Local Refinement and Bias Correction of Regional Climate Simulations using Copulas

Karlsruhe Institute of Technology

P. Laux[1], S. Vogl[2], W. Qiu[1], H.-R. Knoche[1], H. Kunstmann[1,2]
Contact: patrick.laux@kit.edu

1 Karlsruhe Research Center, Institute of Meteorology and Climate Research IMK-IFU, 82467 Garmisch-Partenkirchen, Germany
2 University of Augsburg, Institute for Geography, Regional Climate and Hydrology, 86135 Augsburg, Germany

1. Introduction

In the light of global warming there is a growing concern on regional and local scale **climate change impacts**. Even high resolution regional climate models are usually not capable to reproduce observed small scale climate characteristics, particularly in complex terrains. Significant biases (Fig. 1) in simulated meteorological fields, such as e.g. precipitation and temperature, omit the direct use of model output for subsequent impact models. We investigate the possibility to use a **copula-based technique** to further **refine and bias-correct** regional climate model output on a daily time scale. Copula-based methods allow for a highly flexible consideration of the **dependence between local, small scale climate characteristics and regional or global formation**. The approach allows to model the dependence of variables independently from the choice of the marginal distributions.

2. Research Area & Data

Regional climate simulations for Germany (Fig.1, left) are obtained within the DEKLIM project using the Penn State/NCAR Mesoscale Model MM5 and ECMWF/ERA15 reanalysis data for 1979-1993 at 19.2 km spatial resolution.

Local rainfall information is obtained from the German weather services (DWD). For the Alpine region of Germany, rainfall data of 132 observation station are retrieved. It can be seen from Figure 1 (right) that rainfall is overestimated by MM5 for the whole eastern part of Germany, and strongly underestimated for the Rhine valley and the Alpine region of Germany. The underestimation in the Alpine region is possibly due to the complex terrain with very steep gradients of altitude.

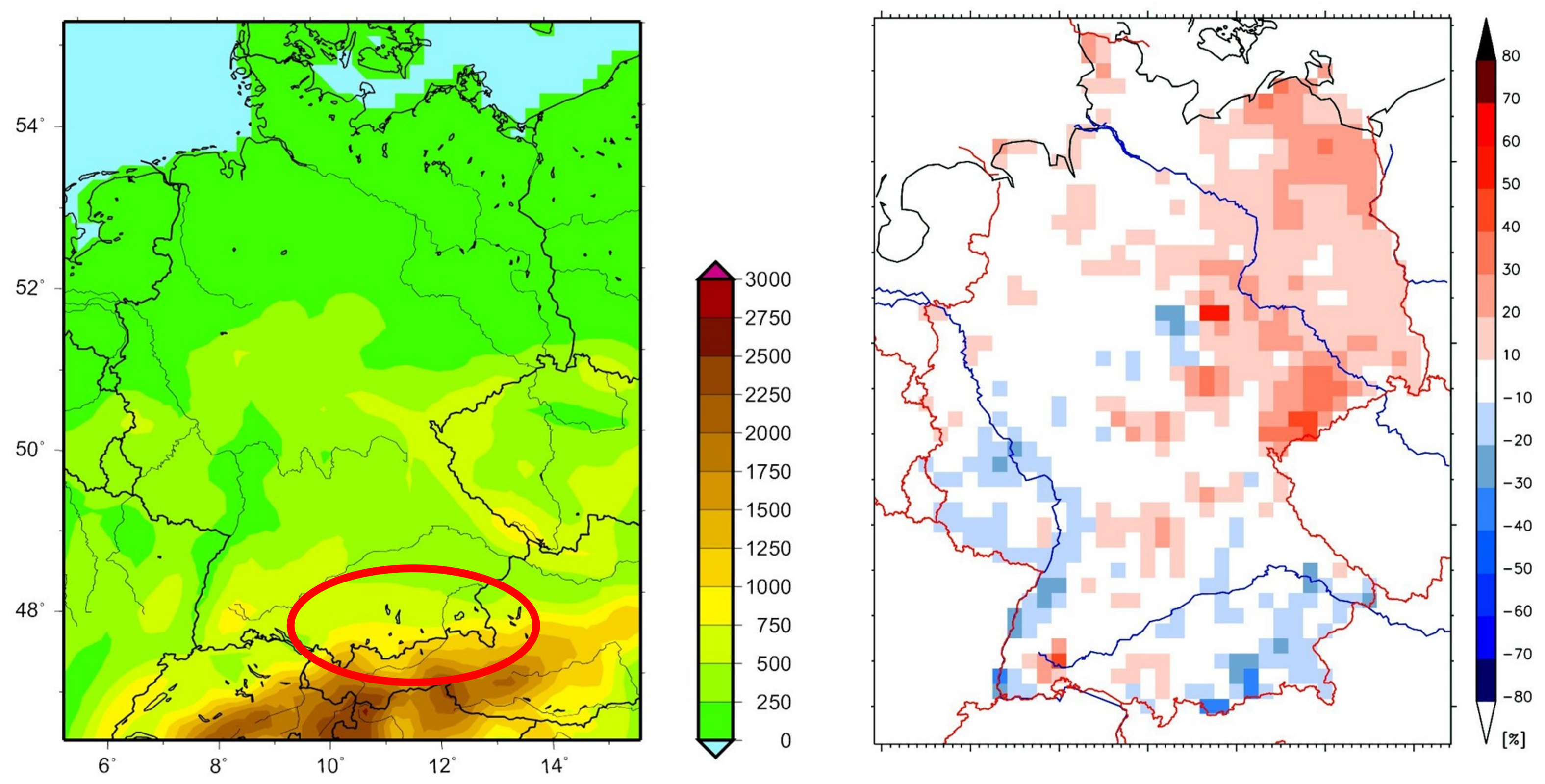

Fig. 1: Domain and topography of regional climate simulations (left). Bias of mean annual total precipitation for the MM5 with respect to the DWD reference data set [%]

3. Modelling the Dependence Structure between Modelled and Observed Rainfall

3.1 Modelling the Marginals

Modelling the single marginal distributions requires the observations to be **independent and identically distributed (iid)**. However, most climatological time series exhibit some degree of **autocorrelation and heteroskedasticity**. The ARMA-GARCH composite model is used to generate iid variables, followed by fitting the marginals, and a joint distribution function (Copula) to model the dependence between modelled and observed rainfall time series.

3.1.1 ARMA-GARCH-Filter

An ARMA model is used to compensate for autocorrelation, and a GARCH model to compensate for heteroskedasticity (time-varying variance). GARCH stands for Generalized Autoregressive Conditional Heteroskedasticity. The term *Conditional* implies explicitely the dependence on a past sequence of observations, and *Autoregressive* describes a feedback mechanism that incorporates past observations into the present

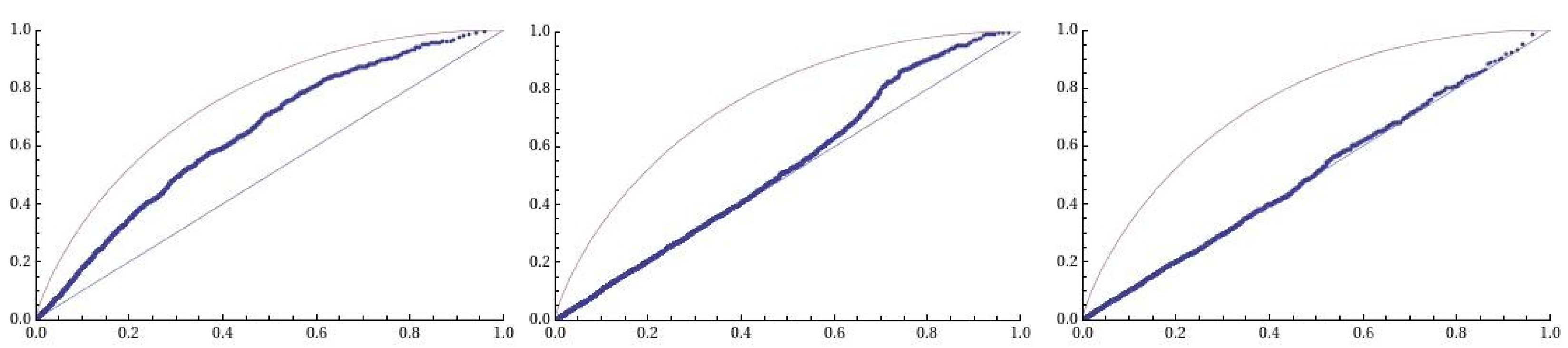

Fig. 2: K-plot of the observed rainfall time series at Garmisch-Partenkirchen (Germany) before ARMA-GARCH transformation (left), after ARMA-GARCH transformation (middle), and a random variable (right). Superimposed on the graphs are a straight line corresponding to the case of perfect independence and a curve corresponding to perfect positive dependence.

3.1.2 Generalized Pareto Distribution

The residuals of the ARMA-GARCH model are fitted to a semi-parametric CDF. First, the empirical CDF of each variable is estimated using Gaussian kernel function. This provides a relatively good fit to the lower tails and the interior of the distribution of the residuals, but this procedure tends to perform poorly when applied to upper tails. Thus, the upper tails are fitted separately from the interior of the distribution. For this reason, the *peaks over threshold method* is applied: A POT value of 0.1 is chosen, i.e. the upper 10% of the residuals are reserved for the upper tail. The extreme residuals are fitted to a parametric GPD using maximum likelihood. Given the exceedance in the upper tail, the negative log-likelihood function is optimized to estimate the tail index and the scale parameter of the GPD. The piecewise distribution object allows for interpolation in the interior of the CDF but also extrapolation in the upper tail.

3.2 Copula Theory: Joint Dependence Structure

Copulas are functions that link univariate distribution functions to form a multivariate distribution function (Sklar,1959). For any bivariate distribution function $F_{XY}(x,y)$ with univariate marginal distribution functions $F_X(x)$ and $F_Y(y)$ there exists a copula C such that:

$$\begin{aligned} F_{XY}(x,y) &= C(F_X(x),F_Y(y)) \qquad x,y \in \mathbb{R} \\ &= C(u,v) \qquad u,v \in [0,1] \end{aligned}$$

In turn, they allow for separating the dependence structure from the marginal distributions in a multivariate distribution if the pdf c is known. The construction of Copulas is therefore reduced to the study of the relationship between the correlated iid variables, giving freedom for the choice of the univariate marginal distributions:

$$f(x,y) = \underbrace{c(F_X(x),F_Y(y))}_{\text{pdf of Copula (copula density)}} \cdot \underbrace{f_X(x) \cdot f_Y(y)}_{\text{pdfs of marginal distributions}} \qquad c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \, \partial v}$$

The dependence structure of daily measured precipitation on simulated precipitation is studied. Since the underlying (theoretical) copula is not known in advance, it is necessary to analyze the empirical density copula, which is solely based on the data.

4. Modelling Approach

Fig. 3 (left) shows the empirical copula density between modelled and measured rainfall residuals for station Garmisch-Partenkirchen. It can be seen that the distribution is strongly asymetrical for the minor diagonal, and that the density in the upper corner is highest. This implies that modelled and observed rainfall are strongly dependent in the higher ranks of the distribution, and the dependence is weaker in the lower ranks.

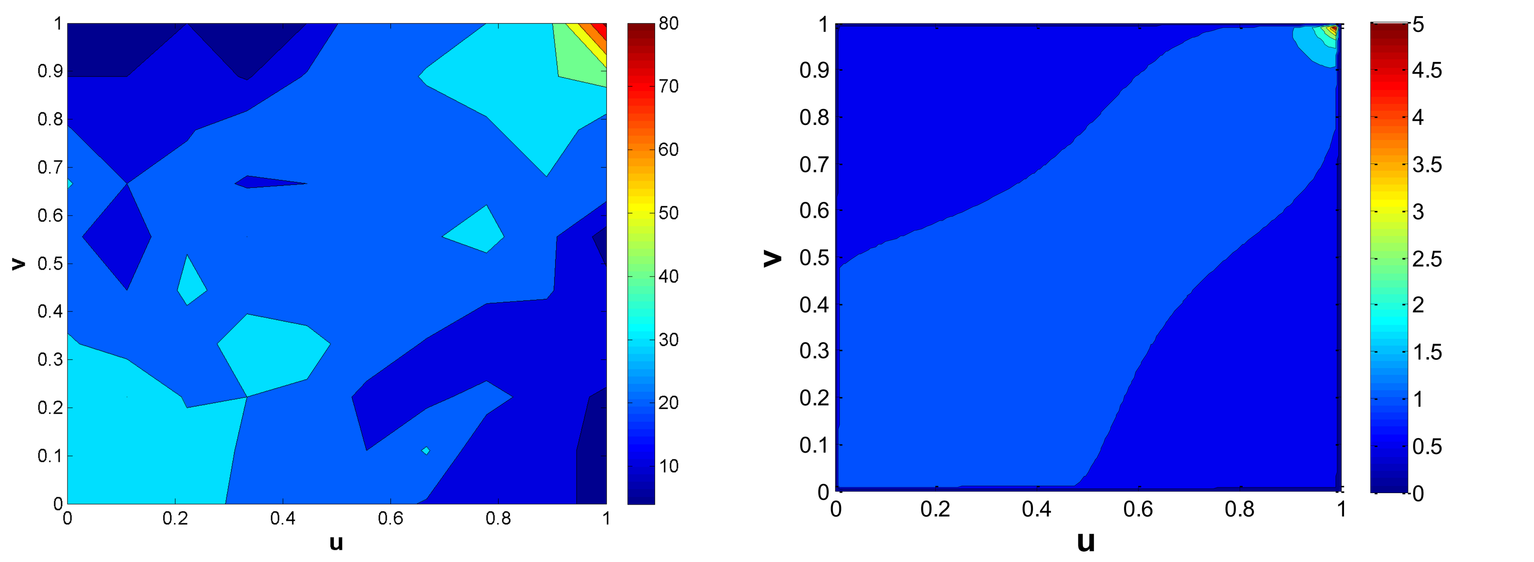

Fig. 3: Empirical copula density between modelled and observed rainfall (left). **Gumbel** copula probability density function (Θ=1.097) (right).

Based on the empirical copula density a theoretical copula model can be estimated using goodness-of-fit tests. Here, the Gumbel copula (Θ=1.097) is chosen (**Fig 3, right**). Once the copula-based joint distribution is estimated conditional random samples from this distribution can be generated through Monte Carlo simulations . The conditional simulation of (x,y) is divided into four steps:

1. Compute $u=F_X(x)$
2. Draw random samples of $v|u$ using the inverse conditional CDF $C_{V|U}^{-1}(v,u)$
3. Invert from v to obtain y
4. Reintroduce the autocorrelation and heteroskedasticity observed in the original time series

5. Outlook

- Analysis of copula density depending on the altitude and flow direction of the observation station
- Comparison of copula-based method with “traditional” methods

Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft

HELMHOLTZ GEMEINSCHAFT

Karlsruhe Institute of Technology (KIT)
The Cooperation between the Forschungszentrum Karlsruhe GmbH and the Universität Karlsruhe (TH)

Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825