Modeling and prediction of extreme wind speeds

Pál RAKONCZAI and Nader TAJVIDI
Mathematical Statistics, Centre for Mathematical Sciences
Lund Institute of Technology, Sweden
pal@maths.lth.se and nader@maths.lth.se

Univariate case

For the univariate extremes there is natural finit-dimesional parametric family (1) but in the multivariate case the parametric family is infinite-dimensional.

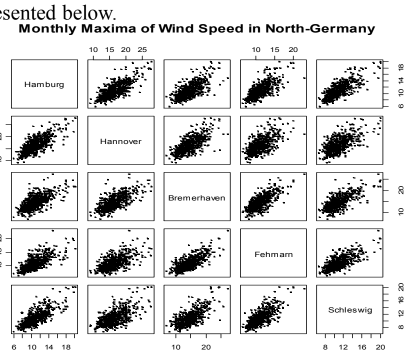
$$G_{\mu,\sigma,\xi}(x)=\exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) \qquad (1)$$

where $1+\xi\frac{x-\mu}{\sigma}>0$.

The most general solution for tackling this problem is if we handle the the **marginals** and the **dependence structure** separetely. Here we review some alternatives for the bivariate case and consider approaches to model the joint behavior of monthly maxima of wind speeds.

Data and Marginals

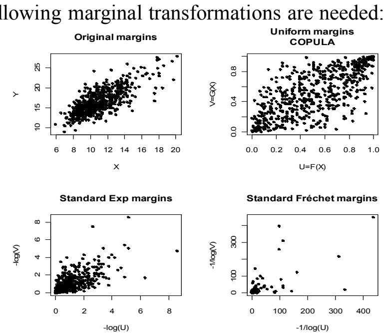
The **monthly wind speed maxima** (in m/s) over the 600 months of the recent 50 years have been coupled simultaneously for the different stations in North-Germany. These pairs of observations are presented below.



As an illustration Hamburg and Hannover have been chosen. The parameters of the fitted marginal distributions (1) as well as the results of hypothesis testing for the goodness of fit are provided in table below. Here we have used the **modified Anderson-Darling statistics** as test statistics focused on the upper tail, showing clearly that the fit is appropriate even for the highest quantiles as well.

City	ξ (shape)	μ (location)	σ (scale)	95% crit.values	AD statistics
Hamburg	-0.03	9.99	1.84	0.4109	>0.1729
Hannover	-0.08	15.14	2.63	0.3255	>0.1969

Finally before fitting the dependence structure the following marginal transformations are needed:



Dependence functions and Copulas

Define (X_i,Y_i) $i=1\ldots n$ a random sample of pairs representing the componentwise maxima of an iid. distributed sequence over a given period of time. Under the appropriate conditions these maxima follow a **bivariate extreme-value** distribution H and both of its margins are necessarily arising from the classical univariate extreme-value distributions F and G as in (1). After transforming the marginals into unit Fréchet distribution H can be characterized as

$$H(x,y)=\exp\left(-\left(\frac{1}{x}+\frac{1}{y}\right)A\left(\frac{x}{x+y}\right)\right) \qquad (2)$$

where $A(t)$ is called **dependence function.**

$$A_{\log}(t)=\left((1-t)^{\alpha}+t^{\alpha}\right)^{1/\alpha}$$

$$A_{asy.\log}(t)=\left(\theta(1-t)^{\alpha}+(\phi t)^{\alpha}\right)^{1/\alpha}+(\theta-\phi)t+1-\theta$$

$$A_{HT}(t)=\left(\frac{1}{n}\sum_{1=i}^{n}\left(\max\{(1-t)\hat{X}_i;t\hat{Y}_i\}\right)^{-1}\right)^{-1}$$

$$A_{CFG}(t)=e^{\left(\frac{1}{n}\sum_{1=i}^{n}(\max\{(1-t)X_i;tY_i\})-(1-t)\sum_{i=1}^{n}X_i\right)-t\sum_{i=1}^{n}Y_i}$$

Besides the dependence functions a fairly different kind of representation is possible. Any continuous bivariate distribution H with given marginals F and G can be written uniquely to the so called **copula form** where C is a bivariate distribution function having support $[0,1]^2$ and uniform marginal distributions:

$$H(x,y)=C(F(x),G(x)) \qquad (3)$$

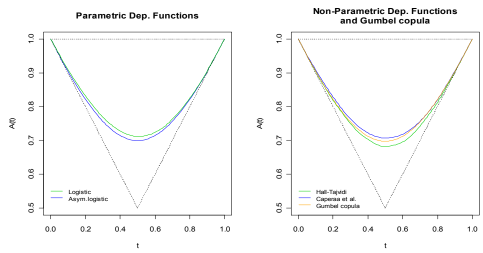
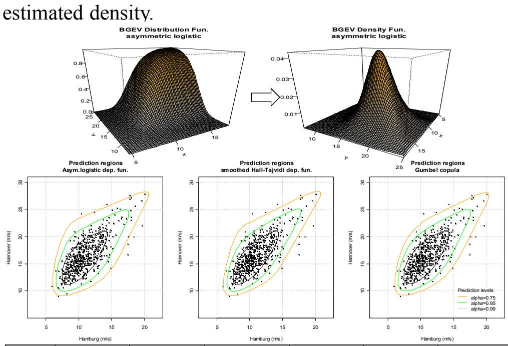


Fig.1 Estimates of dependence functions between the marginals from Hamburg and Hannover. Left: parametric estimates of the symmetric and the asymmetric logistic models. Right: the greatest convex minorants of two non-parametric estimates and the Gumbel copula model.

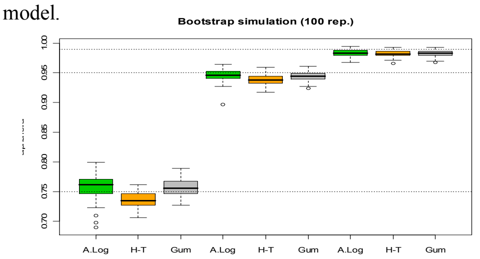
Results

BGEV: In order to complete the required bivariate models we pieced together the estimated marginals with the different dependence structures. For comparison **compact prediction regions** have been counted by integration of the estimated density.



Pred level	Expect. number	Logistic	A.logistic	Caperaa	Hall-Tajvidi	Gumbel
0.75	148	142	134	137	156	143
0.95	29	31	29	31	35	32
0.99	5	12	9	11	12	12

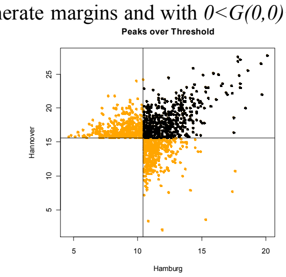
Bootstrap simulation: to avoid the bias caused by concluding only from one realization. The observations have been resampled 100 times with replacement, and considered as another possible wind maximum data to model.

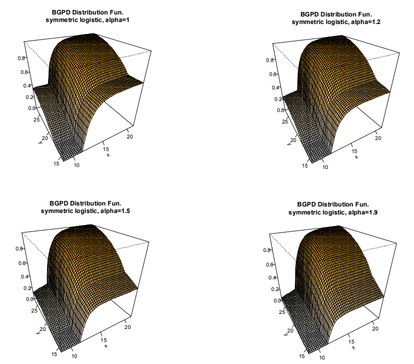


BGPD: Block maxima can hide the time structure within the months so we don not know if the different component maxima occurred simultaneously or not. To avoid this problem also the exceedances over a high threshold can be considered for the daily observations. This way we can describe what happens to the other component when the other exceed its threshold. We define the **bivariate generalized Pareto** distribution as

$$H(x,y)=\frac{1}{-\log G(0,0)}\log\frac{G(x,y)}{G(x\wedge 0,y\wedge 0)}$$

For some bivariate extreme value distribution G with non-degenerate margins and with $0<G(0,0)<1$.





References

[1]Bozso, D. Rakonczai P. and Zempleni A. (2005) **"Extreme value analysis: focusing on the fit and conditions, with hydrological applications"** Conference on 4th Extreme Value Analysis, Gothenburg, Sweden.
[2]Capeeraa, P. Fougeres, A.-L. and Genest, C. (1997) **"Nonparametric estimation procedure for bivariate extreme value copulas"** Biometrika 84(3), p. 567-577.
[3]Hall, P. and Tajvidi, N. (2004) **"Distribution and dependence function estimation for bivariate extreme value distributions"** Bernoulli 6(5), p. 835-844.
[4]Hall, P. and Tajvidi, N. (2000) **"Prediction regions for bivariate extreme events"** Aust. N. Z. J. Stat 46(1), p. 99-112.
[5]Rakonczai, P. and Zempleni, A. (2007) **"Copulas and goodness of fit tests"** Recent Advances in Stochastic Modeling and Data Analysis, p. 198-206.
[6]Rootzen, H. and Tajvidi, N. (2006) **"Multivariate generalized Pareto distributions"** Bernoulli, 12(5), p. 917-930