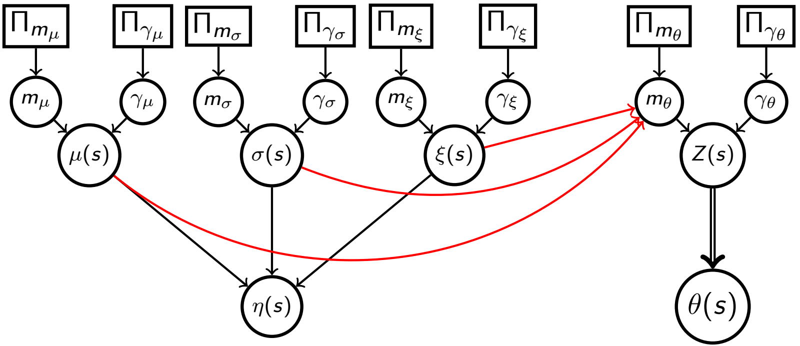
$$\boldsymbol{u} = \begin{bmatrix} \cos t_1 & \dots & \cos t_k \\ \sin t_1 & \dots & \sin t_k \end{bmatrix}^\top$$

Our model: how to combine both approaches ?

$$\eta(s) \overset{\text{Ind}}{\sim} \text{GEV}(\mu(s), \sigma(s), \xi(s)) \qquad \theta \sim \text{PGP}(m_\theta(\cdot), \gamma_\theta(\cdot))$$



$$\eta(s) \overset{\text{Ind}}{\sim} \text{GEV}(\mu(s), \sigma(s), \xi(s)) \qquad \theta \sim \text{PGP}(m_\theta(\cdot), \gamma_\theta(\cdot))$$



Parametrization

- **Regression model** for the mean functions:

$$m_{\mu/\sigma/\xi}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s) + \cdots$$
$$m_\theta^{(1/2)}(s) = \cdots + \beta_4 f(\mu(s), \sigma(s), \xi(s)) + \beta_5 g(\mu(s), \sigma(s), \xi(s)) + \cdots$$

For instance, one can take the quantile function :

$$\begin{aligned} f(\mu(s), \sigma(s), \xi(s)) &= F^{-1}(0.95 | \mu(s), \sigma(s), \xi(s)) \\ &= \mu(s) + \frac{\sigma(s)}{\xi(s)}((-\log(p))^{-\xi(s)} - 1) \end{aligned}$$

Parametrization

- (Cross)covariance function: **stationary**, **isotropic**, **exponentially decreasing**

$$\gamma_{\cdot}(s, s+h) = \tau_{\cdot}\Gamma(\|h\|; \lambda_{\cdot}) \text{ with } \Gamma(t; \lambda_{\cdot}) = e^{-\frac{t}{\lambda_{\cdot}}}$$

and **separable :** $\gamma_\theta(s, s+h) = T \otimes \Gamma_\theta(\|h\|)$

with $T = \begin{bmatrix} \tau_\theta & \rho_\theta\sqrt{\tau_\theta} \\ \rho_\theta\sqrt{\tau_\theta} & 1 \end{bmatrix}$, $\tau_\theta > 0$ and $\rho_\theta \in (-1, 1)$.

Gibbs sampler for inference

- Aim: sampling each parameters and latent variables from posterior distribution $\pi(\Psi \mid \mathcal{D}_n)$.

Gibbs sampler for inference

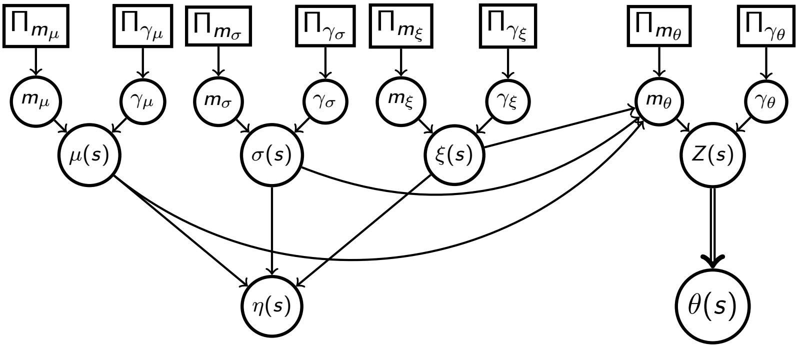
- Aim: sampling each parameters and latent variables from posterior distribution $\pi(\Psi \mid \mathcal{D}_n)$.
- MCMC: creating a Markov Chain with stationary distribution $\pi(\Psi \mid \mathcal{D}_n)$.
  At each step, each variable and parameter is sampled via the full conditional distribution:

$$\pi\{\mu(s_j) \mid \cdot\} \propto \pi\{\mu(\mathrm{s}) \mid m_\mu, \gamma_\mu\}\pi\{\boldsymbol{\eta}(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j)\}\pi\{\mathrm{Z}(\mathrm{s}) \mid m_\theta, \gamma_\theta, \mu(\mathrm{s}), \sigma(\mathrm{s}), \xi(\mathrm{s})\}$$

$$\pi\{R_i(\mathrm{s}) \mid \cdot\} \propto \pi\{R_i(\mathrm{s}) \mid \theta_i(\mathrm{s}), m_\theta, \gamma_\theta, \mu(\mathrm{s}), \sigma(\mathrm{s}), \xi(\mathrm{s})\}$$

$$\pi\{\lambda_\xi \mid \cdot\} \propto \pi\{\xi(\mathrm{s}) \mid \beta_\xi, \tau_\xi, \lambda_\xi\}\Pi\{\lambda_\xi\}$$

Updates: conjugate priors whenever possible, Metropolis-Hastings sampling if not.



Gibbs sampler for inference

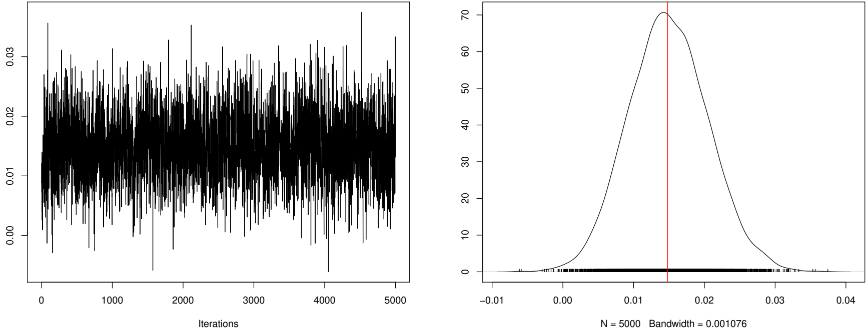


Figure 3: Chain (left) and posterior distribution (right) for $\beta_{2,\theta}$.

- Estimator: the median of the posterior distribution.

Simulation Study

- Multiple configurations: type of dependence, number of site k, number of observations n.
- 100 replicates for each configuration in order to compute a Mean Squared Error
- Two types of asymptotics: infill asymptotic ($k \to \infty$) and repeated observations ($n \to \infty$).

Simulation study

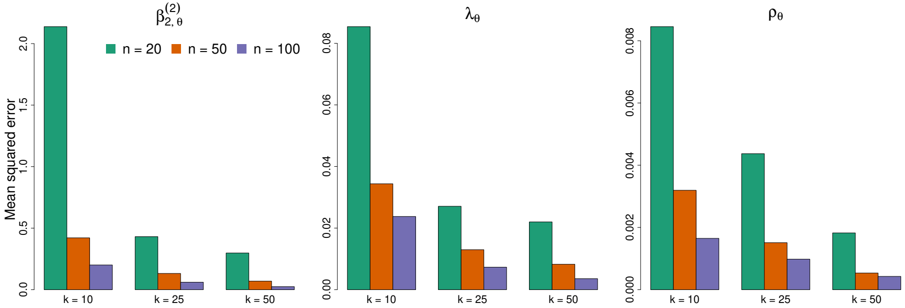


Figure 4: Evolution of the Mean Squared Error for angle related parameters with varying number of observations n and locations k.

- Good performance for both infill asymptotic ($k \to \infty$) and repeated observations ($n \to \infty$).

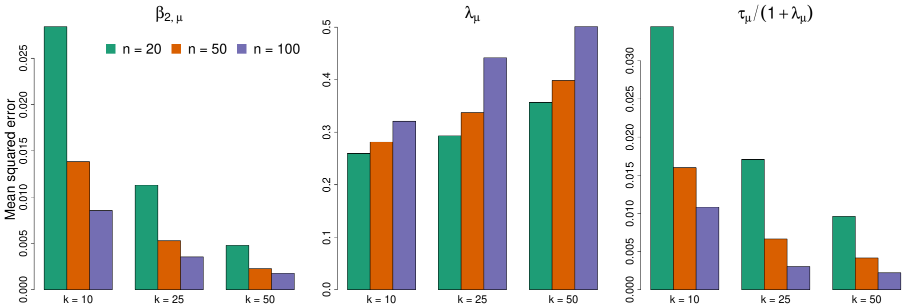
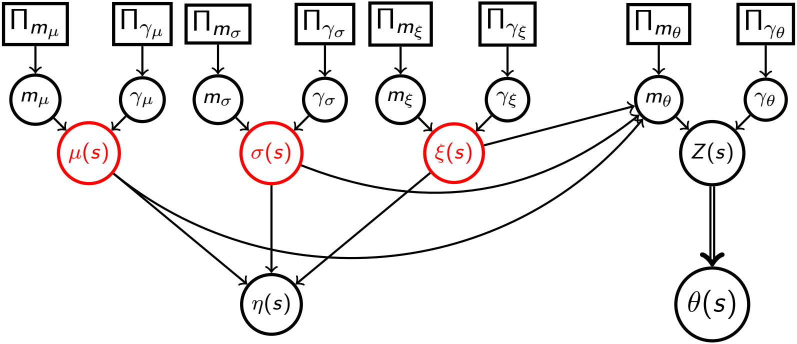


Figure 4: Evolution of the Mean Squared Error for parameters related to μ with varying number of observations n and locations k.

- Inconsistent estimation when $k \mapsto \infty$ for τ (sill) and λ (range) parameters (Zhang (2004)).



- Whatever the number of observation n, **only one value** for μ, σ and ξ.

Simulation study

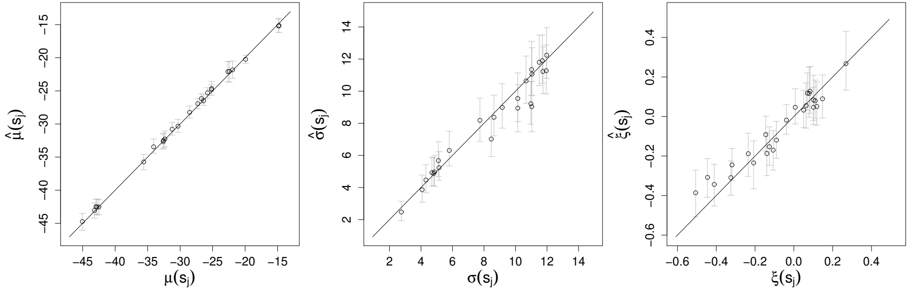


Figure 5: Posterior median estimation (with 95% credible intervals) for the GEV parameters at $k = 25$ locations and with $n = 50$ observations

Wind data: model selection using WAIC

Table 1: Widely Applicable bayesian Information Criteria (WAIC) for selected models. All models have $m_\xi(s) = \beta_0$.

		WAIC
Model 0	$m_\mu(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_3 \text{alt}(s)$	11,751
	$m_\sigma(s) = \beta_0 + \beta_1 \text{lat}(s)$	
	$m_{\theta,1}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s)$	
	$m_{\theta,2}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s)$	
Model 1	*Model 0*	11,557
	$+m_{\theta,1}(s) = \cdots + \beta_3 \mu(s) + \beta_4 \sigma(s)$	
	$+m_{\theta,2}(s) = \cdots + \beta_4 \mu(s) + \beta_5 F^{-1}(0.95 \vert \mu, \sigma, \xi)$	
Model 2	*Model 0*	11,301
	$+m_{\theta,1}(s) = \cdots + \beta_3 F^{-1}(0.95 \vert \mu, \sigma, \xi) + \beta_4 F^{-1}(0.99 \vert \mu, \sigma, \xi)$	
Model 3	*Model 0*	11,283
	$+m_{\theta,1}(s) = \cdots + \beta_3 \mu(s) + \beta_4 F^{-1}(0.95 \vert \mu, \sigma, \xi)$	
Model 4	*Model 3*	11,235
	$+m_\mu(s) = \beta_0 + \beta_1 \text{alt}(s)$	
	$+m_{\theta,2}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{alt}(s)$	
Model 5	*Model 4*	11,635
	$+m_{\theta,1}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \mu(s)$	
Model 6	*Model 4*	11,231
	$+m_\sigma(s) = \beta_0$	
	$+m_{\theta,1}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \mu(s) + \beta_3 F^{-1}(0.95 \vert \mu, \sigma, \xi)$	

Wind data: model selection using WAIC

Table 2: Parameters for the selected model

Regression	Sill	Correlation
$m_\mu(s) = \beta_0^\mu + \beta_{\text{alt}}^\mu \text{alt}(s)$	τ^μ	λ^μ
$m_\sigma(s) = \beta_0^\sigma$	τ^σ	λ^σ
$m_\xi(s) = \beta_0^\xi$	τ^ξ	λ^ξ
$m_{\theta,1}(s) = \beta_0^{\theta,1} + \beta_{\text{lon}}^{\theta,1}\text{lon}(s) + \beta_\mu^{\theta,1}\mu(s)$	τ^θ	
$+\beta_q^{\theta,1} F^{-1}(0.95 \mid \mu, \sigma, \xi)$		$\lambda^\theta, \rho^\theta$
$m_{\theta,2}(s) = \beta_0^{\theta,2} + \beta_{\text{lon}}^{\theta,2}\text{lon}(s) + \beta_{\text{alt}}^{\theta,2}\text{alt}(s)$	1	

Wind data: first results

- $m_\mu = 6.7_{\{-0.3,13.6\}} - 0.6_{\{-3.0,1.7\}}\text{alt}(s)$: slower extreme in high altitude.

Wind data: first results

- $m_\mu = 6.7_{\{-0.3,13.6\}} - 0.6_{\{-3.0,1.7\}}\text{alt}(s)$: slower extreme in high altitude.

Table 3: Posterior mean and 95% credible intervals (in bracket) for the projected gaussian process.

	β_0	β_{alt}	β_μ	β_{quant}
$m_{\theta,1}$	$1.9_{\{0.0,4.5\}}$	—	$1.9_{\{1.4,2.7\}}$	$-1.4_{\{-2.0,-1.0\}}$
$m_{\theta,2}$	$-0.2_{\{-0.3,-0.0\}}$	$0.4_{\{0.1,0.7\}}$	—	—

- As $F^{-1}(0.95|\mu(s),\sigma(s),\xi(s)) = \mu(s) + \frac{\sigma(s)}{\xi(s)}((-\log(p))^{-\xi(s)} - 1)$, we have : $m_{\theta,1} = 1.9 + 0.5\mu(s) - 1.4\frac{\sigma(s)}{\xi(s)}((-\log(p))^{-\xi(s)} - 1)$

Wind data: visualisation

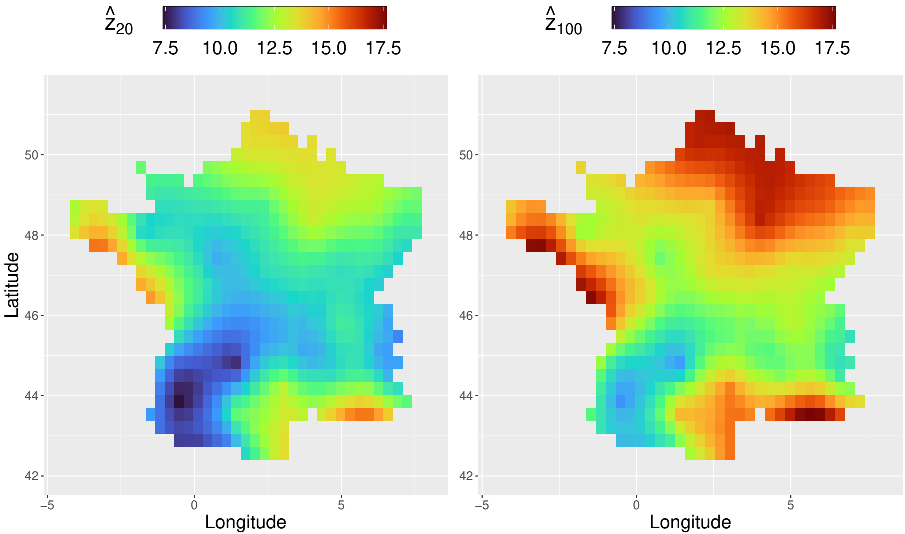


Figure 6: Predicted return levels of return period 20 years (left) and 100 years (right), in m/s.

Wind data: visualisation

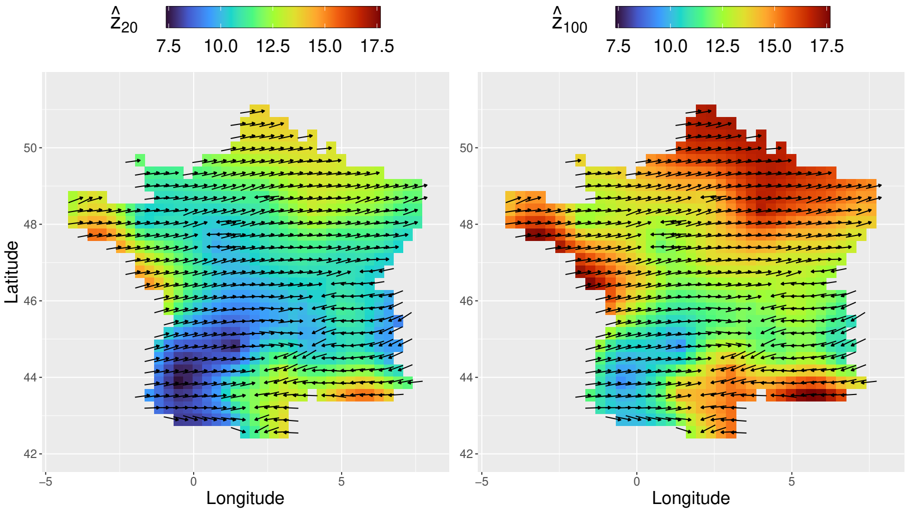


Figure 6: Predicted angles (arrows) and return levels of return period 20 years (left) and 100 years (right), in m/s.

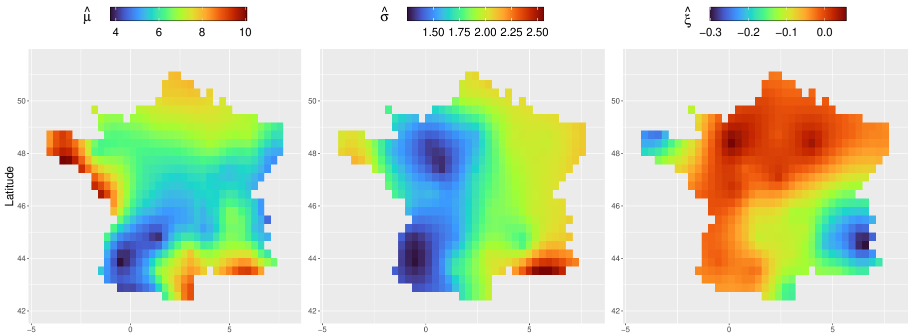


Figure 6: Posterior median of GEV parameters.

- High variation for the shape parameter : big impact on the return level and the angular distribution.

Wind data: visualisation

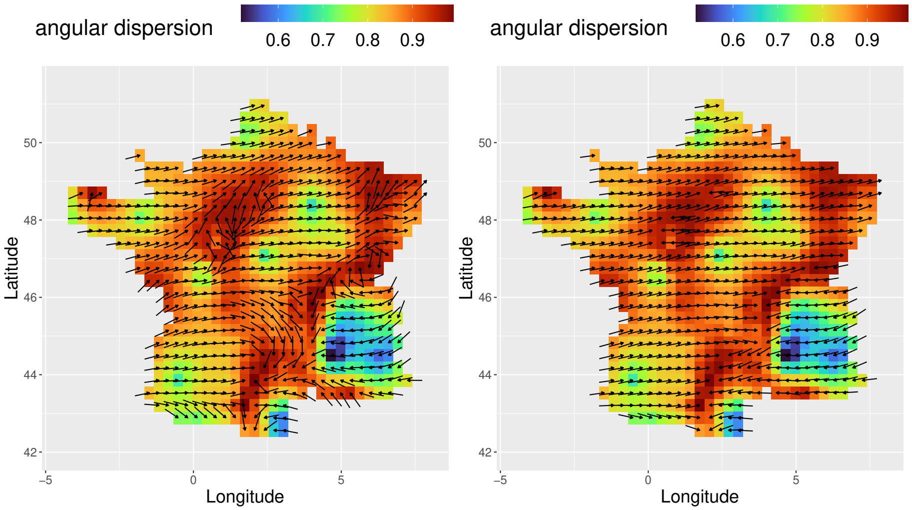


Figure 6: Posterior angular mean (left) and mode (right) and angular dispersion.

Prospect and future works

- Preprint available on arXiv:2306.08940
- Study of climate change potential effect via time related covariates.
- Ongoing work: max-stable with an angular component

Thank you for your attention !

References

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.*, 102(479):824–840.

de Haan, L. and Fereira, A. (2006). *Extreme value theory: An introduction*. Springer Series in Operations Research and Financial Engineering.

Gelfand, A. and Wang, A. (2014). Modeling space and space-time directional data using projected gaussian processes. *Journal of the American Statistical Association*.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14(27):867–897.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Seasonal max-stable process

Question: what is the distribution of η as a random variable in $\mathcal{C}(\mathcal{X}, \mathbb{R})$?

Spectral representation (de Hann, 1984)

η is a max-stable process, if and only if, $\eta \overset{\mathcal{L}}{=} \max_{i>0} \zeta_i W_i$, where:

- $(\zeta_i)_{i>0}$: a realization of a Poisson point process with intensity $r^{-2}dr$ on $\mathbb{R}^+$.
- $(W_i)_{i>0}$: i.i.d. stochastic process in $\mathcal{C}^+(\mathcal{X}, \mathbb{R})$, with $\forall s \in \mathcal{X}, \mathbb{E}(W(s)) = 1$ and $\mathbb{E}(\sup_{s \in \mathcal{X}} W(s)) < \infty$.

For each distribution of W, we have a max-stable model: Smith, Brown-Resnick, extremal-t...

Seasonal max-stable process

Useful lemma

$\forall s \in \mathcal{X}$, there is a unique $i(s) > 0$ such that $\eta(s) = \zeta_{i(s)} W_{i(s)}(s)$.

- Angular max-stable process (extremal-t):

$$\eta = \begin{pmatrix} \eta^{(1)} \\ \eta^{(2)} \\ \eta^{(3)} \end{pmatrix} \text{ with } \begin{cases} \eta^{(1)}(s) = \zeta_{i(s)} c_\nu W^{(1)}_{i(s)}(s)^\nu & : \text{ extreme} \\ \begin{pmatrix} \eta^{(2)}(s) \\ \eta^{(3)}(s) \end{pmatrix} = \zeta_{i(s)}^{1/\nu} \begin{pmatrix} W^{(2)}_{i(s)}(s) \\ W^{(3)}_{i(s)}(s) \end{pmatrix} & : \text{ angle} \end{cases}$$

and $(W_i)_{i>0}$: i.i.d. 3-dimensional Gaussian processes and $\nu > 1$.

- Next steps: compute the likelihood of our model and implement EM algorithm for inference.

Univariate distribution for θ

- The density for projected gaussian variable:

$$f_s(x) = \frac{\Phi_2\left(u(x); 0, \gamma(s,s)\right) + |\gamma(s,s)|^{-1/2} D(x)\phi_1(D(x))\Phi_1\left(a_s(x)^{-1}|\gamma(s,s)|^{-1/2}\mu(s) \wedge u(x)\right)}{a_s^2(x)}$$

where

- ϕ_1, Φ_1: pdf, cdf of standard normal distribution
- $\Phi_2(\cdot; 0, \Sigma)$: cdf of bivariate gaussian vector with covariance matrix Σ
- $u(x) = (\cos x, \sin x)^\top$
- $a_s^2(x) = u(x)^\top \gamma(s,s)^{-1} u(x)$
- $D(x) = a_s(x)^{-1}\mu(s)^\top \gamma(s,s)^{-1} u(x)$.

Widely applicable bayesian information criterion (Watanabe (2013))

- DIC focus on the fitting ability and penalizes model complexity.
- Hierarchical model equivalent of AIC:

$$\mathrm{DIC} = \overline{D(\Psi)} + p_D$$

- The deviance $D(\Psi) = -2 \log L(\Psi|\mathcal{D}_n)$: a low deviance means Ψ fit the training data set $\mathcal{D}_n$.
- The effective number of parameters $p_D = \overline{D(\Psi)} - D(\overline{\Psi})$: a high effective number of parameters means a complex model.

Deviance Information Criteria (Spiegelhalter et al. (2002))

- DIC focus on the fitting ability and penalizes model complexity.
- Hierarchical model equivalent of AIC:

$$\mathrm{DIC} = \overline{D(\Psi)} + p_D$$

- The deviance $D(\Psi) = -2 \log L(\Psi | \mathcal{D}_n)$: a low deviance means Ψ fit the training data set $\mathcal{D}_n$.
- The effective number of parameters $p_D = \overline{D(\Psi)} - D(\overline{\Psi})$: a high effective number of parameters means a complex model.

Continuous ranked probability score (Gneiting and Raftery (2007))

- Asses the prediction capacity of the model on a validation set.
- A score on the predictive posterior distribution F:

$$\text{CRPS} = \frac{1}{n}\sum_{i=1}^{n} \text{CRPS}(F, y_i)$$

with $\text{CRPS}(F, y) = \mathbb{E}_F|X - y| - \frac{1}{2}\mathbb{E}_F|X - X'|$

- Trade-off between calibration and sharpness.

Algorithm 1: Gibbs sampler

input : Target distribution g; initial state for $\text{Var} = (\psi, \mu(s), \sigma(s), \xi(s), R_i(s))$; proposition kernel $K(\cdot, \cdot)$; length of the chain N.

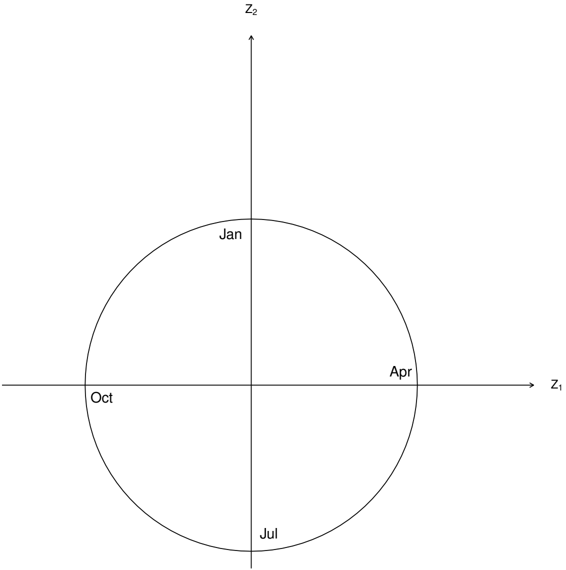
output: Markov chain with stationary distribution g.

for $t = 1, \ldots, N$ **do**
  for $v = 1, \ldots, Var$ **do**
    Sample v_* from $K(v_{t-1}, \cdot)$;
    Compute acceptance probability

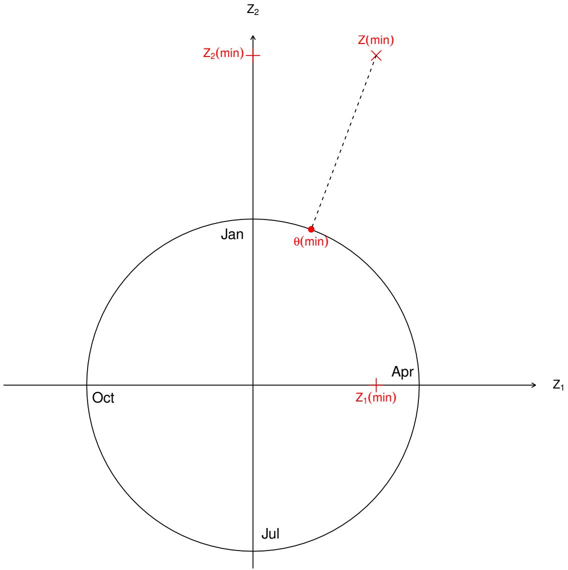
$$\alpha(v_{t-1}, v_*) = \frac{g(v_*)K(v_*, v_{t-1})}{g(v_{t-1})K(v_{t-1}, v_*)}$$

    Take $v_t = \begin{cases} v_*, & \text{with probability } \alpha \\ v_{t-1}, & \text{with probability } 1-\alpha \end{cases}$
  end
end
Return the Markov Chain $\{\text{Var}_t; t = 0, \ldots, N\}$.

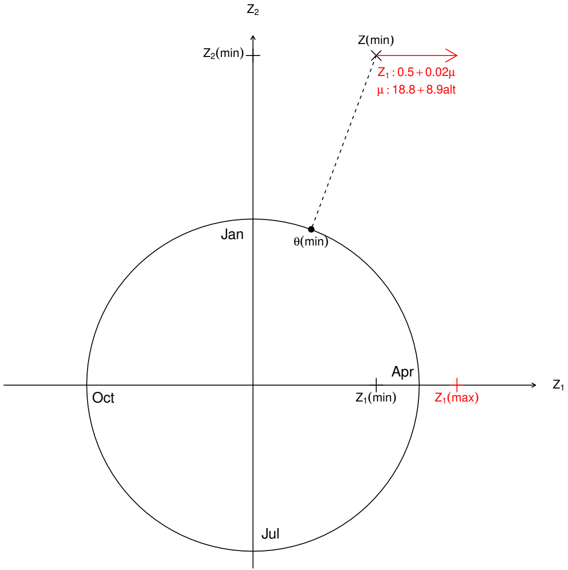
Impact of the altitude on the time of occurrence



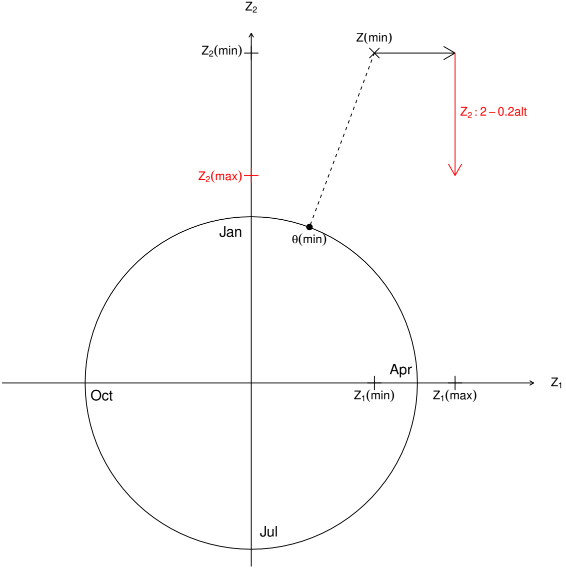
Impact of the altitude on the time of occurrence



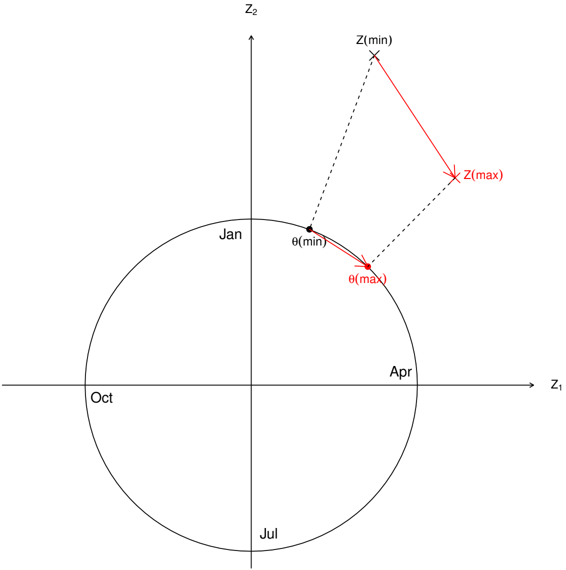
Impact of the altitude on the time of occurrence



Impact of the altitude on the time of occurrence



Impact of the altitude on the time of occurrence



Algorithm 2: Pointwise predictive posterior predictions from the extreme–angular model.

input : A Markov chain $\{\psi_t : t = 1, \ldots, N\}$ sampled from the Gibbs sampler introduced above and a new location $s_* \in \mathcal{X}$.

output: Predictive posterior predictions of an unknown quantity $U(s_*)$.

```
for t = 1, ..., N do
    /* Conditional sampling of the GEV parameters */
    Sample μ_t(s_*) from the conditional distribution μ(s_*) | μ(s), ψ_t;
    Sample σ_t(s_*) and ξ_t(s_*) in the same way;
    /* Conditional sampling for the angular component */
    Sample θ_t(s_*) from the conditional distribution
      θ(s_*) | μ(s), σ(s), ξ(s), ψ_t;
end
Return the U(s^*)-estimator Û(ψ_1(s_*), ..., ψ_N(s_*));
```