HS7.4/AS4.9/CL3.4: Weather Generators: Reviewing the State of the Art

Stephan Thober, Luis Samaniego, Sabine Attinger

UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany (stephan.thober@ufz.de)

1. Abstract

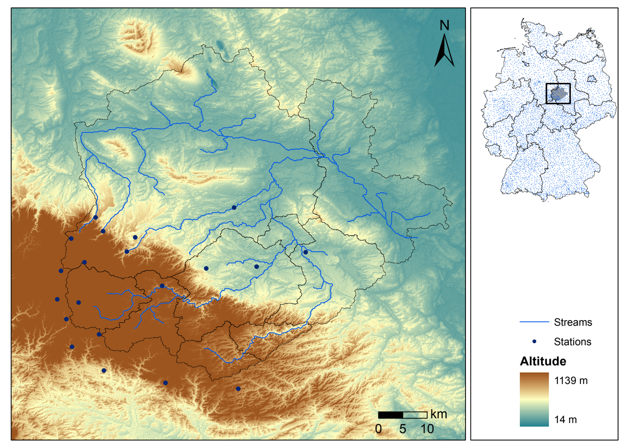
Algorithms for generating synthetic weather time series, especially precipitation, are important tools for hydrological modelling as well as for civil and agricultural engineering. They provide time series of a variable of interest of needed length, which preserve the statistical properties of observations, most importantly to note the spatial and temporal structure. There are numerous techniques and methodologies for Weather-Generators (WG) involving times series models (ARMA), Poisson processes, fuzzy rules, copulas and Markov chains among others.

2. Research Questions

1. How well can a WG reproduce site properties like monthly and annual totals, length of wet and dry spells, autocorrelation functions, etc.?
2. How well can a WG capture the spatial structure of precipitation (e.g. the variability of the first principal components or correlation coefficients of site properties)?
3. How well can a WG capture the extremes (e.g. 95 percentile of precipitation intensity and dry spell length)?

3. Study Area and Data

- **Domain**: Harz Mountain Region, Germany, approx. 10.000 km^2
- **Period**: 1961-1990
- **Observations**:
  - 20 stations provided by German Meteorological Service [2, DWD]
  - Daily precipitation



Location of precipitation stations (blue) operated by DWD 1961-2010 in Germany (right) and used for this study (left)

4.1. Weather Generator A

WG A generates precipitation occurence and intensity seperately. It is based on the method of [1, Brisette].

The occurence process is modelled via a Markov chain $X_s(t)$ of order one

$$X_s(t) = \begin{cases} 1, & \text{if day t at station s is wet,} \\ 0, & \text{if day t at station s is dry.} \end{cases} \tag{1}$$

The Markov chains $X_s(t)$ are drawn via a serially independent, but spatially correlated standard normal multivariate. If $X_s(t)$ is a rainy day, then the intensity is drawn analogously with a multivariate $Y(t)$. $Y_s(t)$ denotes the standard normal variate for station s, which is transformed into the precipitation intensity $pr_s(t)$ via

$$N_{(0,1)}(Y_s(t)) = F(pr_s(t)), \tag{2}$$

where $N_{(0,1)}$ denotes the standard normal cumulative distribution function (cdf) and F the fitted mixed exponential distribution (eq. 3).

$$F(x) = \sum_i \alpha_i (1 - e^{-\lambda_i x}), \tag{3}$$

where i is the occurence class index, which is the ratio of wet neighbouring stations to dry neighbouring stations weighted by their correlation.

4.2. Weather Generator B

WG B generates precipitation occurence and intensity simultaneously and is based on a methodology of [3, Hundecha].

Therefore an Auto-Regressive Model of order one is defined (eq. 4).

$$W(t) = R \cdot W(t-1) + C \cdot \Phi(t), \tag{4}$$

where $W(t)$ is the Auto-Regressive vector, R is a lag 1 Auto-Covariance matrix, C spatial covariance matrix and Φ is a vector of standard normal variates. The s-th entry of $W(t)$, denoted by $W_s(t)$, is indicating the occurence as well as the intensity of precipitation. If $W_s(t)$ is negative, then day t at station s is dry. Otherwise, it rains with intensity $pr_s(t)$, which is derived by

$$N_{(0,1)}(W_s(t)) = \frac{F(pr_s(t)) - (1 - p_w(t))}{p_w(t)}, \tag{5}$$

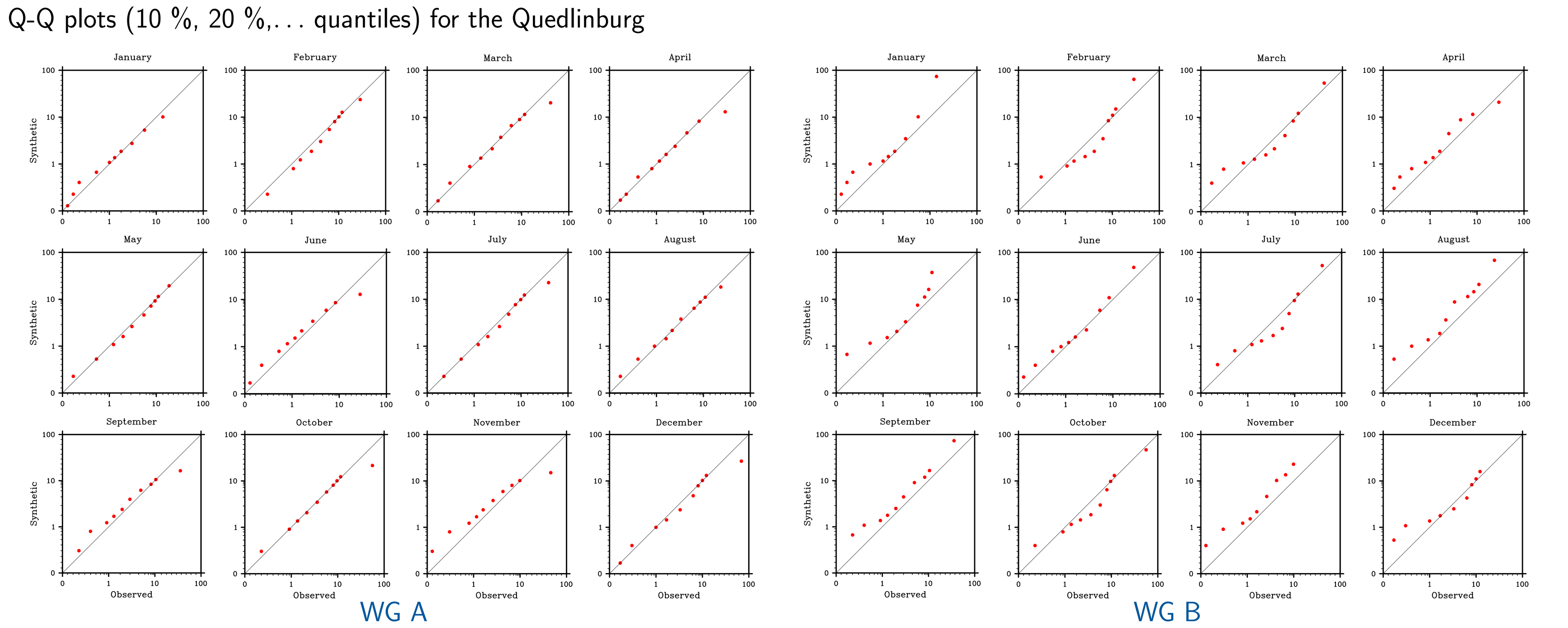
where $N_{(0,1)}$ denotes the standard normal cdf and $p_w(t)$ the probability that this day is wet. F denotes a mixed distribution, which density can be written as

$$f(x) = \frac{g(x) \cdot m(x) + h(x) \cdot (1 - m(x))}{K}, \tag{6}$$

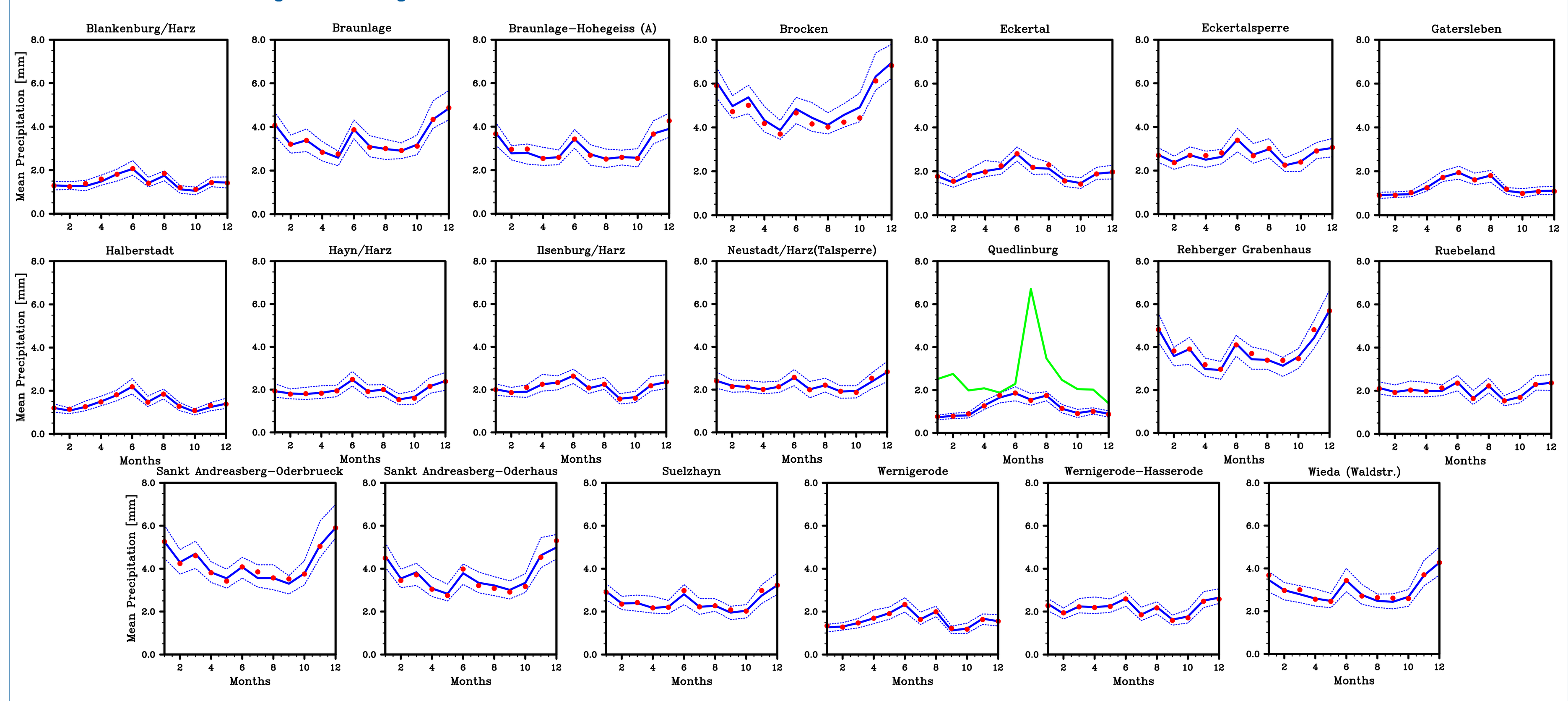
where g denotes the gamma density function, h denotes the generell pareto density function and m denotes a transformed $arctan$ here. K is a normalizing constant.

5. Fitted Intensity distribution functions

Q-Q plots (10 %, 20 %,... quantiles) for the Quedlinburg



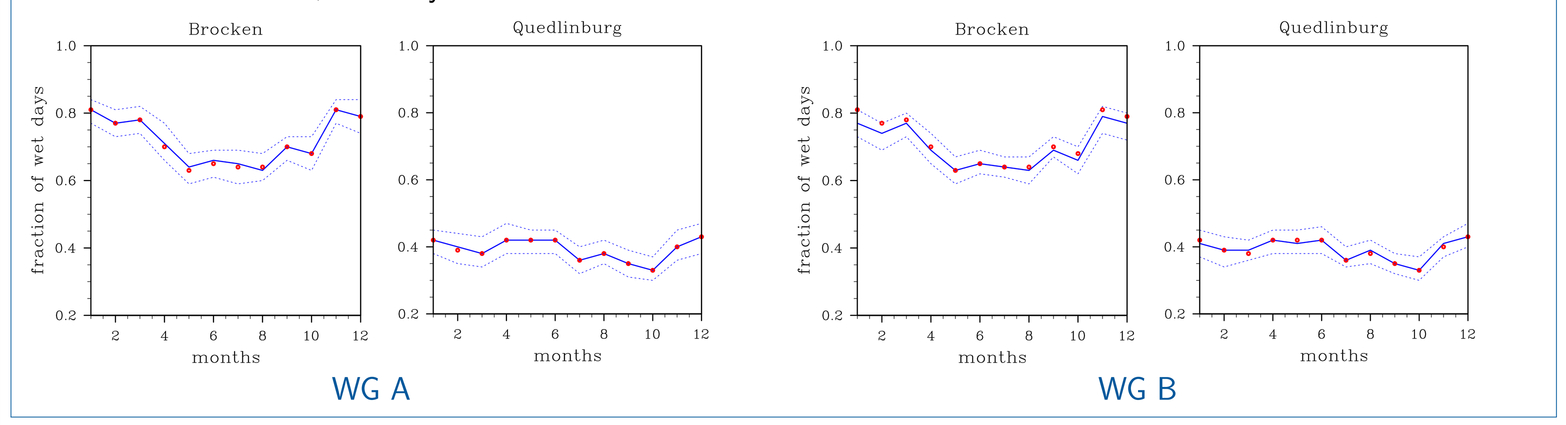
6. Mean precipitation



red - Observed, blue - WG A median with 95 % confidence interval, green - one WG B realisation

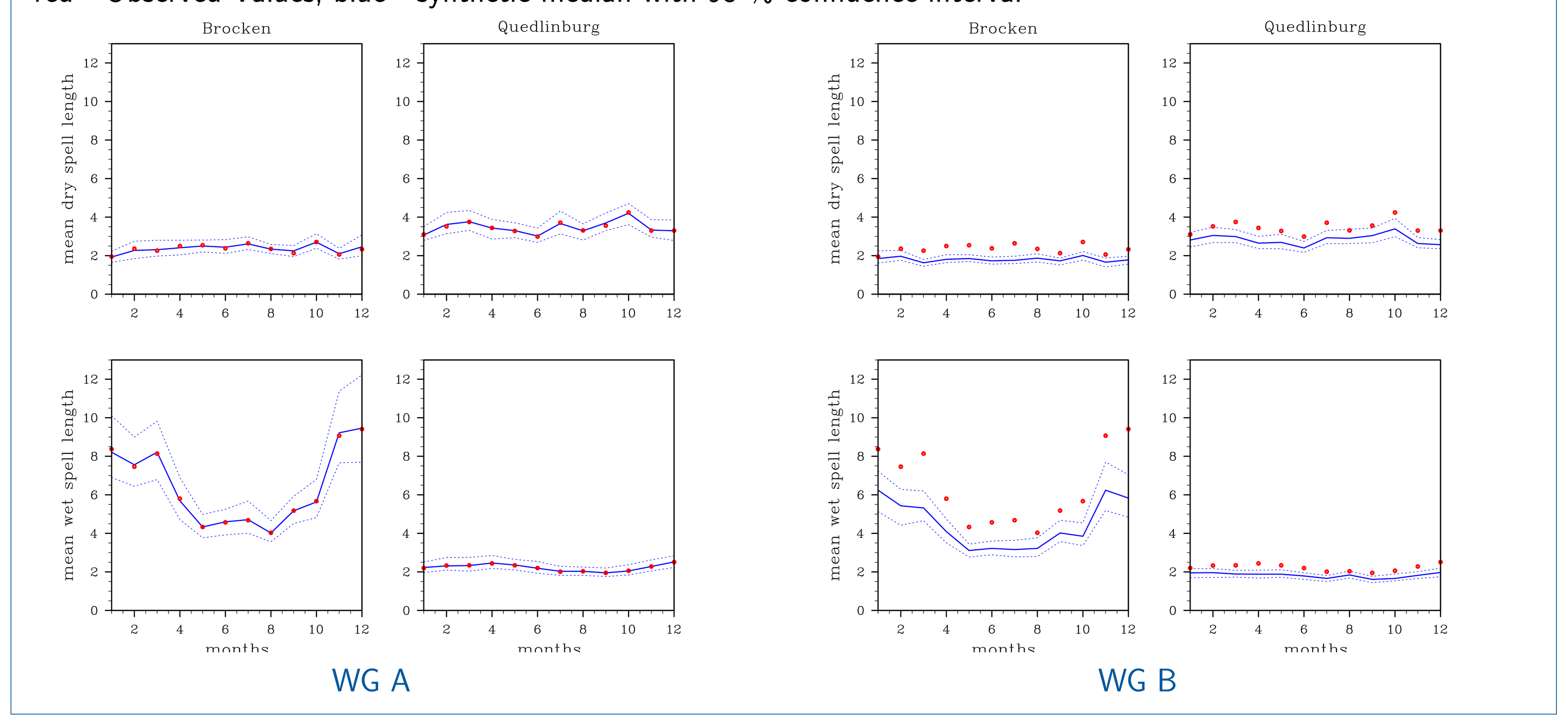
7. Proportion of wet days

red - Observed Values, blue - synthetic median with 95 % confidence interval

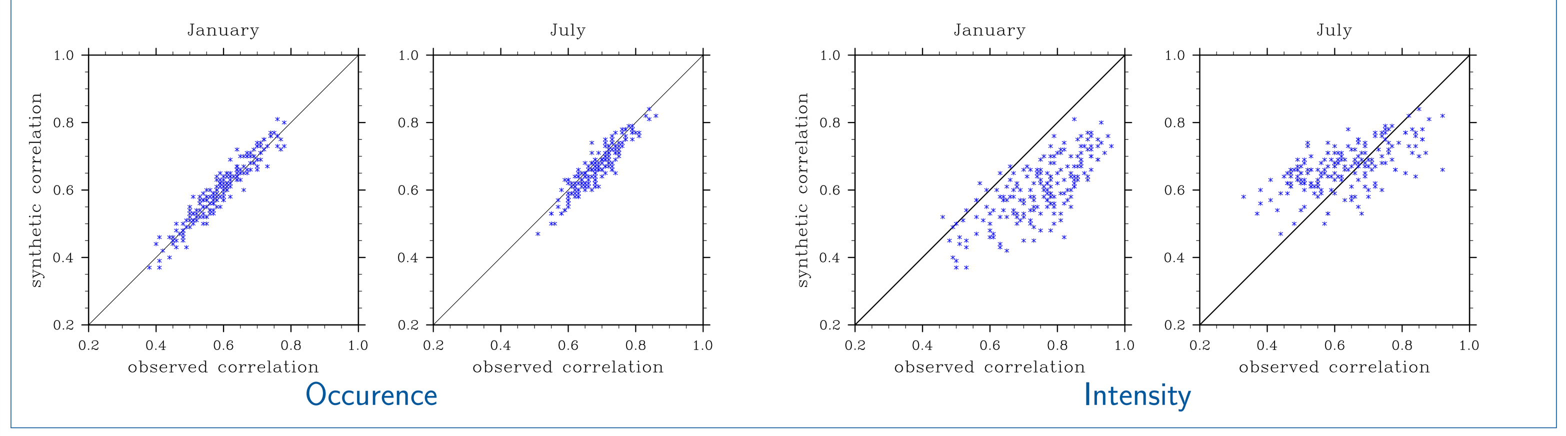


8. Spell Length

red - Observed Values, blue - synthetic median with 95 % confidence interval

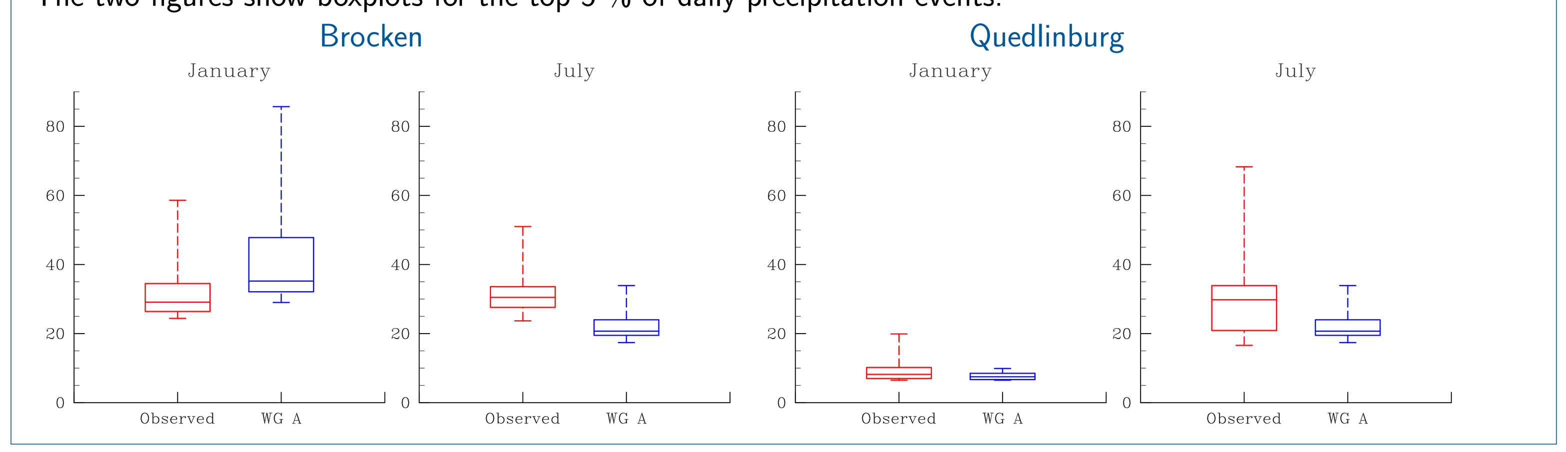


9. Correlation Coefficients of WG A



10. Extremes

The two figures show boxplots for the top 5 % of daily precipitation events.



11. Conclusions

1. We were not able to calibrate WG B in a way that it reproduces the occurence and intensity process. Further investigations are required.
2. WG A is able to reproduce the occurence process characteristics well, but exhibits deficiencies in the extremes and the spatial structure of the amount process.

References

[1] F. P. Brisette M. Khalili R. Leconte, "Efficient stochastic generation of multi-site synthetic precipitation data," *Journal of Hydrology*, vol. 345, pp. 121–133, 2007.

[2] DWD, http://www.dwd.de/.

[3] Y. Hundecha M. Pahlow A. Schumann, "Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize extremes," *Water Resour. Res.*, vol. 45, 2009, w12412.

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ