Statistical modelling of spatial extremes using the SpatialExtremes package

M. Ribatet

University of Montpellier

Rationale for the `SpatialExtremes` package

> "The aim of the `SpatialExtremes` package is to provide tools for the areal modelling of extreme events. The modelling strategies heavily rely on the extreme value theory and in particular block maxima techniques—unless explicitly stated."

As a consequence, most often

- the data used by the package have to be extreme—do not pass daily values for instance;
- the marginal distribution family is fixed, i.e., the generalized extreme value distribution family, but you have hands on how within this family parameters change in space;
- the process family is fixed, i.e., max-stable processes, but you have hands on which type of max-stable processes to use.

0. About the inner structure of max-stable processes

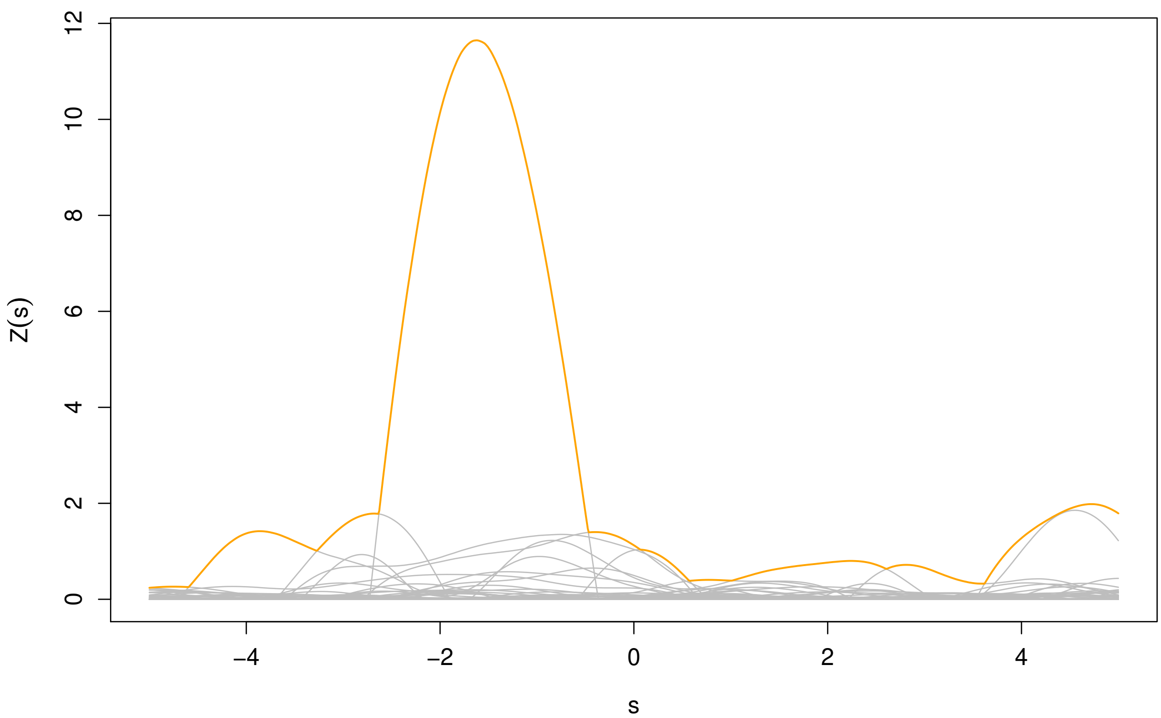
Spectral characterization

$$Z(s) = \max_{i\geq 1} \zeta_i Y_i(s), \qquad s \in \mathcal{X},$$

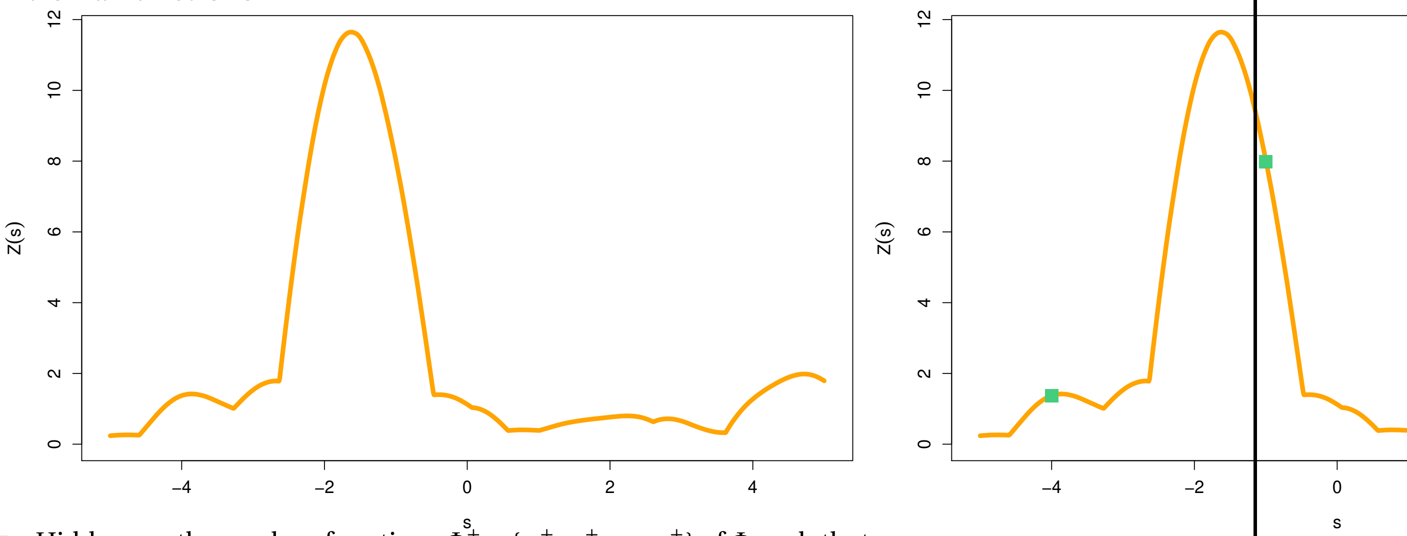
where $\{\zeta_i : i \geq 1\}$ is a Poisson point process on $(0, \infty)$ with intensity measure $\mathrm{d}\Lambda(\zeta) = \zeta^{-2}\mathrm{d}\zeta$ and Y_i independent copies of a (non-negative) stochastic process such that $\mathbb{E}\{Y(s)_+\} = 1$ for all $s \in \mathcal{X}$.

$$Z(s) = \max_{\varphi \in \Phi} \varphi(s), \qquad s \in \mathcal{X},$$

where $\Phi = \{\varphi_i : i \geq 1\}$ is a Poisson point process on $\mathbb{C}_0$ with an appropriate intensity measure.



Extremal functions



- Hidden are the random functions $\Phi^+ = \{\varphi_1^+, \varphi_2^+, \ldots, \varphi_k^+\}$ of Φ such that

$$\varphi_j^+(s_j) = Z(s_j), \qquad j = 1, \ldots, k, \qquad \text{(extremal functions)},$$

- and the random functions $\varphi^- \in \Phi \setminus \Phi^+$, i.e., satisfying

$$\varphi^-(s_j) < Z(s_j), \qquad j = 1, \ldots, k, \qquad \text{(sub-extremal functions)}$$

1. Data and descriptive analysis 6 / 39

Required data

Before introducing more advanced stuffs, let's talk about data format. It is pretty simple

Observations A numeric matrix such that each row is one realization of the spatial field—or if you prefer one column per site;

Coordinates A numeric matrix such that each row is the coordinates of one site—or if you prefer the first column is for instance the longitude of all sites, the second one latitude, …

```
> data
     Valkenburg  Ijmuiden  De Kooy  ...
1971        278        NA      360  ...
1972        334        NA      376  ...
1973        376        NA      365  ...
1974        314        NA      304  ...
1975        278        NA      278  ...
1976        350        NA      345  ...
1977        324        NA      298  ...
1978        298        NA      329  ...
1979        252        NA      298  ...
...
```

```
> coord
                 lon     lat
Valkenburg     4.419  52.165
Ijmuiden       4.575  52.463
De Kooy        4.785  52.924
Schiphol       4.774  52.301
Vlieland       4.942  53.255
Berkhout       4.979  52.644
Hoorn          5.346  53.393
De Bilt        5.177  52.101
...
```

Additional covariates

In addition to the storage of observations and coordinates, you might want to use additional covariates. The latter can be of two types

Spatial A numeric matrix such that each column corresponds to one spatial covariate such as elevation, urban/rural, ...

Temporal A numeric matrix such that each column corresponds to one temporal covariate such as time, annual mean temperature, ...

```
> spat.cov
             alt
Valkenburg  -0.2
Ijmuiden     4.4
De Kooy      0.5
Schiphol    -4.4
Vlieland     0.9
Berkhout    -2.5
Hoorn        0.5
De Bilt      2.0
...
```

```
> temp.cov
        nao
1971   1.87
1972   1.57
1973  -0.20
1974  -0.95
1975  -0.46
1976   2.34
1977  -0.49
1978   0.70
1979   1.11
...
```

✍ *It is always a good idea to name your columns and rows.*

Inspecting data

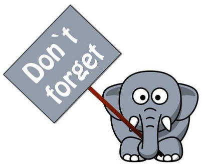
- As usual, you first have to scrutinize your data (weird values, encoding of missing values, check out factors, ...). But you're used to that, aren't you?
- We focus on extremes, so you may wonder
  - are my data extremes, i.e., block maxima?
  - is my block size relevant?
  - what about seasonality? Refine the block or use temporal covariate?
- You might want to check that the generalized extreme value family is sensible for your data—the `evd` package + a few lines of code will do the job for you (homework)
- This will generally be OK, but now you have to go a bit further by analyzing
  - the spatial dependence;
  - and the presence / absence of any spatial trends.

Spatial dependence [Cooley et al., 2006]

- Essentially you want to check if your data exhibit any (spatial) dependence. If not why would you bother with spatial models?
- The most convenient way to do this is through the F-madogram and its connection with the extremal coefficient:

$$\nu_F(h) = \frac{1}{2}\mathbb{E}[|F\{Z(o)\} - F\{Z(h)\}|], \qquad \theta(h) = \frac{1+2\nu_F(h)}{1-2\nu_F(h)}.$$

- The `fmadogram` function will estimate (empirically) the pairwise extremal coefficent from the F–madogram.



$$\theta(h) = -z \log \Pr\{Z(s) \le z, Z(s+h) \le z\}$$

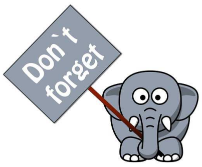
and that $1 \le \theta(h) \le 2$ with complete dependence iff $\theta(h) = 1$ and independence iff $\theta(h) = 2$.

Spatial dependence (2) [Dombry et al., 2017]

- Another recent summary measure of the spatial dependence is the extremal concurrence probability function $p\colon h \mapsto p(h) \in [0,1]$ where

$$p(h) := p(s, s+h) = \Pr\{\exists!\varphi \in \Phi\colon \varphi(s) = Z(s),\ \varphi(s+h) = Z(s+h)\},$$

i.e., there is a single extremal function at position s and $s+h$.



$$\begin{aligned} p(h) &= \mathbb{E}[\text{sign}\{Z(s) - \tilde{Z}(s)\}\text{sign}\{Z(s+h) - \tilde{Z}(s+h)\}] \\ &= \text{Kendall's } \tau, \end{aligned}$$

and that $0 \le p(h) \le 1$ with complete dependence iff $p(h) = 1$ and independence iff $p(h) = 0$.

☞ Note that here we assume a stationnary dependence structure!

The `fmadogram` function

- Run the file `fmadogram.R`. You should get the figure below.
- You can also use a binned version with `n.bins = 300`...

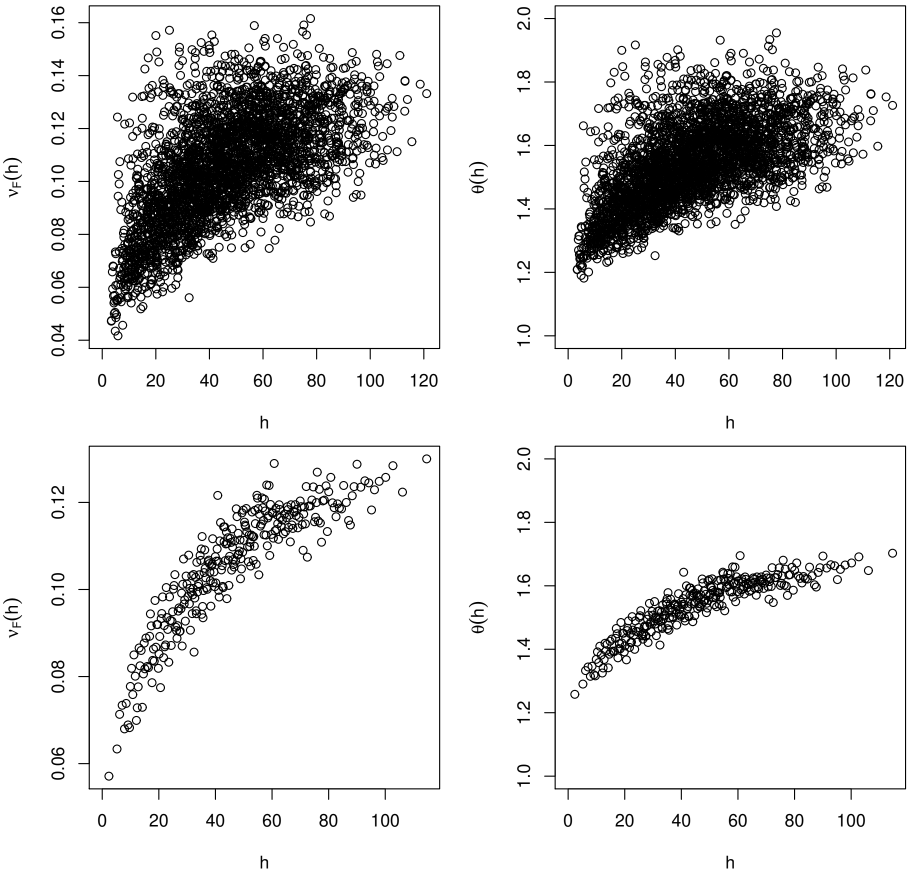


Figure 1: *Use of the* `fmadogram` *function to assess the spatial dependence.*

The `concprob` function

- □ Run the file `concprob.R`. You should get the figure below.
- □ You can also use a binned version with `n.bins = 300`...

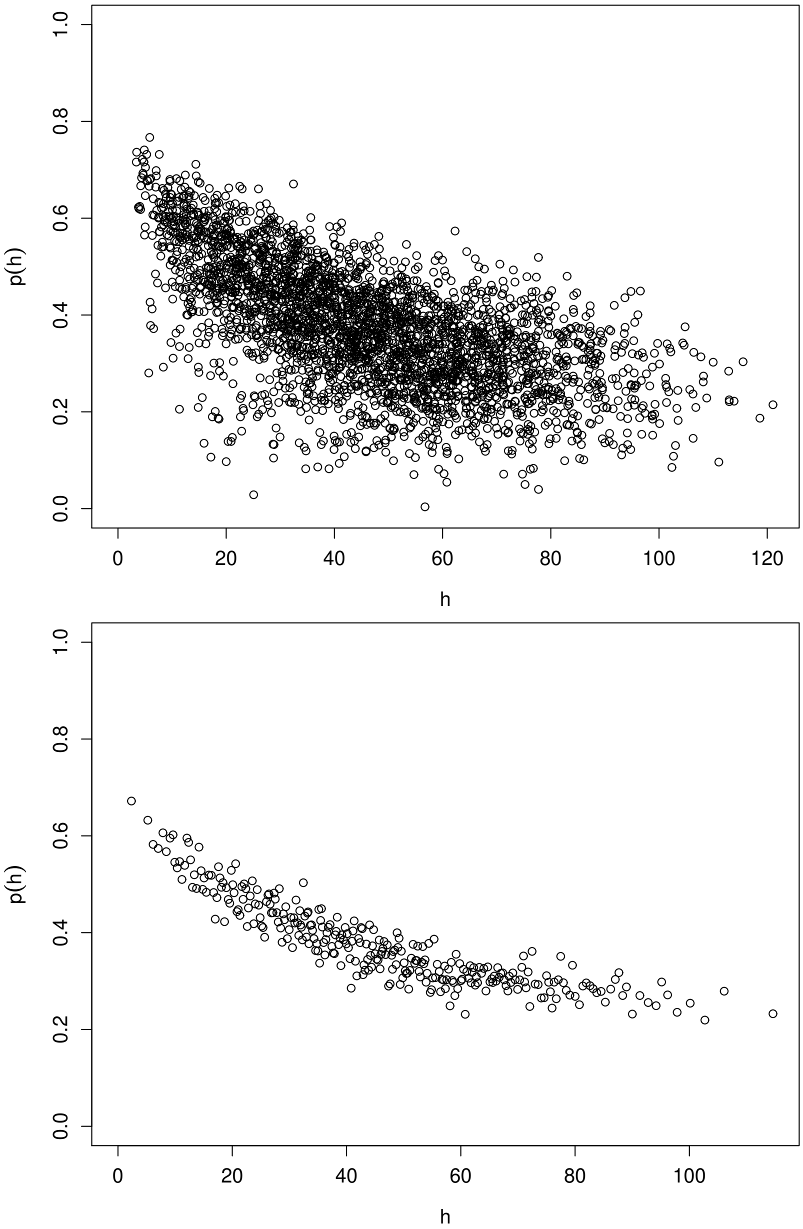


Figure 2: *Use of the* `concprob` *function to assess the spatial dependence.*

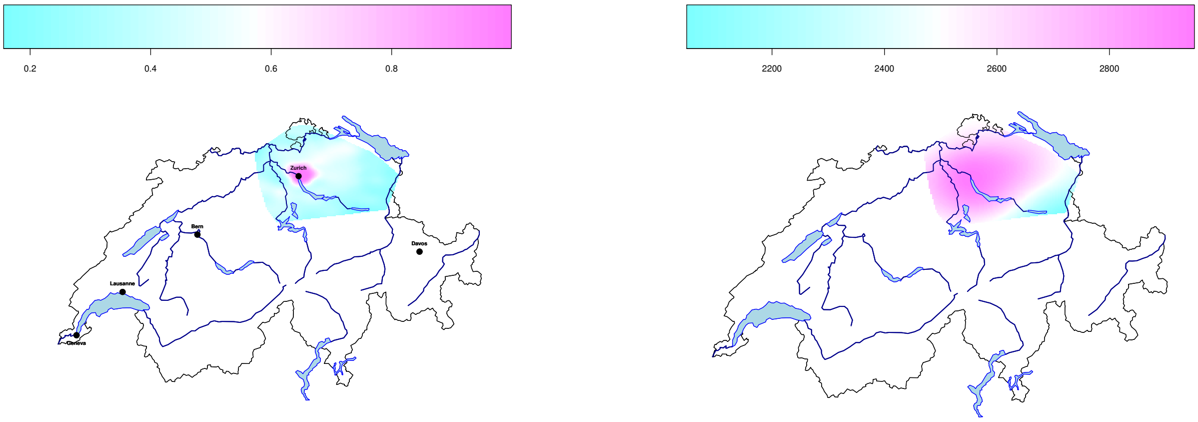
As an aside (`AsAnAside.R`) [Dombry et al., 2017]

- We can estimate the spatial distribution of the extremal concurrence probability w.r.t. a given weather station, e.g., plotting

$$\left\{(s, p(\text{Zurich}, s)) : s \in \mathcal{X}\right\},$$

- or estimate the expected area of concurrence cells

$$A(s_0) = \mathbb{E}\left\{\int_{\mathcal{X}} 1_{\{s_0 \text{ and } s \text{ are concurrent}\}} \mathrm{d}s\right\} = \int_{\mathcal{X}} p(s_0, s) \mathrm{d}s$$



Spatial trends

- We can do a symbol plot see the file `SpatialTrends.R`.

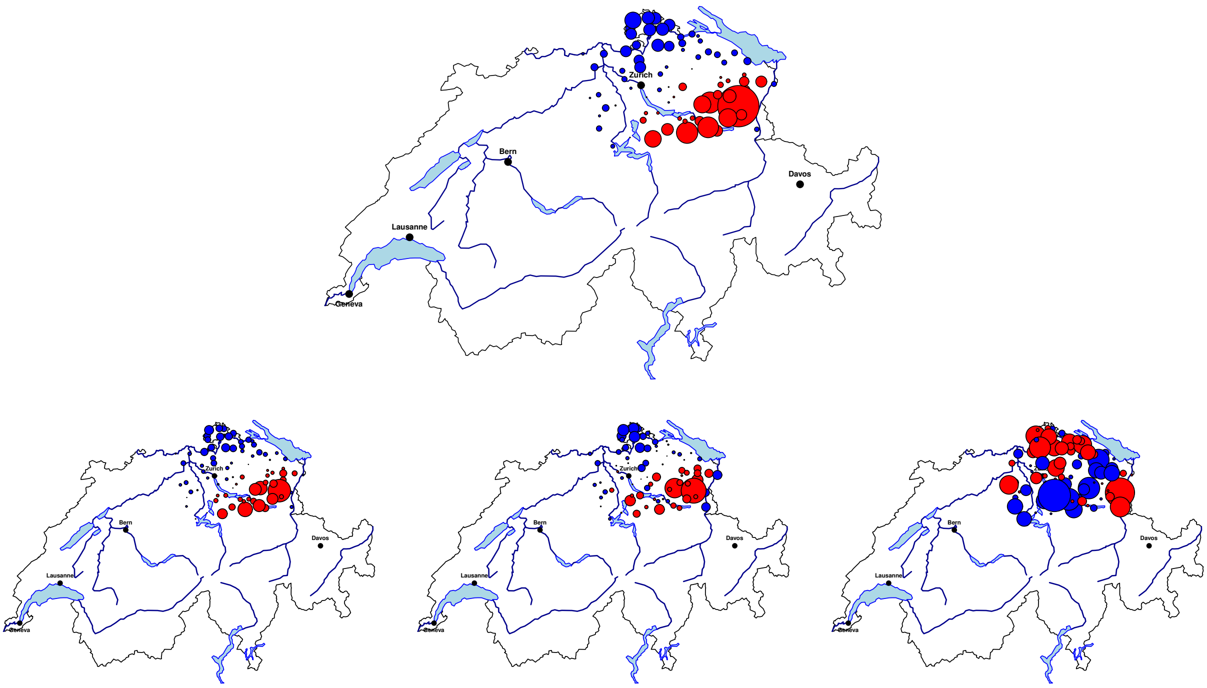


Figure 3: *Symbol plot for the swiss precipitation data.*

✍ *When exporting figures into eps/pdf, always pay attention to the aspect ratio.*

What we have learned so far (apart from using SpatialExtremes)

- The data exhibit some spatial dependence and there is still some (weak) dependance at a separation lag of 100km.
- There's a clear north-west / south-east gradient in the intensities of rainfall storms.
- In conclusion it makes sense to use max-stable processes whose marginal parameters are not constant across space.
- More specifically, we have:
  - a clear north-west / south-east gradient for the location and scale parameters;
  - no clear pattern for the shape parameter.

2. Simple max-stable processes

17 / 39

Max-stable models

- In this section we focus only on the spatial dependence and so assume that the margins are known and unit Fréchet—this is a standard choice in extreme value theory.
- From the spectral characterization

$$Z(s) = \max_{i\geq 1} \zeta_i Y_i(s), \qquad s \in \mathcal{X},$$

we can propose several parametric models for spatial extremes. Hence by letting Y to be

Gaussian densities with random displacements we get the Smith process;
Gaussian we get the Schlather process;
Log-normal (with a drift) we get the Brown–Resnick process;
Gaussian but elevated to some power we get the Extremal-t process.

Dependence parameters

Smith Elements of the covariance matrix appearing in the Gaussian densities;
Schlather Parameters of the correlation function;
Brown–Resnick Parameters of the semi-variogram;
Extremal–t Parameters of the correlation function and degrees of freedom.

- Since the margins are fixed, we only need to get estimates for the dependence parameters.
- How can we do that?

Least squares (`leastSquares.R`) **[Smith, 1990]**

$$\operatorname*{arg\,min}_{\psi \in \Psi} \sum_{1 \le i < j \le k} \left\{ \theta(s_j - s_j; \psi) - \hat{\theta}(s_i - s_j) \right\}^2,$$

where $\theta(\cdot;\psi)$ is the extremal coefficient obtained from the max-stable model with dependence parameters set to ψ and $\hat{\theta}(\cdot)$ is any empirical estimates of the extremal coefficient, e.g., F–madogram based.

```
> M0
        Estimator: Least Squares
            Model: Schlather
         Weighted: TRUE
  Objective Value: 3592.429
Covariance Family: Whittle-Matern

Estimates
  Marginal Parameters:
  Assuming unit Frechet.

  Dependence Parameters:
  range    smooth
54.3239   0.4026

Optimization Information
  Convergence: successful
  Function Evaluations: 61
```

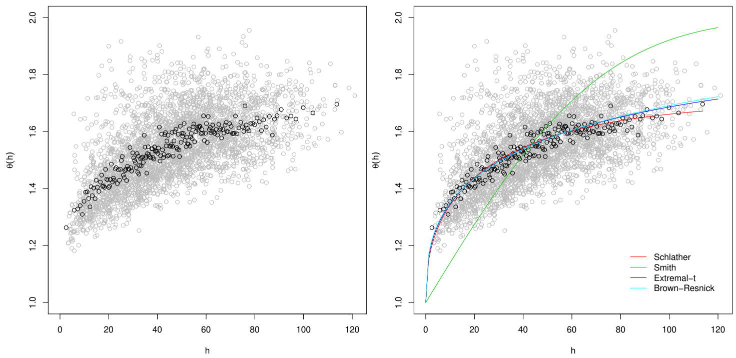


Figure 4: *Fitting simple max-stable processes from least squares.*

Pairwise likelihood (`pairwiseLlik.R`) **[Padoan et al., 2010]**

$$\operatorname*{arg\,max}_{\psi \in \Psi} \sum_{\ell=1}^{n} \sum_{1 \le i < j \le k} \log f\{z_\ell(s_i), z_\ell(s_j); \psi\},$$

where $f(\cdot,\cdot;\psi)$ is the bivariate density of the considered max-stable model.

```
        Estimator: MPLE
            Model: Schlather
         Weighted: FALSE
   Pair. Deviance: 1136863
              TIC: 1137456
Covariance Family: Whittle-Matern

Estimates
  Marginal Parameters:
  Assuming unit Frechet.

  Dependence Parameters:
  range    smooth
50.1976   0.3713

Standard Errors
  range    smooth
20.7085   0.0789

Asymptotic Variance Covariance
        range       smooth
range   428.841018   -1.570081
smooth   -1.570081    0.006225
Optimization Information
  Convergence: successful
  Function Evaluations: 67
```

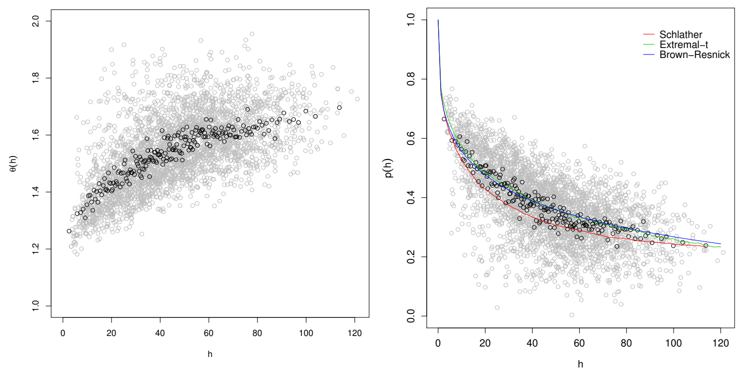


Figure 5: *Fitting simple max-stable processes maximizing pairwise likelihood.*

Model Selection [Varin and Vidoni, 2005]

- The advantage of the pairwise likelihood estimator over the least squares one is that you can do model selection.
- For instance one can use the TIC, Takeuchi Information Criterion or sometimes known as CLIC, Composite Likelihood Information Criterion,

$$\text{TIC} = 2\ell_{\text{pairwise}}(\hat{\psi}) - 2\text{tr}\{J(\hat{\psi})H^{-1}(\hat{\psi})\},$$

  $H(\hat{\psi}) = \mathbb{E}\{\nabla^2 \ell_{\text{pairwise}}(Y;\hat{\psi})\}$, $J(\hat{\psi}) = \text{Var}\{\nabla \ell_{\text{pairwise}}(Y;\hat{\psi})\}$.
- From our previous fitted models, we get

```
> TIC(M0,M1,M2)
     M1      M2      M0
1133660 1134823 1137449
```


Simulating simple max-stable processes (simulation.R) [Schlather, 2002; Dombry et al., 2016]

- Once you have fitted a suitable model, you usually want to simulate from it.
- Simulation from max-stable models is rather complex, recall that

$$Z(s) = \max_{i\geq 1} \zeta_i Y_i(s), \qquad s \in \mathscr{X}.$$

```
sim <- rmaxstab(n.obs, cbind(x, y), "twhitmat", DoF = 4,
+               nugget = 0, range = 3, smooth = 1)
> sim
          [,1]      [,2]      [,3]      [,4]      [,5]
 [1,] 3.8048914 0.4767980 6.3613989 1.4548317 1.0433912
 [2,] 1.2200332 0.6711422 0.8078701 2.0928629 0.7537061
 [3,] 0.5466466 2.0498561 4.8852572 2.3497976 0.6857268
 [4,] ...
```

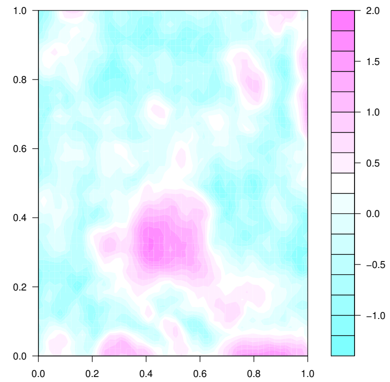


Figure 6: *One simulation on a 50 x 50 grid from the extremal–t model. (log scale)*

What we have learned so far (apart from using SpatialExtremes)

- The Smith model is clearly not a sensible model for our data—because of its linear behaviour near the origin;
- Schlather, Brown–Resnick and Extremal–t seems relevant;
- According to the TIC, the Extremal–t should be preferred.

From generalized extreme value margins to unit Fréchet ones

- Alright! We are able to handle the spatial dependence, but we assume that our data have unit Fréchet margins. This is not realistic at all!
- Fortunately, if $Y \sim \text{GEV}(\mu, \sigma, \xi)$ then

$$Z = \left(1 + \xi \frac{Y - \mu}{\sigma}\right)^{1/\xi} \sim \text{Unit Fréchet} = \text{GEV}(1, 1, 1).$$

- And since we are extreme value and spatial guys

$$Z(s) = \left\{1 + \xi(s) \frac{Y(s) - \mu(s)}{\sigma(s)}\right\}^{1/\xi(s)}, \qquad s \in \mathcal{X},$$

  is a simple max-stable process.
- Hence we can use the maximum pairwise likelihood estimator as before—up to an additional Jacobian term.

Omitting the spatial dependence

- With simple max-stable models, we omitted the marginal parameters.
- Here we will omit the spatial dependence for a while and consider locations as being mutually independent, i.e., use independence likelihood

$$\arg\max_{\psi \in \Psi} \sum_{i=1}^{k} \ell_{\text{GEV}}\{y(s_i); \psi\}.$$

- This is a kind of "*spatial GEV*" where ψ is a vector of marginal parameters.

Defining trend surfaces

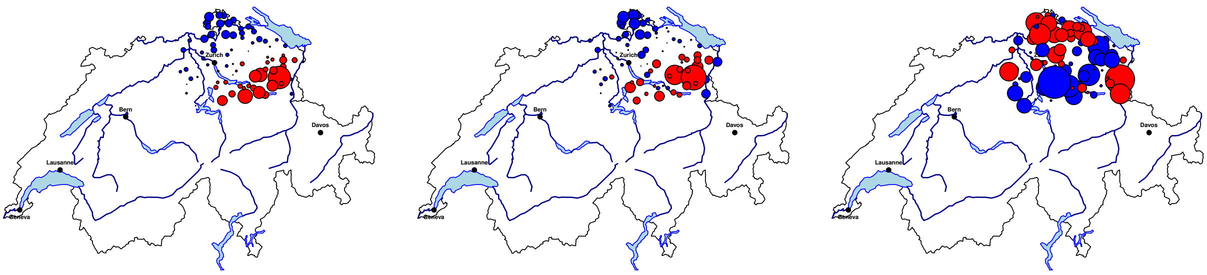


Figure 7: *Symbol plot for the swiss precipitation data.*

This suggests that

$$\begin{aligned}
\mu(s) &= \beta_{0,\mu} + \beta_{1,\mu}\text{lon}(s) + \beta_{2,\mu}\text{lat}(x) + \beta_{3,\mu}\text{lon}(s) \times \text{lat}(s),\\
\sigma(s) &= \beta_{0,\sigma} + \beta_{1,\sigma}\text{lon}(s) + \beta_{2,\sigma}\text{lat}(s) + \beta_{3,\sigma}\text{lon}(s) \times \text{lat}(s),\\
\xi(s) &= \beta_{0,\xi},
\end{aligned}$$

or equivalently with the `R` language

```
loc.form <- scale.form <- y ~ lon * lat; shape.form <- y ~ 1
```


Fitting the *spatial GEV* model (`spatialGEV.R`) [Davison et al., 2012]

```
      Model: Spatial GEV model
   Deviance: 29303.81
        TIC: 29499.38

    Location Parameters:
locCoeff1  locCoeff2  locCoeff3  locCoeff4
   27.132      1.846     -3.656     -1.080
       Scale Parameters:
scaleCoeff1  scaleCoeff2  scaleCoeff3  scaleCoeff4
     9.7850       0.7023      -1.0858      -0.5531
       Shape Parameters:
shapeCoeff1
     0.1572

Standard Errors
  locCoeff1    locCoeff2    locCoeff3    locCoeff4  scaleCoeff1  scaleCoeff2
    1.13326      0.34864      0.45216      0.38361      0.76484      0.28446
scaleCoeff3  scaleCoeff4  shapeCoeff1
    0.31267      0.27566      0.05878

Asymptotic Variance Covariance
             locCoeff1   locCoeff2   locCoeff3   locCoeff4   scaleCoeff1
locCoeff1     1.2842711   0.1131400  -0.1740921  -0.0729564   0.6570988
locCoeff2     0.1131400   0.1215498  -0.0623759   0.0149596   0.0521630
locCoeff3    -0.1740921  -0.0623759   0.2044448   0.0576622  -0.1086629
locCoeff4    -0.0729564   0.0149596   0.0576622   0.1471593  -0.0346376
scaleCoeff1   0.6570988   0.0521630  -0.1086629  -0.0346376   0.5849729
...

Optimization Information
  Convergence: successful
  Function Evaluations: 2135
```

Get predictions (`predictionSpatialGEV.R`)

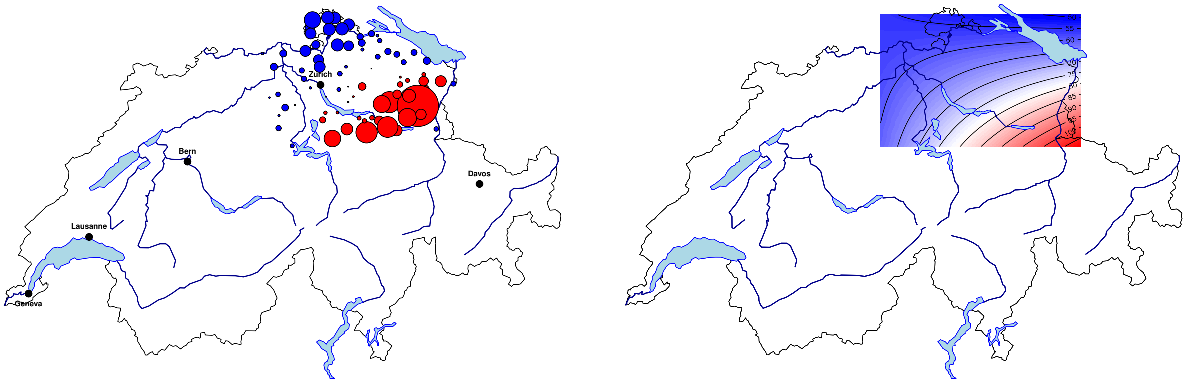


Figure 8: *Left: symbol plot. Right: Prediction of the pointwise 25-year return levels from a fitted spatial GEV model.*

- □ But don't we forget something???
- □ Model selection?

Model selection #2 (`modelSelection.R`) [Chandler and Bate, 2007; Kent, 1982]

- □ Typically here we would like to test if a given covariate is required or not
- □ Hence we're dealing with nested model for which composite likelihood ratio test are especially suited

$$2\{\ell_{\text{composite}}(\hat{\psi}) - \ell_{\text{composite}}(\hat{\phi}_{\lambda_0}, \lambda_0)\} \longrightarrow \sum_{j=1}^{p} \lambda_j X_i, \qquad n \to \infty.$$

```
Eigenvalue(s): 2.7 1.95

Analysis of Variance Table
   MDf Deviance Df  Chisq Pr(> sum lambda Chisq)
M2   7    29328
M0   9    29306  2 22.265               0.008273 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

✍ *Always check that your models are nested. The code won't do that for you!*

What we have learned so far (apart from using `SpatialExtremes`)

- □ Based on the `spatial GEV` model, we identify what seems to be relevant trend surfaces for the marginal parameters:

$$\mu(s) = \beta_{0,\mu} + \beta_{1,\mu}\text{lon}(s) + \beta_{2,\mu}\text{lat}(s) + \beta_{3,\mu}\text{lon}(s)\text{lat}(s),$$
$$\sigma(s) = \beta_{0,\sigma} + \beta_{1,\sigma}\text{lon}(s) + \beta_{2,\sigma}\text{lat}(s),$$
$$\xi(s) = \beta_{0,\xi},$$

Fitting a max-stable process with trend surfaces

- Now it's time to combine everything, i.e., trend surfaces + dependence.
- The syntax won't be a big surprise

```
M0 <- fitmaxstab(rain, coord[,1:2], "twhitmat", nugget = 0, loc.form, scale.form, shape.form)

       Estimator: MPLE
           Model: Extremal-t
        Weighted: FALSE
   Pair. Deviance: 2237562
             TIC: 2249206
Covariance Family: Whittle-Matern

Estimates
  Marginal Parameters:
    Location Parameters:
locCoeff1  locCoeff2  locCoeff3  locCoeff4
27.136295   0.060145  -0.164755  -0.001117
       Scale Parameters:
scaleCoeff1  scaleCoeff2  scaleCoeff3
    9.88857      0.02869     -0.04581
       Shape Parameters:
shapeCoeff1
     0.1727
  Dependence Parameters:
   range     smooth        DoF
225.9452     0.3645     4.1566
...
```


Model checking [Davison et al., 2012]

- When you want to check your fitted max-stable model, you usually want to check if
  - observations at each single location are well modelled: return level plot;
  - the dependence is well captured: extremal coefficient function.
- This can be done using a single line of code

```
> plot(M0)
```

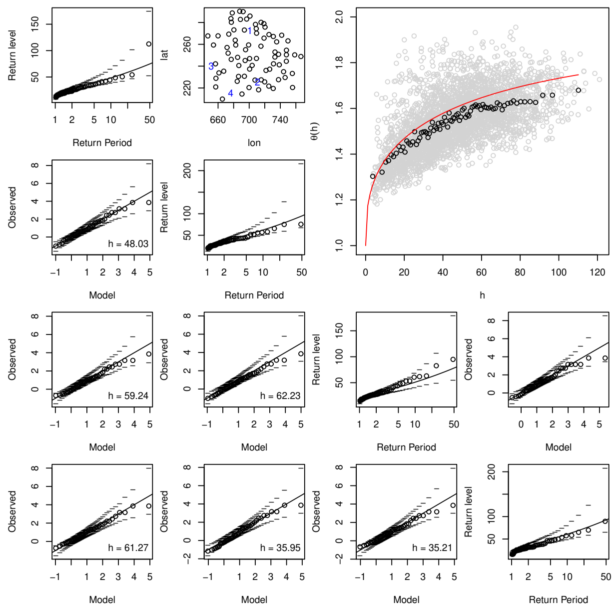


Figure 9: *Model checking for a fitted max-stable process having trend surfaces.*

Predictions (`simulationFinal.R`)

- Prediction works as for the *spatial GEV model* thanks to the `predict` function.
- But beware these predictions are pointwise—no spatial dependence at all!!!
- If you want to do take into account spatial dependence then you need to simulate from your fitted model.

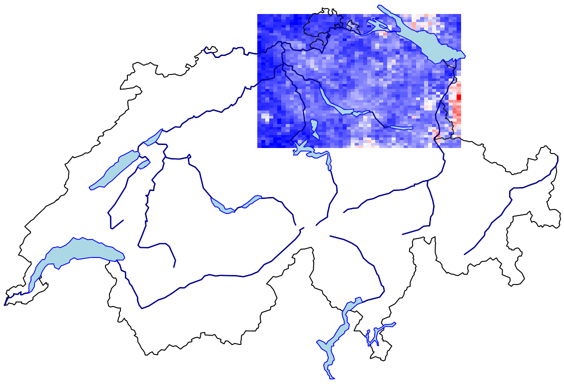


Figure 10: *One simulation from our fitted extremal–t model with trend surfaces.*

5. Conclusion

What we haven't seen

- Using weighted pairwise likelihood;
- Many (many!) utility functions. Highly recommended to have a look at the documentation;
- The package has a vignette: `vignette("SpatialExtremesGuide")`;
- Copula models—although I do not recommend their use for spatial extremes;
- Bayesian hierarchical models;
- Unconditional simulations: several implementations (including exact simulations)
- Conditional simulations—really CPU demanding.

☞ A rather recent review on max-stable processes with R code is given by Ribatet [2013]

THANK YOU!

References

- ✓ Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183
- ✓ Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Soulier, P., Doukhan, P., Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., and Zeger, S., editors, *Dependence in Probability and Statistics*, volume 187 of *Lecture Notes in Statistics*, pages 373–390. Springer New York
- ✓ Davison, A., Padoan, S., and Ribatet, M. (2012). Statistical modelling of spatial extremes. *Statistical Science*, 7(2):161–186
- ✓ Dombry, C., Engelke, S., and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317
- ✓ Dombry, C., Ribatet, M., and Stoev, S. (2017). Probabilities of concurrent extremes. Journal of the American Statistical Association (Theory & Methods), 113(524):1565–1582
- ✓ Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association (Theory & Methods)*, 105(489):263–277
- ✓ Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique*, 154(2):156–177
- ✓ Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44
- ✓ Smith, R. L. (1990). Max-stable processes and spatial extreme. *Unpublished manuscript*
- ✓ Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528