Knowledge for Climate

Regional peaks-over-threshold modeling with respect to climate change

Martin Roth (KNMI / EURANDOM)

joint work with A. Buishand (KNMI), G. Jongbloed (TU Delft), A. Klein Tank (KNMI), and H. v. Zanten (TU Eindhoven)

9 September 2012

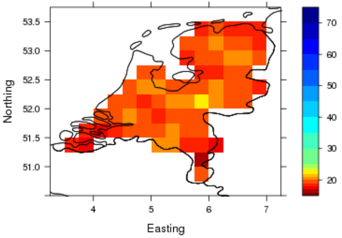
Goals

- Estimate site specific quantiles / return levels
- Assess the temporal trends in these quantiles
- Reduce the estimation uncertainty by spatial pooling

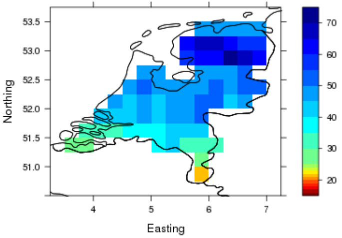
Inspired by the work of M. Hanel, A. Buishand and C. Ferro (2009) for block maxima data.

Data

- Daily, gridded precipitation data (E-OBS v. 5.0)
- Netherlands (high station density)
- Winter (DJF) data from 1950 - 2010

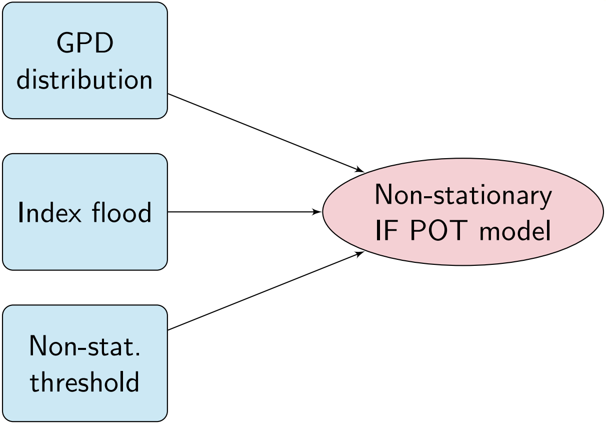


Mean winter maxima



Event on December 3, 1960

Components of the Model



GEV and GPD

Generalized Extreme Value distribution (GEV) for block maxima (BM)

$$
\begin{aligned}
P(M \leq x) &= H_{\xi^*,\sigma^*,\mu^*}(x) \\
&= \begin{cases} \exp\left\{ -\left[1 + \xi^* \left(\frac{x-\mu^*}{\sigma^*} \right) \right]^{-1/\xi^*} \right\}, & \xi^* \neq 0, \\ \exp\left[-\exp(-\frac{x-\mu^*}{\sigma^*}) \right], & \xi^* = 0, \end{cases}
\end{aligned}
$$

Generalized Pareto distribution (GPD) for excesses

$$
\begin{aligned}
P(Y \leq y | Y \geq 0) &= G_{\xi,\sigma}(y) \\
&= \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma} \right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\sigma} \right), & \xi = 0, \end{cases}
\end{aligned}
$$

GEV/GPD relation

If, for a threshold u, the excesses follow a GPD distribution with shape ξ and scale σ (denoted by $G_{\xi,\sigma}$) and the exceedance times follow a Poisson process with intensity λ, then we have that the maxima above u are GEV distributed with the following parameters:

$$\begin{aligned}
\mu^* &= \begin{cases} u - \frac{\sigma}{\xi}(1-\lambda^{\xi}), & \xi \neq 0, \\ u + \sigma \ln(\lambda), & \xi = 0, \end{cases} \\
\sigma^* &= \sigma\lambda^{\xi}, \\
\xi^* &= \xi,
\end{aligned} \tag{1}$$

Index flood for POT I

The index flood method assumes that all site specific distributions are identical apart from a site specific scaling factor, the index variable[1]. For exceedances this means, that

$$P\left(\frac{X_s}{\eta_s} \leq x | X_s \geq u_s\right) = \psi(x) \quad \forall s \in \mathcal{S}, \tag{2}$$

where X_s is a random variable representing the site-specific daily precipitation, u_s is the site specific threshold, η_s is the index variable and ψ does not depend on site s.

[1] Hosking and Wallis (1997)

Index flood for POT I

The index flood method assumes that all site specific distributions are identical apart from a site specific scaling factor, the index variable[1]. For exceedances this means, that

$$P\left(\frac{X_s}{\eta_s} \leq x | X_s \geq u_s\right) = \psi(x) \quad \forall s \in \mathcal{S}, \tag{2}$$

where X_s is a random variable representing the site-specific daily precipitation, u_s is the site specific threshold, η_s is the index variable and ψ does not depend on site s.

[1] Hosking and Wallis (1997)

Index flood for POT II

Knowledge for Climate

Index variable equals threshold

$$\psi(u_s/\eta_s) = P\left(X_s \leq u_s | X_s \geq u_s\right) = 0, \qquad \forall s \in \mathcal{S}.$$
$$\Rightarrow u_s/\eta_s = c.$$

Without loss of generality we can set $\eta_s = u_s$.

Index flood also for the excesses

$$P\left(\frac{Y_s}{\eta_s} \leq y | Y_s \geq 0\right) = \widetilde{\psi}(y) \quad \forall s \in \mathcal{S}, \tag{3}$$

where $\widetilde{\psi}(y) := \psi(y+1)$ is independent of site s.

Index flood for POT III

Site specific threshold

The τ-th quantile ($\tau >> 0.9$) of the daily precipitation amounts is a natural choice for a site specific threshold.
$\Rightarrow$ λ_s will be approximately constant over the region.

Restriction on the GPD parameters

The distribution of the scaled excesses has the following form:

$$P\left(\frac{Y_s}{\eta_s} \leq y \middle| Y_s \geq 0\right) = G_{\xi_s, \frac{\sigma_s}{u_s}}(y) \equiv \widetilde{\psi}(y). \tag{4}$$

Therefore we have:

$$\frac{\sigma_s}{u_s} \equiv \gamma, \qquad \xi_s \equiv \xi \quad \forall s \in \mathcal{S}. \tag{5}$$

We refer to γ as the dispersion coefficient.

Index flood for block maxima

Assuming constant λ, γ and ξ gives for the GEV parameters:

$$\xi_s^* \equiv \xi$$

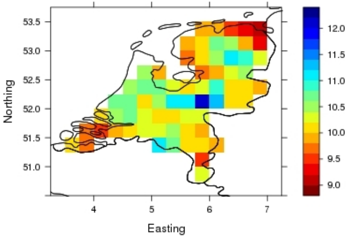
$$\gamma_s^* := \frac{\sigma_s^*}{\mu_s^*} = \left\{ \begin{array}{ll} \frac{\lambda^\xi}{\gamma^{-1} - \frac{1}{\xi}(1-\lambda^\xi)}, & \xi \neq 0 \\ \frac{1}{\gamma^{-1} + \ln(\lambda)}, & \xi = 0. \end{array} \right\} \equiv \gamma^*$$

Therefore the parameters ξ^* and γ^* fulfill the IF assumption for BM data[2]. This does not apply for the IF model for POT data proposed by Madsen and Rosbjerg (1997).

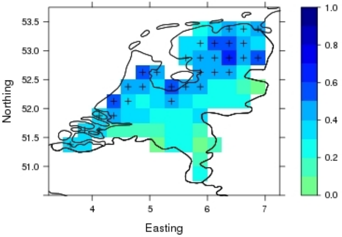
[2] Hanel, Buishand and Ferro (2009)

Nonstationary Threshold

The threshold is determined as the 0.96 linear regression quantile[3]:



Mean of the threshold for the 1950–2010 period in mm.



Trend in the threshold for the 1950–2010 period in mm per decade.

[3]Koenker (2005)

Nonstationary Version of the IF Model

Knowledge for Climate

IF restrictions on the GPD parameters

$$\xi_s(t) \equiv \xi(t), \quad \frac{\sigma_s(t)}{u_s(t)} \equiv \gamma(t).$$

Quantile estimates

$$\begin{aligned} q_\alpha(s,t) &= u_s(t) + G^{-1}_{\xi(t),\sigma_s(t)}\big(1 - \frac{\alpha}{\lambda}\big) \\ &= \begin{cases} u_s(t) \cdot \Big(1 - \frac{\gamma(t)}{\xi(t)}\big[1 - \big(\frac{\alpha}{\lambda}\big)^{-\xi(t)}\big]\Big), & \xi(t) \neq 0, \\ u_s(t) \cdot \big(1 + \gamma(t)\ln(\lambda/\alpha)\big), & \xi(t) = 0. \end{cases} \end{aligned}$$

Note the factorization into a time and site dependent index variable and a quantile function, which depends on time only.

Nonstationary Version of the IF Model

Knowledge for Climate

IF restrictions on the GPD parameters

$$\xi_s(t) \equiv \xi(t), \quad \frac{\sigma_s(t)}{u_s(t)} \equiv \gamma(t).$$

Quantile estimates

$$\begin{aligned} q_\alpha(s,t) &= u_s(t) + G^{-1}_{\xi(t),\sigma_s(t)}\big(1-\frac{\alpha}{\lambda}\big) \\ &= \begin{cases} u_s(t)\cdot\big(1-\frac{\gamma(t)}{\xi(t)}\big[1-\big(\frac{\alpha}{\lambda}\big)^{-\xi(t)}\big]\big), & \xi(t)\neq 0, \\ u_s(t)\cdot\big(1+\gamma(t)\ln(\lambda/\alpha)\big), & \xi(t)=0. \end{cases} \end{aligned}$$

Note the factorization into a time and site dependent index variable and a quantile function, which depends on time only.

Independence Likelihood

- Maximum likelihood estimation (MLE) popular method
- Since late 1980s used for regional estimation approaches
- Difficult dependence structure was neglected using an artificial independence assumption (independence likelihood)
- Dependence influences mainly the uncertainty
- Smith (1990) studies the uncertainty in an extended manner
- Special case of composite likelihood[4], which is a class of simplified (not true) likelihoods (e.g. also pairwise likelihood)

[4] Varin, Reid and Firth (2011)

Independence Likelihood

- Maximum likelihood estimation (MLE) popular method
- Since late 1980s used for regional estimation approaches
- Difficult dependence structure was neglected using an artificial independence assumption (independence likelihood)
- Dependence influences mainly the uncertainty
- Smith (1990) studies the uncertainty in an extended manner
- Special case of composite likelihood[4], which is a class of simplified (not true) likelihoods (e.g. also pairwise likelihood)

[4] Varin, Reid and Firth (2011)

Composite Likelihood

- Allows to assess for spatial dependence
- Specify a certain structure for the parameters, e.g.

$$\gamma(t) = \gamma_1 + \gamma_2 \cdot (t - \bar{t}), \qquad \xi(t) = \xi_1.$$

- Maximize:

$$\ell_I(\theta) = \sum_{s=1}^{S} \sum_{\substack{t=1 \\ y_s(t) \geq 0}}^{T} \ln\Big(f_{\underbrace{\gamma(t) u_s(t)}_{\sigma_s(t)}, \xi(t)}(y_s(t)) \Big),$$

where $f_{\sigma,\xi}(y)$ is the density of the GPD distribution.

Asymptotic Normality

$\hat{\theta}_I$ is asymptotically normal with mean θ and covariance matrix $G^{-1}(\theta)$

Godambe (sandwich) information

$$G(\theta) = H(\theta) J^{-1}(\theta) H(\theta)$$

- $H(\theta)$ is the expected negative Hessian of $\ell_I(\theta, \mathbf{Y})$
  Fisher information or sensitivity matrix
- $J(\theta)$ is the covariance matrix of the score $\nabla_\theta \ell_I(\theta, \mathbf{Y})$
  referred to as variability matrix
- In the independent case we have

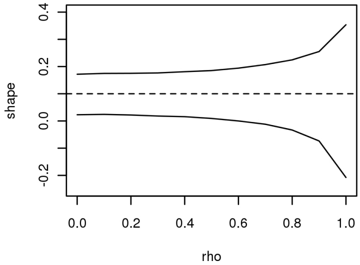
$$J(\theta) = H(\theta) \Rightarrow G(\theta) = H(\theta)$$

Simulation

- Marginal parameters:
  $u \sim \mathcal{N}(10, 1)$, $\xi = 0.1$ and $\gamma = 0.5$
- Dependence model: Normal copula with auto-regressive correlation structure governed by one parameter rho
- Dimension: 10 sites and 100 (common) excesses and 2500 samples

Simulation

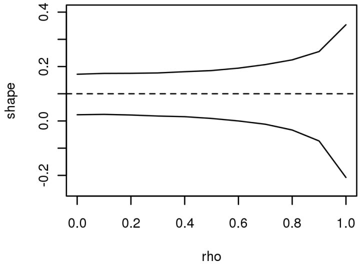
- Marginal parameters:
  $u \sim \mathcal{N}(10, 1)$, $\xi = 0.1$ and $\gamma = 0.5$
- Dependence model: Normal copula with auto-regressive correlation structure governed by one parameter rho
- Dimension: 10 sites and 100 (common) excesses and 2500 samples



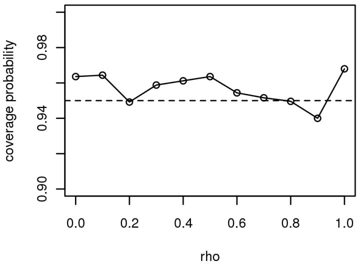
Mean confidence interval (shape parameter)

Simulation

- Marginal parameters:
  $u \sim \mathcal{N}(10, 1)$, $\xi = 0.1$ and $\gamma = 0.5$
- Dependence model: Normal copula with auto-regressive correlation structure governed by one parameter rho
- Dimension: 10 sites and 100 (common) excesses and 2500 samples



Mean confidence interval (shape parameter)



Coverage probability

Composite Information Criteria

Composite likelihood adaptations of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)

$$
\begin{aligned}
AIC &= -2\ell_I(\hat{\theta}_I, Y) + 2\dim(\theta), \\
BIC &= -2\ell_I(\hat{\theta}_I, Y) + \ln(n)\dim(\theta),
\end{aligned}
$$

where $\dim(\theta)$ is an effective number of parameters:

$$
\dim(\theta) = \mathrm{tr}\left(\hat{H}(\theta)\hat{G}(\theta)^{-1}\right),
$$

Composite Likelihood Ratio Test

Adaptation of the likelihood ratio test

$$W = 2\left[\ell_I\left(\hat{\theta}_{M_1}; y\right) - \ell_I\left(\hat{\theta}_{M_0}; y\right)\right].$$

The asymptotic distribution of W is given by a linear combination of independent χ^2 variables, and can be determined using the Godambe information.

Bootstrap

- Transform excesses to standard exponentials using the full model M_1
- Sample monthly blocks of the whole region
- Transform the sampled data back using the nested model M_0

Composite Likelihood Ratio Test

Adaptation of the likelihood ratio test

$$W = 2\left[\ell_I\left(\hat{\theta}_{M_1}; y\right) - \ell_I\left(\hat{\theta}_{M_0}; y\right)\right].$$

The asymptotic distribution of W is given by a linear combination of independent χ^2 variables, and can be determined using the Godambe information.

Bootstrap

- Transform excesses to standard exponentials using the full model M_1
- Sample monthly blocks of the whole region
- Transform the sampled data back using the nested model M_0

Models and Information Criteria

IF models used

Model	**dispersion** γ	**shape** ξ
no trend	γ_1	ξ_1
trend in dispersion	$\gamma_1 + \gamma_2 * (t - \bar{t})$	ξ_1
trend in shape	γ_1	$\xi_1 + \xi_2 * (t - \bar{t})$

Information criteria for the IF models

Model	AIC	BIC
no trend	78387.28	**78715.59**
trend in dispersion	78435.60	78880.41
trend in shape	**78333.28**	78748.95

Models and Information Criteria

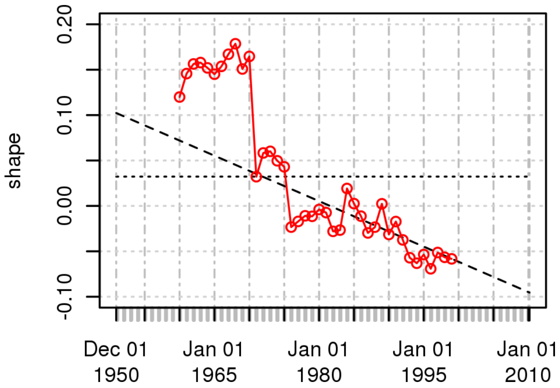
IF models used

Model	**dispersion** γ	**shape** ξ
no trend	γ_1	ξ_1
trend in dispersion	$\gamma_1 + \gamma_2 * (t - \bar{t})$	ξ_1
trend in shape	γ_1	$\xi_1 + \xi_2 * (t - \bar{t})$

Information criteria for the IF models

Model	AIC	BIC
no trend	78387.28	**78715.59**
trend in dispersion	78435.60	78880.41
trend in shape	**78333.28**	78748.95

Shape



Shape parameter for different models (dotted – constant, dashed – linear trend, solid red – 20 year moving window estimates)

Significance Tests

Trend in the GPD parameters

p-values of the test for trend in the GPD parameters

Model	asymptotic	bootstrap
trend in dispersion	82.9%	81.3%
trend in shape	26.7%	12.2%

Index flood assumption

We compare the composite likelihoods of an IF model without trend in the parameters with that of a model with site specific dispersion coefficient and common shape parameter using the bootstrap method. We obtain a *p*-value of 0.103, i.e. the IF assumption is not rejected.

Significance Tests

Trend in the GPD parameters

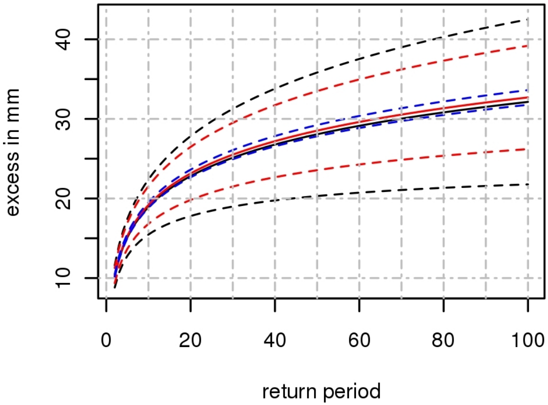
p-values of the test for trend in the GPD parameters

Model	asymptotic	bootstrap
trend in dispersion	82.9%	81.3%
trend in shape	26.7%	12.2%

Index flood assumption

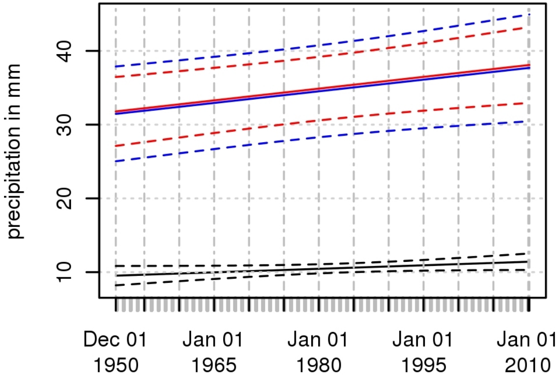
We compare the composite likelihoods of an IF model without trend in the parameters with that of a model with site specific dispersion coefficient and common shape parameter using the bootstrap method. We obtain a *p*-value of 0.103, i.e. the IF assumption is not rejected.

Uncertainty I - Excess distribution



Estimated return levels of the excesses (solid lines) with 95% pointwise confidence bands (dashed lines) for the year 1980 at the grid box around De Bilt (black – site-specific, red – IF, blue – no correction of the standard error for spatial dependence)

Uncertainty II - Threshold and Return Level



Knowledge for Climate

Estimated threshold with 95% pointwise confidence band (black) and 25-year return level based on the at-site estimation (blue) and the IF approach (red), together with pointwise confidence bands, for the grid box around De Bilt.

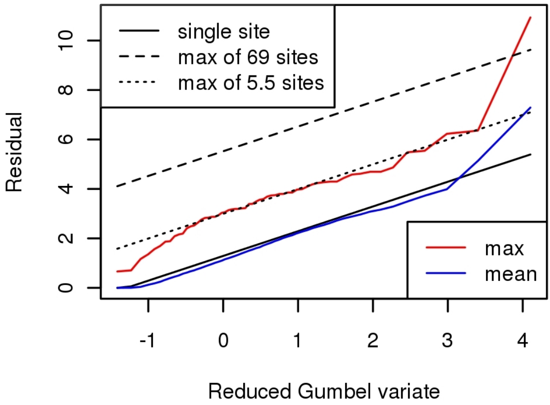
Spatial Dependence I

- Transform observed peaks to standard exponentials
- Consider the maximum $M_{s,j}$ for each site s and winter season j. This maximum is approximately Gumbel distributed with location parameter $\ln(\lambda)$ and scale parameter 1
- Determine the spatial mean Gumbel plot of these maxima
- Determine Gumbel plot of $\max_s M_{s,j}$ over the grid. For independent observations this should be Gumbel distributed with location parameter $\ln(\lambda * S)$ and scale parameter 1[5]

[5] Reed and Stewart (1994)

Spatial Dependence II

Knowledge for Climate



Spatial mean Gumbel plot (blue), Gumbel plot of the maxima over the grid (red) and theoretical distributions for the maximum of a different number of Gumbel variables (black)

Conclusions and Further Research

- Positive trends in the threshold are observed, which are significant in the coastal region
- No trend in the dispersion coefficient, i.e. proportional increase of the GPD scale parameter
- Negative trend in the shape parameter not significant
- Uncertainty is substantially reduced by regional modeling

- Application to climate model data
- Validity of the bootstrap needs to be explored

Literature

Hanel, M., T. A. Buishand, and C. A. T. Ferro (2009),
A nonstationary index flood model for precipitation extremes in transient regional climate model simulations.
J. Geophys. Res., 114(D15):D15107.

Hosking, J. R. M. and J. R. Wallis (1997),
Regional frequency analysis.
Cambridge University Press, Cambridge, UK.

Koenker, R. (2005)
Quantile Regression.
Cambridge University Press, Cambridge, UK.

Madsen, H. and D. Rosbjerg (1997),
The partial duration series method in regional index-flood modeling.
Water Resour. Res., 33(4):737–746.

Reed, D. W. and E. J. Stewart (1994),
Inter-site and inter-duration dependence of rainfall extremes.
in Barnett and Turkman *Statistics for the Environment 2*, Wiley, Chichester.

Smith, R. L. (1990),
Regional estimation from spatially dependent data.
Unpublished manuscript, University of Chapel Hill.

Varin, C., N. Reid, and D. Firth (2011),
An overview of composite likelihood methods.
Statistica Sinica, 21:5–42.