Geophysical Research Abstracts,
Vol. 11, EGU2009-9160, 2009
EGU General Assembly 2009

Towards a new surrogate-based non-Gaussian non-Markovian multi-station weather generator using self organising maps for weather classification applied to daily rain sums

T. Sauter (1) and V. Venema (2)

(1) RWTH Aachen University, Geography, Aachen, Germany (tobias.sauter@geo.rwth-aachen.de), (2) Meteorological Institute of Bonn University, Germany (Victor.Venema@uni-bonn.de)

We are developing a new multi-station weather generator, i.e. an algorithm that generates time series for a number of climate stations in a region conditioned on the large-scale circulation. The algorithm is based on the so-called surrogate data approach. It is very similar to the Iterative Amplitude Adjusted Fourier Transform (IAAFT) algorithm used to generate surrogate data. Surrogate data is synthetic data with the distribution of values and the power spectrum of an example data set. The algorithm can thus work with any distribution and power spectrum. As the power spectrum is equivalent to the auto-correlation function (ACF) any temporal auto-correlation can be specified. If desired the ACF can also be approximated by a power law or an exponential law. We have adapted the algorithm such that the distribution can be different for different categories. In this way we can specify the distributions belonging to the large-scale circulations patterns (Großwetterlagen). Furthermore, a modification was made to improve the reproduction of the cross correlations between the stations.

The large-scale circulation patterns were classified using so-called Self Organising Maps (SOM). Since no assumptions regarding the underlying data are required, SOMs present an objective unsupervised clustering method which is able to map any arbitrary linear or non-linear data distribution. Unlike common approaches the atmospheric states are clustered to get a preferable high number of unique precipitation probability density functions at each station. The cost function was redefined so that the precipitation distributions were as different as possible. We optimized the input domain and the number of categories using a simulated annealing algorithm.

The approach was tested on 30 years of data from a network of 8 stations with daily precipitation sums in the Rhineland region. In a first test, we have generated a surrogate network of 30 years based on 30 years of data. In this way we can study the reconstruction of the network under optimal conditions. Later the algorithm will be trained, applied and validated using disjoint consecutive data. The temporal variability of the average precipitation is well described by this approach. Distributions of the daily rain sums are reconstructed exactly. Admittedly, the average length of dry spells are slightly underestimated. The cross correlation matrix was reproduced with deviations of a few percent.