# Investigation of hydrological time series using copulas for detecting catchment characteristics and anthropogenic impacts

Takayuki Sugimoto[1], András Bárdossy[1,2], Geoffrey G. S. Pegram[2] and Johannes Cullmann[3]

1 Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany

2 Civil Engineering Program, University of KwaZulu-Natal, Durban, South Africa

3 Water & Climate Department, World Meteorological Organization, Geneva, Switzerland

**Abstract.** Global climate change can have impacts on characteristics of rainfall-runoff events and subsequently on the hydrological regime. Meanwhile, the catchment itself changes due to anthropogenic influences. However, it is not easy to prove the link between the hydrology and the forcings. In this context, it might be meaningful to detect the temporal changes of catchments independent from climate change by investigating existing long term discharge records. For this purpose, a new stochastic system based on copulas for time series analysis is introduced in this study.

A statistical tool like copula has the advantage to scrutinize the dependence structure of the data and, thus, can be used to attribute the catchment behavior by focusing on the following aspects of the statistics defined in the copula domain: (1) Copula asymmetry, which can capture the non symmetric property of discharge data, differs from one catchment to another due to the intrinsic nature of both runoff and catchment. (2) Copula distances can assist in identifying catchment change by revealing the variability and interdependency of dependence structures.

These measures were calculated for 100 years of daily discharges for the Rhine rivers and these analyses detected epochs of change in the flow sequences. In a follow-up study, we compared the results of copula asymmetry and copula distance applied to two flow models: (i) Antecedent Precipitation Index (API) and (ii) simulated discharge time series

generated by a hydrological model. The results of copula based analysis of hydrological time series seem to support the assumption that the Neckar catchment had started to change around 1976 and stayed unusual until 1990.

**Keywords :** Catchment discharge characteristics, Copula stochastic analysis, API, Model uncertainty

## 1. Introduction

In order to understand the water cycle behavior of a region, it is important to determine its characteristics, but this is difficult to achieve due to the diversity of the system response at different time and space scales. In particular, temporal variability makes parameter estimation difficult and the assessment of model uncertainty essential. As a part of the endeavor to understand the hydrological system, the objective of this research, assessing the anthropogenic impacts on the catchment characteristic independent of the climate change, is therefore important, yet hard to accomplish.

The first possible approach is to statistically test the existence or change of trend in hydrological time series which can be related to climate changes or anthropogenic impacts. Mann-Kendall's Test was performed to confirm the existence of a trend in the annual discharge, precipitation and sediment loads, then the human intervention and climate impacts based on the available information of the catchments were discussed (Wu et al., 2012). Pettitt's Method (Pettitt, 1979) can be used to detect the time point of trend alternation and analyze the impacts based on a double mass curve (Gao et al., 2012) or a hydrological model (Karlsson et al., 2014). These non-parametric methods for detecting the signal seem, however, not capable enough of explaining when and how much the system had changed, thus making it still difficult to relate the change due to human activities.

On the other hand, runoff events are initiated by precipitation then modified by the state and physical features of the catchment. This implies that the integrated information of catchment status might be retrieved by analyzing the discharge time series itself. Focusing on this property, the attempts can be made for capturing the temporal dependence structure of runoff by time series models. The classical time series model, autoregressive integrated moving average (ARIMA), is designed to describe a stationary stochastic process based on the temporal correlation structure of Gaussian random variables (Box and Jenkins, 1976). However, the stationarity of the data is not guaranteed in reality, thus a number of alternative approaches have been suggested. While the application of Fourier analysis is basically for stationary process, the analysis using empirical mode decomposition (Huang et al., 1998) overcomes the restriction of stationarity by allowing the frequency and local variance of a time series to vary within a component and to separate the signals adaptively by scale. Autoregressive Conditional Heteroskedasticity (ARCH) models lose the assumption of stationarity to a certain extent so that variance is not constant, however models the variance in a similar way to ARIMA. Although inventions and efforts to overcome the limitation of stationarity have been made, it seems still inadequate to model dynamic changes of hydrological processes with these time series models.

Alternatively there is a statistical concept, the copula, which has advantages to model the multivariate dependence independently from marginals and recently adopted in the field of hydrology. A Copula (Sklar, 1959) is a multivariate probability distribution designed to flexibly model dependence structure in the uniform (quantile) domain. The use of copulas in hydrology can be found for the assessment of extreme events by considering flooding as a joint behavior of peak and volume (De Michele and Salvadori, 2003). Copulas have been applied to describe the spatio-temporal uncertainty of precipitation (Bárdossy and Pegram, 2009) or the inhomogeneity of groundwater parameters (Bárdossy and Li, 2008). Asymmetry of dependence in a time series can be tested in the framework of a finite state Markov chain's transition probability matrix (Sharifdoost et al., 2009). Dissimilarity measures can be defined by means of a copula modelling the correlation structure of pairs of discharge time series in order to identify the similarity of

73 catchments with the purpose of transferring catchment properties from one to the other (Samaniego et al.,
74 2010). We aim at utilizing copulas as an alternative to classical time series models and an efficient tool for
75 time series analysis to overcome these hydrological challenges.

76 The main interest of this study is to assess the human intervention and climate change impacts on
77 hydrological regime for the strategy of future development in the region. For achieving this goal, 7 daily
78 discharge gauging stations in South-West Germany (Fig. 1), which have 100 years daily discharge records,
79 were chosen and extensively analyzed. The gauging stations Andernach, Kaub, Worms and Maxau are
80 located in the main stream of the Rhine, while Kalkofen, Cochem and Plochingen are located on tributaries.
81 For further analysis, daily precipitation and temperature records in the Baden-Württemberg state of
82 Germany for the last 50 years were obtained from the German Weather Service. Also, 77 discharge records
83 obtained from the Global Runoff Date Centre in Germany were utilized.

84 The following are the novel aspects introduced in this study: (1) The catchment characteristics are
85 defined based on copulas and estimated from discharge data. Also the changes of catchment characteristics
86 are investigated by tracing the temporal change of these statistics. (2) A method to model systematic
87 changes of dependence structure with the help of copulas is suggested, then its variability and
88 interrelationship with the time series are examined. (3) Anthropogenic impacts are assessed by the
89 discharge - precipitation relation using API and a hydrological model with copula based measures.

90 This article is divided into five sections. After the introduction, the basic methodology for applying
91 copulas to discharge time series is introduced in the second section. Thirdly, the measures of asymmetry in
92 copulas are defined and estimated for the discharges of the River Rhine and other catchments. The
93 determination of the temporal change of the asymmetry of the copulas is treated in the third section as well.
94 In the fourth section two topics are treated: (i) the analysis based on copula distances for the observed
95 discharges and (ii) the comparison of observed discharge with API (Antecedent Precipitation Index) time
96 series and simulated discharge time series with a hydrological model. The conclusion is given in the fifth
97 section.

## 2. Methodology

In this section, the application of copulas to time series is articulated after a brief introduction. The very basics about copulas are presented here ; further information can be obtained from (Joe, 1997) or (Nelsen, 2006).

### 2.1 Basic Methodology

In probability theory and statistics, a copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform.

$$C:[0,1]^n \to [0,1] \tag{1}$$

$$C\left(\mathbf{u}^{(i)}\right) = u_i \qquad if\ u^{(i)} = (1,\ldots,1,u_i,1,\ldots 1) \tag{2}$$

Any multivariate distribution can be described by a copula and its marginal distributions as was proven by Sklar's theorem (Sklar, 1959):

$$F(\mathbf{x}) = C\left(F_{X_1}(x_1),\ldots,F_{X_n}(x_n)\right) \tag{3}$$

where $F_{\mathbf{X}_i}(x_i)$ represents the *i*-th marginal distribution of a multivariate random variable $\mathbf{X}$. The copula density can be derived by taking partial derivatives of the copula:

$$c(u_1,...,u_n) = \frac{\partial^n C(u_1,...,u_n)}{\partial u_1 ... \partial u_n} \tag{4}$$

The advantage of using copulas is that the marginal is detached from the multivariate distribution and the dependence structure can be examined in the uniform compact domain for different types of data.

### 2.2 Basic Hypothesis of Temporal Copulas

For the application of copulas to time series analysis, a stochastic system should be presumed to be similar to the case of spatial copulas (Bárdossy and Li, 2008): the random variable at time $t$ is described as $Z(t)$ and in general there may exist non-Gaussian dependency among the elements of $Z(t)$. Then

stationarity is defined for each subset of times $t_1,\ldots,t_n \subset N$ and time lag $k$ such that $\{t_1+k\ldots,t_n+k\}\subset N$ and for each set of possible values $z_1,\ldots,z_n$:

$$P\left(Z\left(t_1\right)<z_1,\ldots,Z\left(t_n\right)<z_n\right)= P\left(Z\left(t_1+k\right)<z_1,\ldots,Z\left(t_n+k\right)<z_n\right) \quad (5)$$

For the given random function $Z(t)$, a set $S(k)$ containing pairs of ranked values is defined as a function of time lag $k$ as follows:

$$S(k)=\left\{\left(F_Z\left(z(t)\right)\right),\left(F_Z\left(z(t+k)\right)\right)\right\} \quad (6)$$

Thus, a 2-dimentional autocopula for stochastic time series is a function of time lag $k$ for the set $S(k)$ similar to the case of spatial copula (Bárdossy and Li, 2008):

$$\mathbf{C}_t\left(k,u_1,u_2\right)=P\left[F_z\left(Z\left(t\right)\right)<u_1,F_z\left(Z\left(t+k\right)\right)<u_2\right] \quad (7)$$

where $(u_1,u_2)\in S(k)$. Thus, a 2-dimentional empirical copula density can be constructed based on conditional empirical frequencies on a regular $g\times g$ grid and kernel density smoothing (Bárdossy, 2006):

$$c^*\left(\frac{2i-1}{2g},\frac{2j-1}{2g}\right)=\frac{g^2}{\left|S(k)\right|}\cdot\left|\left\{\left(u_1,u_2\right)\in S(k);\frac{i-1}{g}<u_1<\frac{i}{g}\,and\,\frac{j-1}{g}<u_2<\frac{j}{g}\right\}\right| \quad (8)$$

where $\left|S(k)\right|$ denotes the cardinality (the number of elements in a set) of set $S(k)$.

## 3. Copula Asymmetry in Discharge Time Series

High and low values might have different dependences in general. Measuring the asymmetry of copulas could reveal substantial aspects of time series data, which are not illuminated in the Gaussian approach. Statistics defined on copula shape and calculated from observed discharge time series we believe to be a new idea. The two types of asymmetry, "asymmetry1" and "asymmetry2", are considered for two diagonals on 2-dimensional copulas, which can be described as a function of time lag *k* (Li, 2010):

138
$$A_1(k) = E[(U_t - 0.5)(U_{t+k} - 0.5)((U_t - 0.5) + (U_{t+k} - 0.5))] = \int_0^1 \int_0^1 (u_t - 0.5)(u_{t+k} - 0.5)(u_t + u_{t+k} - 1)c(u_t, u_{t+k})du_t du_{t+k} \quad (9)$$

139
$$A_2(k) = E[-(U_t - 0.5)(U_{t+k} - 0.5)((U_t - 0.5) - (U_{t+k} - 0.5))] = \int_0^1 \int_0^1 -(u_t - 0.5)(u_{t+k} - 0.5)(u_t - u_{t+k})c(u_t, u_{t+k})du_t du_{t+k} \quad (10)$$

140 where $u_t = F_Z(z(t))$, $u_{t+k} = F_Z(z(t+k))$. $A_1(k)$ and $A_2(k)$ are asymmetry functions corresponding to
141 asymmetry1 and asymmetry2 respectively. Figure 2 shows an idealization of the asymmetries between a
142 pair of variables $U(t)$ and $U(t+k)$, showing that the tails of the distributions have a large impact on
143 each type of asymmetry. The measure of asymmetry compares the dependency between low and high
144 values and quantifies how much it is not symmetric. For example, in a 2-dimensional copula, $A_1(k)$ is
145 positive if the probability density is higher in the upper right corner than in the lower left corner. On the
146 contrary, $A_1(k)$ is negative if the probability density is higher in the lower left corner. $A_2(k)$ is the
147 asymmetry for the other diagonal of a 2-dimensional copula.

148 Figure 3 shows the scatterplot of ranked values of a discharge time series with time lag $k$ = 1 as a sample
149 of an empirical autocopula and its relation with storm hydrographs. This figure displays (i) where each pair
150 of values on a hydrograph can be plotted on an empirical copula, demonstrating that (ii) the dependence
151 structure is not symmetric especially for $A_2(k)$. This illustration provides the insight that asymmetry can
152 be related to the shape of a unit hydrograph as well as the notion that asymmetry might be used for
153 advanced modeling of hydrological time series.

154

155 **3.1 Asymmetry and catchment characteristics**

156 Asymmetry functions can be considered as statistics calculated from the observed discharge time series
157 and an important assumption can be made: “assymetry2 is related to catchment characteristics”. This idea

will be discussed and demonstrated in this section. Figure 5 (upper left) shows parts of the hydrographs of 7 gauging stations in southwest Germany.

First, an important natural property of discharge seen in this figure is that the durations of high flow and low flow periods are not symmetric: Flood events, which are initiated by rainfall or snowmelt, do not continue for a long time because the duration of runoff to rivers is comparatively short. On the other hand, discharge keeps decreasing and stays low for no rain periods. This means that, if two consecutive values in a time series are chosen for small time lag *k* (day), these two values are likely to be less correlated for high values but more correlated for low values, which leads to negative value of $A_1(k)$. This implies that the intrinsic temporal distribution of precipitation can be investigated based on this asymmetry, possibly with advanced asymmetry functions such as bivariate moments based on L-moments (Brahimi et al., 2015).

Second, the rates of increase and decrease of discharge are not symmetrical in the upper limb compared to the lower limb of the hydrograph (Fig. 3): Soon after the rainfall, the river flow rises sharply, but once the rain stops and peak discharge is observed, then the water level starts to decrease, typically more slowly on the recession than the rising limb of the hydrograph. This leads to the negative values of $A_2(k)$ for small time lags *k* (day) and the notion that asymmetry2 can be related to the shape of the hydrograph, and therefore the characteristics of the runoff and catchment. In addition, it can be said the annual cycle in Fig. 4 is not symmetric in the same sense a hydrograph is not symmetric.

The change of $A_2(k)$ with time lag *k* is now discussed. The point is that these statistics for small time lags *k* can be more related to the catchment and rainfall characteristics of the region, while asymmetry for larger time lags *k* can capture the inter-seasonal characteristic of the climate in the region.

In order to reduce such seasonal impacts on the analysis of hydrological time series, deseasonalization measures can be applied, for example, for daily stream flow (Grimaldi, 2004). Adopting Grimaldi's method, all the time series are normalized in this study. First, the mean $\mu_i$ on the *i*-th calendar day is calculated as the expectation of the random variable $X_i$. Then, the annual cycle of the mean $\mu_i^*$ (Fig. 4

left) is calculated as a smoothed version of $\mu_i$ by linearly weighting the neighboring values along *i* and summing them up. The smoothed annual cycle of standard deviations $\sigma_i^*$ (Fig. 4 right) can be obtained in the same way. Then the normalized time series is defined by dividing the original time series $Z(t)$ by $\sigma_i^*$ after subtracting $\mu_i^*$ as follows

$$Z_{\mathrm{norm}}(t) = \frac{Z(t) - \mu^*_{t|365}}{\sigma^*_{t|365}} \tag{11}$$

where $t \,|\, 365$ is $t \,(\mathrm{mod}\, 365)$ and represents calendar day at time *t* (day). Figure 5 (upper right) shows parts of normalized discharge time series from the 7 gauging stations. It should be noted that the process still appears to be non-Gaussian after this transformation and the seasonality for small time lags *k* might not have been fully eliminated. Figure 5 (bottom left and bottom right) shows the variation of asymmetry functions for 7 discharge time series corresponding to time lag *k*, similar to correlograms, in addition to the confidence interval of Gaussian process.

The confidence intervals in the figures are gained by calculating $A_2(k)$ for 100 realizations of stationary Gaussian process which are fitted to the observed discharge of Andernach. The result shows that the process is clearly different from Gaussian and the influence of asymmetry is significantly large.

It can be seen that the variation of $A_2(k)$ of discharge without normalization (Fig. 5 bottom left) has a larger impact of seasonality for bigger *k* ($k > 40$), while its impacts are mitigated after the normalization (Fig. 5 bottom right). Furthermore, as a consequence of normalization, a sharp drop down of $A_2(k)$ for small time lags *k* emerged which might be regarded as a catchment indicator. Therefore, the selected/critical properties for small time lags *k* is formulated by (i) taking the minimum value of $A_2(k)$ for the time lag $k < 50$ and (ii) the lag *k* at the minimum of $A_2(k)$:

$$A_{2,\min} = \min_{k<50} A_2(k) \tag{12}$$

$$L_{2,\min} = \min_{0<k<50} \{k; A_2(k) = A_{2,\min}\} \tag{13}$$

The question is whether they are really related to catchment characteristics. Now, these statistics estimated for 77 discharge data recorded at the gauging stations in Germany are compared with the catchment area as one of the simplest possible indicators of the catchment as shown in Fig. 6. $A_{2,\min}$ - area (Fig. 6 top) and $L_{2,\min}$ - area (Fig. 6 middle) both show a linear relationship with the log-scaled x-axis of catchment area, with positive correlation. There seems also to be a linear relation between $A_{2,\min}$ and $L_{2,\min}$ (Fig. 6 bottom) as a consequence of the above relationships.

This result demonstrates that the information extracted from discharge is related to the basic information of its catchment to a certain extent. Since the principal objective is to assess anthropogenic impacts, the idea introduced now is to use this measure for evaluating the catchment change by calculating chronological changes of $A_{2,\min}$.

### 3.2 Time Series Analysis with Asymmetry

Temporal change of asymmetry2 is defined $A_2(k,t)$ on the set representing a moving time window of size *w*.

$$S^*(k,t) = \left\{ \left(F_Z(z(a))\right), \left(F_Z(z(a+k))\right); t - \frac{w}{2} < a < t + \frac{w}{2} \right\} \tag{14}$$

$$\begin{aligned} A_2(k,t) &= E\left[-(U_t - 0.5)(U_{t+k} - 0.5)((U_t - 0.5) - (U_{t+k} - 0.5))\right] \\ &= \int_0^1 \int_0^1 -(u_t - 0.5)(u_{t+k} - 0.5)(u_t - u_{t+k})c(u_t, u_{t+k})du_t du_{t+k} \end{aligned} \tag{15}$$

where $u_t \in U_t, u_{t+k} \in U_{t+k}, (u_t, u_{t+k}) \in S^*(k,t)$. Then the minimum of $A_2(k)$ and lag *k* at the at time *t* are given by

$$A_{2,\min}(t) = \min_{k<30} A_2(k,t) \tag{16}$$

$$L_{2,\min}(t) = \min_{0<k<30} \{k; A_2(k,t) = A_{2,\min}(t)\} \tag{17}$$

223 Figure 7 shows the temporal changes of $A_{2,\min}(t)$ with window size $w = 3000$ (days) for 7 gauging
224 stations in southwest Germany in addition to the confidence interval calculated for 100 times
225 independently generated Gaussian process.

226 The comparison of $A_{2,\min}(t)$ from observed discharges with $A_{2,\min}(t)$ from a Gaussian process exhibits
227 (i) the influence of asymmetry in discharge is significantly large as was seen in Fig. 5, (ii) The fluctuations
228 of $A_{2,\min}(t)$ of 7 observed discharge time series appear to be bigger than the one calculated for a
229 realization of a Gaussian process and (iii) $A_{2,\min}(t)$ of these 7 discharge records shows a similar trend:
230 there are big drop-downs around 1945 and after 1980 for all the discharges.

231 However, it cannot be ascertained whether this is caused by the simultaneous change of the catchments,
232 the long term meteorological behavior in the region or just randomness in the stationary process. To
233 overcome this, temporal behavior of discharge and temperature were first checked by calculating the mean,
234 the standard deviation and the minimum of discharge and temperature in a time window centered on time
235 $t$ defined by

236
$$\begin{aligned} \mathrm{Mean}(t) &= \frac{1}{w}\int_{t-w/2}^{t+w/2} z(a)\,da \\ \mathrm{Std}(t) &= \sqrt{\mathrm{Var}(t)} = \frac{1}{w}\left(\int_{t-w/2}^{t+w/2}\left(z(a)-E[Z(t)]\right)^2 da\right)^{\frac{1}{2}} \\ \mathrm{Min}(t) &= \min\left\{z(a); t-\frac{w}{2} < a < t+\frac{w}{2}\right\} \end{aligned} \tag{18}$$

237 where $w$ is the size of time window. Figure 8 shows the moving average and moving standard deviation of
238 discharge records with windows size $w = 3000$ (days), but it is hard to say whether the behavior around
239 1945 and after 1980 is unusual. Figure 9 shows mean and minimum of temperature in the time window of
240 size 365 days which correspond to annual mean and minimum. Roughly speaking, there are certain cold
241 periods around 1940, 1955 and 1985, which might influence the snow accumulation and melting in the
242 region, but the relation with $A_2(k)$ is rather obscure.

What seems to be a useful outcome from the above exploratory analysis is that (i) the behavior of $A_2(k)$ is different from catchment to catchment showing a statistical relation with the catchment area and (ii) temporal behaviors of $A_2(k)$ of 7 discharges time series are dependent on each other, which implies the existence of a background mechanism common to the region.

## 4. Analysis of hydrological time series with Copula Distance

As an alternative to copula asymmetry, which emphasizes the behavior in the corners of copulas, copula distance is here suggested so that the characteristic behavior can be captured in the entire domain of the copula. Calculating this for each time step for different time series and comparing them hopefully exhibits the changes of dependence structure and therefore the catchment change.

### 4.1 Introduction of Copula Distance

The basic idea behind the copula distance is to apply the Cramér-von Mises type distance

$$D = \int_0^1 \int_0^1 \left( C^*(u_1, u_2) - C(u_1, u_2) \right)^2 du_1 du_2 \tag{19}$$

which by design measures the goodness of fit between two distribution functions to two copulas. This type of distance was tested to measure the difference between empirical and theoretical copulas in the bootstrap framework for the evaluation of spatial dependence of ground water quality (Bárdossy, 2006). For the analysis of time series data, it still needs to be carefully thought out how (and which) copulas should be chosen.

#### 4.1.1 Introduction of Copula Distance to single time series

In order to apply the concept of copula distance to time series, the adoption of two copulas in different time scales is considered. An empirical copula can be obtained from an entire time series which contains the averaged information of all the time points (*global* copula). Another empirical copula can be obtained

for a certain time window of width *w* centered at time step $t$ (*local* copula). In order to make the concept clear, two sets containing pairs of ranked values with different time scales are specified.

$$S_{\text{global}}(k) = \{(F_Z(z(t))), (F_Z(z(t+k))); t_1 < t < t_n\} \tag{20}$$

$$S_{\text{local}}(k,t) = \left\{(F_Z(z(a))), (F_Z(z(a+k))); t - \frac{w}{2} < a < t + \frac{w}{2}\right\} \tag{21}$$

$S_{\text{local}}(k,t)$ can be interpreted as a moving time window where the reference time *t* is set to the middle of the window of size *w*, while $S_{\text{global}}(k)$ represents a set of the entire time series. Global copula and local copula are the empirical autocopula densities defined on these sets based on Eq. (8), there denoted by $c^*_{\text{global}}(\mathbf{u})$ and $c^*_{\text{local}}(\mathbf{u},t,w)$ respectively for the n-dimensional case. In this analysis, 3000 days for the time window *w* and a 3-dimensional copula separated with 1 day gap between each variable are employed. This means that

$$\mathbf{u} = (u_0, u_1, u_2) \tag{22}$$

where $u_0 = F_z(Z(t)), u_1 = F_z(Z(t+1)), u_2 = F_z(Z(t+2))$, then the deviation of local copula from global copula is defined by

$$\Delta c(\mathbf{u},t) = c^*_{\text{local}}(\mathbf{u},t) - c^*_{\text{global}}(\mathbf{u}) \tag{23}$$

For the first approach, the comparison of dependence structures between entire and local time series is done for detecting unusual dependence structures. To this end, copula distance type1 is defined by taking the copula distance between global and local copulas at each time step $t$

$$\begin{aligned} D_1(c,t) &= \int_0^1 \ldots \int_0^1 \left(c^*_{\text{global}}(\mathbf{u}) - c^*_{\text{local}}(\mathbf{u},t)\right)^2 du_1 \ldots du_n \\ &= \int_0^1 \ldots \int_0^1 \Delta c(\mathbf{u},t)^2 du_1 \ldots du_n \end{aligned} \tag{24}$$

Second, copula distance type2 is introduced for indicating the point at which the structure of copulas starts to change. For this method, the distance between two local copulas is calculated at two instants:

$$D_2(c,t) = \int_0^1 \ldots \int_0^1 \left(c^*_{\text{local}}\left(\mathbf{u}, t - \frac{w}{2}\right) - c^*_{\text{local}}\left(\mathbf{u}, t + \frac{w}{2}\right)\right)^2 du_1 \ldots du_n \tag{25}$$

Note that reference time is set to the middle of both time windows and shifted for *w/2* (days) from each other where the size of the time windows is *w*. Therefore, there is no overlapping part between the two time intervals of these two local copulas. For the comparison, the moving variance is introduced as follows:

$$E[Z(t)] = \frac{1}{w}\int_{t-w/2}^{t+w/2} z(a)da$$
$$\mathrm{Var}(t) = \frac{1}{w}\int_{t-w/2}^{t+w/2} (z(a) - E[Z(t)])^2 da \quad (26)$$

Figure 10 shows the result of $D_1(t), D_2(t)$ and $\mathrm{Var}(t)$ in the moving time window for the normalized discharge time series between 1940 to 2000 at 4 gauging stations located in the main stream of the Rhine (Andernach, Maxau ) and its two different tributaries (Cochem, Plochingen) in addition to the 90 % confidence intervals calculated for the Gaussian process fitted to the discharge data of Andernach.

First of all, the values of $D_1(t)$ and $D_2(t)$ at Cochem and Plochingen are bigger and more fluctuating than in general. The reason could be that their catchments and discharges are smaller, thus more sensitive to changes. Second, it can be said that the dependence structure is not homogeneous over the time period, but the local copula clearly deviates from the global copula for certain time periods. For example, the value of $D_1(t)$ is remarkably big around 1947, 1982 and 2000 for all the 4 discharge records (indicated by white arrows). $D_2(t)$ is also big around 1977 for all the data. The signal of $D_2(t)$ implies that a simultaneous change of runoff behavior occurred in this region in 1977, which can be related to the high value of $D_1(t)$ at 1982. $\mathrm{Var}(t)$ is also changing, but a direct relation with $D_1(t)$ and $D_2(t)$ is hard to recognize. Also the confidence interval of the Gaussian process is clearly smaller than the observed one. This indicates the copula distances of the stationary process are small while the nature process is non-stationary and its dependence structure is more varying.

For copula distance type1, the global copula can be considered as an average state of the copula, while the local copula can be regarded as a realization of a possible state of a copula at time step *t*. This concept

306 can be comparable to variance and leads to a new measure, copula variance, which is the summation of
307 copula distances between global and local copula over the time.

308 $$\mathrm{Var}_{\mathrm{cop}}(c) = \frac{1}{t_n - t_1}\int_{t_1}^{t_n} D_1(c,t)dt \tag{27}$$

309 Table 1 shows the variance and copula variance calculated for the 4 discharge data. The result
310 demonstrates that copula variance of the time series can be higher, even if the conventional variance is
311 lower for example in case of Maxau.

312 **4.1.2 Copula Distance for two time series**

313 In the previous section, copula variance was defined as a measure of the variability characteristic of the
314 copula itself. Here, it is determined whether covariance can be defined for two copula densities $c_1$ and $c_2$
315 from two time series as copula distance type3, which shows whether the variability characteristic of
316 copulas is related to each other. The measure introduced is:

317 $$D_3(c_1, c_2, t) = \int_0^1 ... \int_0^1 \Delta c_1(\mathbf{u}, t)\, \Delta c_2(\mathbf{u}, t) du_1 ... du_n \tag{28}$$

318 where

319 $$\begin{aligned} \Delta c_1(\mathbf{u}, t) &= c^*_{1,\mathrm{local}}(\mathbf{u}, t) - c^*_{1,\mathrm{global}}(\mathbf{u}) \\ \Delta c_2(\mathbf{u}, t) &= c^*_{2,\mathrm{local}}(\mathbf{u}, t) - c^*_{2,\mathrm{global}}(\mathbf{u}) \end{aligned} \tag{29}$$

320 By its definition, the value of $D_3(t)$ can be related to $D_1(t)$ because $D_3(t)$ compares the deviation of local
321 copulas from global copulas in a similar way to $D_1(t)$ in Eq. (26). In order to reduce the influence of
322 $D_1(t)$ on $D_3(t)$, copula distance type4 is introduced as a normalized measure bounded between -1 and 1
323 analogous to correlation.

324 $$D_4(c_1, c_2, t) = \frac{D_3(c_1, c_2, t)}{\sqrt{D_1(c_1, t)} \cdot \sqrt{D_1(c_2, t)}} \tag{30}$$

325 where $|D_4(c_1, c_2, t)| \le 1$. For comparison, covariance and correlation in a moving window are introduced for
326 two random variables $Z_1(t)$ and $Z_2(t)$ as follows:

$$\mathrm{Cov}(t) = \int_{t-w/2}^{t+w/2} (z_1(a) - E[Z_1(t)])(z_2(a) - E[Z_2(t)])da \tag{31}$$

$$\mathrm{Corr}(t) = \frac{\mathrm{Cov}(t)}{\sqrt{\mathrm{Var}(Z_1(t))} \cdot \sqrt{\mathrm{Var}(Z_2(t))}} \tag{32}$$

Figure 11 shows the copula distance between two time series $D_3(t)$ and $D_4(t)$ in addition to the covariance and correlation in a moving time window.

First, it can be said that the behavior of covariance and correlation in a moving window are different from $D_3(t)$ and $D_4(t)$. This implies these two copula based statistics exhibit different properties of the time series from ordinary statistics. Second, $D_3(t)$ shows high values around 1947, 1982 and 2000, which is similar to the case of $D_1(t)$ in Fig. 10. This indicates that unusual states of copulas in 4 discharge time series can be related to each other. Third, $D_4(t)$ is in general high except for the period around 1970 and 1990. This means, the temporal behavior of dependence structures for these 4 discharges are actually similar except for these periods even if $D_1(t)$ and $D_3(t)$ are small.

Copula covariance and copula correlation can be defined similar to copula variance in order to quantify the overall behavior of two time series.

$$\mathrm{Cov}_{\mathrm{cop}}(c_1, c_2) = \frac{1}{t_2 - t_1} \int_{t_2}^{t_1} D_3(t)dt \tag{33}$$

$$\mathrm{Corr}_{\mathrm{cop}}(c_1, c_2) = \frac{\mathrm{Cov}_{\mathrm{cop}}(c_1, c_2)}{\sqrt{\mathrm{Var}_{\mathrm{cop}}(c_1)} \cdot \sqrt{\mathrm{Var}_{\mathrm{cop}}(c_2)}} \tag{34}$$

where $\left|\mathrm{Corr}_{\mathrm{cop}}(c_1, c_2)\right| \le 1$ and its derivation can be found in appendix A. In Table 2, these copula based statistics are compared with ordinary statistics. For example, Cochem and Plochingen are located remotely in different tributaries, thus covariance and correlation are lower than the others, but copula covariance and copula correlation are not the lowest.

The measures using copula distance are different from the conventional statistics. This behavior can be explained by the fact that the autocopula has more substantial information about temporal dependence

structure than the autocorrelation. Using these measures might enable us to take advantage of a different way of seeing the dependence between time series.

What is new in the analysis of this section is that (i) measures based on copula distance show the different properties of time series in comparison to conventional statistics and (ii) there are significant signals of copula distances for certain time periods in common to all the discharge data.

### 4.2 Copula based Stochastic Analysis with API and a Hydrological Model

The difficulty of analyzing discharge time series in order to detect catchment change is that it is not clear whether the temporal change of stochastic information is caused by catchment change or merely by random behavior of precipitation. To gain an understanding of this process, we attempted to eliminate the influence of precipitation using, first, an Antecedent Precipitation Index (API) for comparison with discharge, second, using a hydrological model with the parameter sets calibrated and fixed for the entire simulation time period.

#### 4.2.1 Copula Distance Analysis with API

An API time series, which is generated from observed precipitation time series and behaves similarly to discharge, is used instead of precipitation.

$$\mathrm{API}(t+1) = \alpha \mathrm{API}(t) + P(t+1) \tag{35}$$

where $P(t)$ is daily precipitation (mm/day), $\mathrm{API}(t)$ is time series of API (mm/day) and $\alpha = 0.85$ was chosen. The assumption for this method is that the API time series has the stochastic information purely originated from the precipitation, while observed discharge is influenced by both catchment and precipitation characteristics. If the stochastic information derived from these two data sets is the same, this indicates that the stochastic turbulence is originating from precipitation; otherwise the change is from the catchment.

For this investigation, precipitation data was carefully chosen for 4 regions northwest, northeast, southwest and central) of Baden-Württemberg (Germany) so that they have several almost continuous

daily records between 1935 and 2005. Figure 12 shows the locations of measuring stations. The precipitation time series were aggregated into one for each region by taking their daily average, then 4 API time series were calculated in total by Eq. (35). Figure 13 shows the resulting copula distances $D_1(t), D_2(t)$ and moving average $\mathrm{Var}(t)$ for API time series with the 90% confidence intervals of the Gaussian process. Figure 14 shows the result of copula distances $D_3(t), D_4(t)$ and moving covariance and correlation for API time series.

What can be recognized first from Fig. 13 is that the magnitudes of $D_1(t)$ and $D_2(t)$ are smaller than the case of discharge. This is considered to be a result of aggregation of precipitation time series and adoption of API, but some signals can be still identified: $D_1(t)$ around 1947 and 2000 is high, but not as high for 1982. The signal of $D_2(t)$ which was detected around 1977 in Fig. 11 does not seem to exist for API. This is even more clear for $D_3(t)$ in Fig. 14 in that there is no common change of the dependence structure around 1982 in API time series. This is interesting due to the following implications: (i) the noises of $D_1(t)$ in Fig. 13 were reduced and signals in common were amplified (ii) the unusual state of the copula around 1982 is not caused by precipitation, but could be caused by the catchment change.

For further verification, copula distance type3 and type4 between discharge and API time series were calculated as shown in Fig. 15. This result also shows there is no clear relation between API and discharge time series around 1982.

### 4.2.2 Copula based analysis with a hydrological model

In this section, simulated discharges time series are generated by a conceptual hydrological model, HBV (Bergström 1976 ; Bergström, Singh, and others 1995), which takes daily precipitation and temperature records as input and simulates discharges for smaller catchments as an example of discharge, to compare with observed discharge, in order to check if differences might occur due to the method.

Thus the idea behind this methodology is similar to the case of API: a hydrological model with the parameters fixed for the entire time period represents the catchment not influenced by anthropogenic impacts. Then, the discharges simulated by this model should not depend on catchment change, while observed discharge is assumed to be influenced by both catchment and precipitation.

For the study area, the Upper Neckar Catchment was chosen as shown in Fig. 12. One parameter set needed for this model constitutes of 13 parameters which are calibrated based on the Nash–Sutcliffe model efficiency coefficient using the simulated annealing algorithm for the period between 1960 and 2000. Then, 30 parameter sets are independently calibrated in total and, subsequently, 30 simulated discharges time series are generated to compare with one observed discharge.

Figure 16 shows the result of copula based analysis calculated for single time series ($D_1(t), D_2(t), A_{2,\min}(t)$). It can be seen that $A_{2,\min}(t)$ in Fig. 16 (top) that (i) fluctuations of $A_{2,\min}(t)$ of observed and simulated discharge are locally identical. This implies that the short term behavior of $A_{2,\min}(t)$ is originated from the temporal behavior of precipitation but (ii) there exists a change of trend around 1976: $A_{2,\min}(t)$ of observed discharge is slightly bigger than simulated before 1976, while $A_{2,\min}(t)$ of observed discharge clearly undershoot the simulated ones of after 1976. This change of trend was also seen in the previous analyses ($D_2(t)$ in Fig. 10). Furthermore, $D_1(t)$ in Fig. 16 (middle) is high before 1976 which indicates the state of the copula is different from the rest, while the result of simulated discharges does not show such tendency. $D_2(t)$ in Fig. 16 (bottom) indicates the change of dependence structure happened around 1970 and 1977. These results using the HBV model indicate the change of the dependence structure detected using copulas around 1976 is not caused by the random behavior of precipitation, but by the behavior of the catchment itself.

The fact and the notion obtained in this section is that (i) both results from API and HBV based on copula measures indicate that the catchment changed around 1976 and (ii), by comparing the simulated discharge with observed discharge, the origin of the change of stochastical information can be assessed.

## Conclusion

In this paper the application of copulas for hydrological time series data is newly explored for the detection of catchment characteristics and their temporal changes.

1. A Copula based measure of asymmetry, $A_1(k)$ and $A_2(k)$, was defined and newly applied for the identification of catchment characteristics. Indeed, it was presumed that asymmetry2 is related to the runoff characteristics.

2. The relation between asymmetry2 and catchment characteristics was tested for 77 discharge records. $A_{2,\min}$ has a certain relation especially with the size of catchments and this strengthens the notion that asymmetry2 of discharge data can be used to describe the catchment characteristic and state.

3. $A_{2,\min}(t)$ was defined for evaluating the temporal change of asymmetry2 and calculated as an indicator of the catchment state. The result indicates $A_{2,\min}(t)$ keeps changing coincidentally with time. However, it is difficult to explain the causality, at least, by long term behavior of discharge and temperature time series.

4. A method based on copula distance was examined for the investigation of temporal behavior of hydrological time series. This measure can detect the time period where dependence structure is unusual and its interdependency between different time series. Clear signals were detected that the dependence structure is unusual for a certain time period and this signal was not found by investigating the time series with variance, covariance or correlation.

5. API time series were calculated for each region in the Baden-Württemberg state and simulated discharge time series were generated using the HBV model for the Upper Neckar Catchment. These are the data not influenced by catchment change, thus compared with observed discharge to assess the anthropogenic impacts. The results showed that there was a signal detected only in the observed discharge

441 around 1982, but not in the API or simulated time series, which implies the anthropogenic impacts on the
442 catchment. Also it was shown in the results of copula asymmetry that the trend clearly changed around
443 1976.

444 The results of copula based analysis of hydrological time series seem to support the assumption that the
445 catchment had started to change around 1976 and stayed unusual until 1990. These changes could
446 correspond to the construction of flood retention basins started around 1982 (Lammersen et al., 2002) and
447 ecological flooding strategy, which let small floods to happen for the rehabilitation of ecological systems
448 in the floodplain, introduced in the Upper Rhine since 1989 (Siepe, 2006).

449 Copulas can be seen as an alternative method to analyze hydrological time series data by focusing on the
450 dependence structure, but further exploratory applications and theoretical developments are expected. The
451 copula based measures introduced in this study can be related to the potential model uncertainty, that is,
452 how much the natural system is varying. Empirical autocopula analysis is a more data driven approach
453 which retains more information than the copulas estimated with parametric methods, but it is also
454 numerically demanding. The effective way to analyze time series and build up a time series model based
455 on copulas can be further explored.

456

## Appendix A

Suppose that a random variable at time *t* is denoted as $X(t)$ and $c_X(\mathbf{u},t)$ is an autocopula obtained from $X(t)$. Assuming $c_{X,\text{mean}}(\mathbf{u})$ as an average state of $c_X(\mathbf{u},t)$, deviation of copula $\Delta c_X(\mathbf{u},t)$ at time *t* is defined by

$$\Delta c_X(\mathbf{u},t) = c_X(\mathbf{u},t) - c_{X,\text{mean}}(\mathbf{u}) \tag{A1}$$

For the empirical case, $c_X(\mathbf{u},t)$ and $c_{X,\text{mean}}(\mathbf{u})$ can be regarded as local copula and global copula respectively similar to Eq. (29). Since global and local copula are empirical copula density as defined in Eq. (8), $\Delta c_X(\mathbf{u},t)$ can be regarded as a vector of values on finite number of grids:

$$\boldsymbol{\Delta}\mathbf{c}_X(t) = \left(\Delta c_{X,1}(t), \Delta c_{X,2}(t), \ldots, \Delta c_{X,i}(t), \ldots, \Delta c_{X,N}(t)\right) \tag{A2}$$

where $\Delta c_{X,i}(t)$ denotes the value of copula density at *i*-th grid and *N* is the number of grids. From Cauchy-Schwarz inequality

$$\left\|\boldsymbol{\Delta}\mathbf{c}_X(t)\right\|\left\|\boldsymbol{\Delta}\mathbf{c}_Y(t)\right\| \geq \left|\left\langle \boldsymbol{\Delta}\mathbf{c}_X(t), \boldsymbol{\Delta}\mathbf{c}_Y(t)\right\rangle\right|^2 \tag{A3}$$

where $\left\|\boldsymbol{\Delta}\mathbf{c}_X(t)\right\|$ is norm and $\left\langle \boldsymbol{\Delta}\mathbf{c}_X(t), \boldsymbol{\Delta}\mathbf{c}_Y(t)\right\rangle$ is inner product of vector $\boldsymbol{\Delta}\mathbf{c}_X(t)$ and $\boldsymbol{\Delta}\mathbf{c}_Y(t)$ .Then

$$\begin{aligned}\left\|\boldsymbol{\Delta}\mathbf{c}_X(t)\right\| &= \sum_{i=1}^{N} \Delta c_{X,i}(t)^2 \\ &= \int_0^1 \ldots \int_0^1 \left(\Delta c_X(\mathbf{u},t)\right)^2 du_1 \ldots du_n = D_1(c_X,t)\end{aligned} \tag{A4}$$

$$\begin{aligned}&\left|\left\langle \boldsymbol{\Delta}\mathbf{c}_X(t), \boldsymbol{\Delta}\mathbf{c}_Y(t)\right\rangle\right|^2 \\ &= \left\langle \boldsymbol{\Delta}\mathbf{c}_X(t), \boldsymbol{\Delta}\mathbf{c}_Y(t)\right\rangle = \sum_{i=1}^{N} \Delta c_{X,i}(t) \cdot \Delta c_{Y,i}(t) \\ &= \int_0^1 \cdots \int_0^1 \Delta c_X(\mathbf{u},t)\, \Delta c_Y(\mathbf{u},t) du_1 \ldots du_n = D_3(c_X, c_Y, t)\end{aligned} \tag{A5}$$

$$\frac{\left|\left\langle \boldsymbol{\Delta}\mathbf{c}_X(t), \boldsymbol{\Delta}\mathbf{c}_Y(t)\right\rangle\right|^2}{\left\|\boldsymbol{\Delta}\mathbf{c}_X(t)\right\|\left\|\boldsymbol{\Delta}\mathbf{c}_Y(t)\right\|} = \frac{D_3(c_X, c_Y, t)^2}{D_1(c_X,t) \cdot D_1(c_Y,t)} = D_4(c_X, c_Y, t)^2 \leq 1 \tag{A6}$$

Therefore $\left|D_4\left(c_X, c_Y, t\right)\right| \le 1$in Eq. (30). Above inequality is valid for certain time point *t* and summing up Eq. (A6) for all the time steps *t* leads to

$$\sum_{t=1}^{T}\left(\left\|\Delta\mathbf{c}_X\left(t\right)\right\| \cdot \left\|\Delta\mathbf{c}_Y\left(t\right)\right\|\right) \ge \sum_{t=1}^{T}\left|\left\langle\Delta\mathbf{c}_X\left(t\right), \Delta\mathbf{c}_Y\left(t\right)\right\rangle\right|^2 \tag{A7}$$

where $T$ is the number of time steps. $\left\|\Delta\mathbf{c}_X\left(t\right)\right\|$ is a norm and can be denoted for simplicity as $x_t = \left\|\Delta\mathbf{c}_X\left(t\right)\right\|$. Then

$$\sum_{t=1}^{T}\left(\left\|\Delta\mathbf{c}_X\left(t\right)\right\| \cdot \left\|\Delta\mathbf{c}_Y\left(t\right)\right\|\right) = \left\langle\mathbf{x}, \mathbf{y}\right\rangle \tag{A8}$$

where $\mathbf{x} = \left(x_1, x_2, ...x_T\right), \mathbf{y} = \left(y_1, y_2, ...y_T\right)$ for $t = 1...T$. Again from Cauchy-Schwarz inequality

$$\left|\left\langle\mathbf{x}, \mathbf{y}\right\rangle\right|^2 \le \left\|\mathbf{x}\right\| \cdot \left\|\mathbf{y}\right\| \tag{A9}$$

where

$$\begin{aligned}\left\|\mathbf{x}\right\| \cdot \left\|\mathbf{y}\right\| &= \sum_{t=1}^{T} x_t^{\,2} \cdot \sum_{t=1}^{T} y_t^{\,2} = \sum_{t=1}^{T}\left\|\Delta\mathbf{c}_X\left(t\right)\right\|^2 \cdot \sum_{t=1}^{T}\left\|\Delta\mathbf{c}_Y\left(t\right)\right\|^2 \\ &= \sum_{t=1}^{T} D_1\left(c_X, t\right)^2 \cdot \sum_{t=1}^{T} D_1\left(c_Y, t\right)^2 = T^2 \cdot \mathrm{Var}_{\mathrm{cop}}\left(c_X\right) \cdot \mathrm{Var}_{\mathrm{cop}}\left(c_Y\right)\end{aligned} \tag{A10}$$

$$\begin{aligned}\left\langle\mathbf{x}, \mathbf{y}\right\rangle &= \sum_{t=1}^{T}\left(x_t \cdot y_t\right) = \sum_{t=1}^{T}\left(\left\|\Delta\mathbf{c}_X\left(t\right)\right\| \cdot \left\|\Delta\mathbf{c}_Y\left(t\right)\right\|\right) \ge \sum_{t=1}^{T}\left|\left\langle\Delta\mathbf{c}_X\left(t\right), \Delta\mathbf{c}_Y\left(t\right)\right\rangle\right|^2 \\ &= \sum_{t=1}^{T} D_3\left(c_X, c_Y, t\right) = T \cdot \mathrm{Cov}_{\mathrm{cop}}\left(c_X, c_Y\right)\end{aligned} \tag{A11}$$

Then $\left|\left\langle\mathbf{x}, \mathbf{y}\right\rangle\right|^2 \le \left\|\mathbf{x}\right\| \cdot \left\|\mathbf{y}\right\|$ indicates

$$\begin{aligned}\left|\mathrm{Cov}_{\mathrm{cop}}\left(c_X, c_Y\right)\right|^2 &\le \mathrm{Var}_{\mathrm{cop}}\left(c_X\right) \cdot \mathrm{Var}_{\mathrm{cop}}\left(c_Y\right) \\ \left|\mathrm{Corr}_{\mathrm{cop}}\right| &= \frac{\mathrm{Cov}_{\mathrm{cop}}\left(c_X, c_Y\right)}{\sqrt{\mathrm{Var}_{\mathrm{cop}}\left(c_X\right)} \cdot \sqrt{\mathrm{Var}_{\mathrm{cop}}\left(c_Y\right)}} \le 1\end{aligned} \tag{A12}$$

## Acknowledgment

Fundamental research of this paper was initiated by the German Federal Institute of Hydrology with financial support. Special thanks are given to the Global Runoff Data Centre (GRDC) in Germany for offering the discharge data and the German Meteorological Service (DWD) for precipitation and temperature data. The authors thank the reviewers for their care in examining this work.

## References

Bárdossy, a., Pegram, G., 2009. Copula based multisite model for daily precipitation simulation. Hydrol. Earth Syst. Sci. Discuss. 6, 4485–4534. doi:10.5194/hessd-6-4485-2009

Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. Water Resour. Res. 42, W11416. doi:10.1029/2005WR004754

Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copulas. Water Resour. Res. 44, W07412. doi:10.1029/2007WR006115

Bergrström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments, Bulletin Series A, A]: [Bulletin series. Department of Water Resources Engineering, Lund Institute of Technology, University of Lund.

Bergstrom, S., 1995. The HBV Model. Singh, V.P. (Ed.), Comput. Model. Watershed Hydrol. 443–476.

Box, G.E.P., Jenkins, G.M., 1976. Time series analysis: forecasting and control, revised ed. Holden-Day, San Francisco,USA.

Brahimi, B., Chebana, F., Necir, A., 2014. Copula representation of bivariate L-moments: a new estimation method for multiparameter two-dimensional copula models. Statistics (Ber). 1–25.

De Michele, C., Salvadori, G., 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. J. Geophys. Res. Atmos. 108, 4067. doi:10.1029/2002JD002534

Gao, P., Geissen, V., Ritsema, C., Mu, X.-M., Wang, F., 2012. Impact of climate change and anthropogenic activities on stream flow and sediment discharge in the Wei River basin, China. Hydrol. Earth Syst. Sci. Discuss. 9, 3933–3959. doi:10.5194/hessd-9-3933-2012

Grimaldi, S., 2004. Linear parametric models applied to daily hydrological series. J. Hydrol. Eng. 9, 383–391. doi:10.1061/(ASCE)1084-0699(2004)9:5(383)

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci. 454, 903–995.

519 Joe, H., 1997. Multivariate models and multivariate dependence concepts. Chapman&Hall, London.

520 Karlsson, I.B., Sonnenborg, T.O., Jensen, K.H., Refsgaard, J.C., 2014. Historical trends in precipitation
521 and stream discharge at the Skjern River catchment, Denmark. Hydrol. Earth Syst. Sci. 18, 595–610.
522 doi:10.5194/hess-18-595-2014

523 Lammersen, R., Engel, H., Van de Langemheen, W., Buiteveld, H., 2002. Impact of river training and
524 retention measures on flood peaks along the Rhine. J. Hydrol. 267, 115–124. doi:10.1016/S0022-
525 1694(02)00144-0

526 Li, J., 2010. Application of copulas as a new geostatistical tool. PhD Thesis. Nr. 187. University of
527 Stuttgart, Germany

528 Nelsen, R.B., 2006. An Introduction to Copulas. Springer, New York. doi:10.1007/0-387-28678-0

529 Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. Appl. Stat. 126–135.

530 Samaniego, L., Bárdossy, A., Kumar, R., 2010. Streamflow prediction in ungauged catchments using
531 copula-based dissimilarity measures. Water Resour. Res. 46, W02506. doi:10.1029/2008WR007695

532 Serfling, R., Xiao, P., 2007. A contribution to multivariate L-moments: L-comoment matrices. J. Multivar.
533 Anal. 98, 1765–1781. doi:10.1016/j.jmva.2007.01.008

534 Sharifdoost, M., Mahmoodi, S., Pasha, E., 2009. A statistical test for time reversibility of stationary finite
535 state Markov chains. Appl. Math. Sci. 52, 2563–2574.

536 Siepe, A., 2006. Dynamische Überflutungen am Oberrhein : Entwicklungs-Motor für die Auwald-Fauna.
537 Stand 149–158.

538 Singh, S.K., McMillan, H., Bárdossy, A., 2013. Use of the data depth function to differentiate between
539 case of interpolation and extrapolation in hydrological model prediction. J. Hydrol. 477, 213–228.
540 doi:10.1016/j.jhydrol.2012.11.034

541 Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de
542 statistique de l'Université de Paris. Publications de l'Institut de Statistique de L'Université de Paris 8.

543 Sugimoto, T., 2014. Copula based stochastic analysis of discharge time series. PhD Thesis. Nr. 232.
544 University of Stuttgart, Germany

545 Wu, C.S., Yang, S.L., Lei, Y.P., 2012. Quantifying the anthropogenic and climatic impacts on water
546 discharge and sediment load in the Pearl River (Zhujiang), China (1954-2009). J. Hydrol. 452-453, 190–
547 204. doi:10.1016/j.jhydrol.2012.05.064

548

**Table 1.** Variance and copula variance calculated for 4 discharge time series (ANDE:Andernach, COCH:Cochem, MAXA:Maxau, PLOC:Plochingen)

| | ANDE | COCH | MAXA | PLOC |
|---|---|---|---|---|
| Var | 1.79 | 2.24 | 1.75 | 2.72 |
| $Var_{cop}$ ($\times 10^{-5}$) | 3.01 | 1.64 | 5.39 | 1.27 |

**Table 2.** Covariance, correlation, copula covariance and copula correlation between 4 discharge data (AN:Andernach, CO:Cochem, MA:Maxau, PL:Plochingen)

| | AN-CO | AN-MA | AN-PL | CO-MA | CO-PL | MA-PL |
|---|---|---|---|---|---|---|
| Cov | 1.68 | 1.60 | 1.33 | 1.38 | 1.31 | 1.41 |
| Cor | 0.84 | 0.90 | 0.60 | 0.70 | 0.53 | 0.64 |
| $Cov_{cop}$ ($\times 10^{-6}$) | 4.90 | 3.40 | 3.39 | 7.16 | 9.90 | 5.47 |
| $Corr_{cop}$ | 0.60 | 0.77 | 0.46 | 0.71 | 0.60 | 0.59 |

**Table 3.** Variance and copula variance calculated for API time series of 4 regions in the Baden-Württemberg state of Germany (C: Central, SW: South-West, NW: North-West, NE: North-East)

| | C | SW | NW | NE |
|---|---|---|---|---|
| Var | 1.70 | 1.66 | 1.72 | 1.78 |
| $Var_{cop}$ ($\times 10^{-6}$) | 3.00 | 4.02 | 3.35 | 3.21 |

**Table 4.** Covariance, correlation, copula covariance and copula correlation between API time series from 4 regions in the Baden-Württemberg state of Germany (C: Central, SW: South-West, NW: North-West, NE: North-East)

| | C-SW | C-NW | C-NE | SW-NW | SW-NE | NW-NE |
|---|---|---|---|---|---|---|
| Cov | 1.35 | 1.33 | 1.44 | 1.25 | 1.41 | 1.42 |
| Cor | 0.80 | 0.77 | 0.84 | 0.74 | 0.84 | 0.83 |
| $Cov_{cop}$ ($\times 10^{-7}$) | 1.46 | 1.16 | 8.94 | 4.42 | 1.11 | 8.80 |
| $Corr_{cop}$ | 0.36 | 0.29 | 0.29 | 0.09 | 0.26 | 0.24 |

572 Figure Captions

573

574 **Fig. 1.** Locations of 7 discharge gauging stations in the Upper Rhine Region

575 **Fig. 2.** Visualization of $a_1(u_t, u_{t+k}) = (u_t - 0.5)(u_{t+k} - 0.5)((u_t - 0.5) + (u_{t+k} - 0.5))$ (left) and
576 $a_2(u_t, u_{t+k}) = (u_t - 0.5)(u_{t+k} - 0.5)((u_t - 0.5) - (u_{t+k} - 0.5))$ (right) which displays the contribution of single
577 realization of $(U_t, U_{t+k})$ to asymmetry functions
578 $A_1(k) = E\left[(U_t - 0.5)(U_{t+k} - 0.5)\left((U_t - 0.5) + (U_{t+k} - 0.5)\right)\right]$ and
579 $A_2(k) = E[-(U_t - 0.5)(U_{t+k} - 0.5)((U_t - 0.5) - (U_{t+k} - 0.5))]$

580 **Fig. 3.** Sketch of the transformation of the values from sample hydrograph (left) to the points on scatterplot
581 of ranks (right): empirical copula calculated from two values separated by time lag $k$ = 1 (days) in a
582 discharge time series of Andernach where rank correlation is 0.9870, $A_1(k=1) = -0.0002398$ and
583 $A_2(k=1) = -0.00011037$. The possible combinations of high and low values, which has large impacts on
584 asymmetry, are numbered: (1) low to high, (2) high to high, (3) high to low, (4) low to low. Negative
585 contribution to $A_2$ is drawn with red circle and positive contribution with blue oval.

586 **Fig. 4.** Annual cycles of mean discharge measured at seven sites in the Rhine basin after smoothing (left)
587 and annual cycle of standard deviation after smoothing (right)

588 **Fig. 5.** Discharge time series measured at seven sites in the Rhine basin between 1950 and 1955 before
589 applying normalization (upper left) and after applying normalization (upper right). $A_2(k)$ calculated for
590 entire time series before applying normalization (bottom left) and after applying normalization (bottom
591 right) with 90% confidence intervals (grey) calculated for 100 realizations of Gaussian process (dashed
592 line is $A_2(k)$ calculated for one of the realization of Gaussian process ).

593 **Fig. 6.** Relation between asymmetry of discharge data and catchment characteristics: $A_{2,\min}$ of discharge
594 and catchment area (top), $L_{2,\min}$ of discharge and catchment area (middle), $A_{2,\min}$ of discharge and $L_{2,\min}$ of
595 discharge (bottom)

596 **Fig. 7.** Temporal change of asymmetry2 : $A_{2,\min}(t)$ for 7 discharge records and, for comparison, confidence
597 intervals calculated from the Gaussian process (90% confidence interval with grey color and 60%
598 confidence interval with dark grey color) and one of its realizations (dashed line)

599 **Fig. 8.** Moving average and standard deviation of the 7 daily discharge records for the window size $w$ =
600 3000 (days)

601 **Fig. 9.** Annual minimum (upper panel) and mean of aggregated daily temperature (lower pannel) in the
602 Baden-Württemberg state of Germany

603 **Fig. 10.** Copula distances of discharge time series in moving time window: variance (top), distance type1
604 (middle) and distance type2 (bottom), each panel containing the 80% confidence interval of Gaussian
605 process and one of its realization (dashed line). The arrows point 1947, 1982, 2000 and 1977 in which the

clear signals of anomalies are detected for all four discharge time series: Andernach(ANDE), Cochem(COCH), Maxau(MAXA) and Plochingen(PLOC)).

**Fig. 11.** Copula distances of discharge time series in moving time window: covariance (top), correlation (second), copula distance type3 (third) and copula distance type4 (bottom). The arrows point 1947, 1982 and 2000 in which the clear signals of anomalies are detected for the comparisons between 4 discharge time series: Andernach(ANDE), Cochem(COCH), Maxau(MAXA), Plochingen(PLOC)

**Fig. 12.** Locations of the precipitation gauge stations within Baden-Württemberg (Germany) indicated by coloured circles. Upper Neckar catchment is identified by the light green area and the location of the gauging station is indicated by a square

**Fig. 13.** Copula distances of API time series in moving time window: variance (top), copula distance type1 (middle) and copula distance type2 (bottom) where 'C' denotes central, 'SW' denotes southwest, 'NW' denotes northwest and 'NE' denotes northeast part of the Baden-Württemberg State of Germany, each containing 80% confidence interval of Gaussian process and one of its realization (dashed line). The arrows indicate the years in which anomalies are detected in the previous analysis (Fig. 10)

**Fig. 14.** Copula distances of API time series in moving time window: covariance (top), correlation (second), copula distance type3 (third) and copula distance type4 (bottom). The arrows indicate the years in which anomalies are detected in the previous analysis (Fig. 11)

**Fig. 15.** Copula distance type3 (top) and type4 (bottom) between 4 discharge and 1 API time series which is aggregated for all the daily precipitations depicted in Fig. 12. The arrows indicate the years in which anomalies are detected in the previous analysis (Fig. 11)

**Fig. 16.** Copula asymmetry and copula distances for 30 simulated and one observed discharge time series at Plochingen between 1965 and 2000: $A_{2,\min}$ for the time lag $k = 2$ days (top), copula distance type1 (middle), copula distance type2 (bottom)

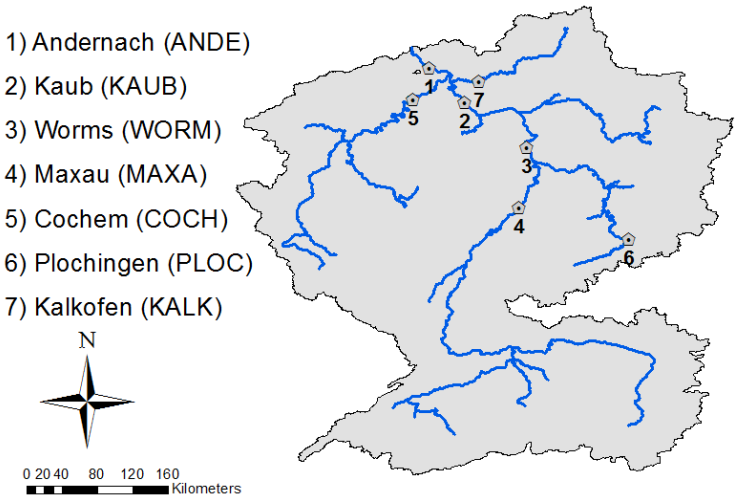



**Fig. 1.** Locations of 7 discharge gauging stations in the Upper Rhine Region

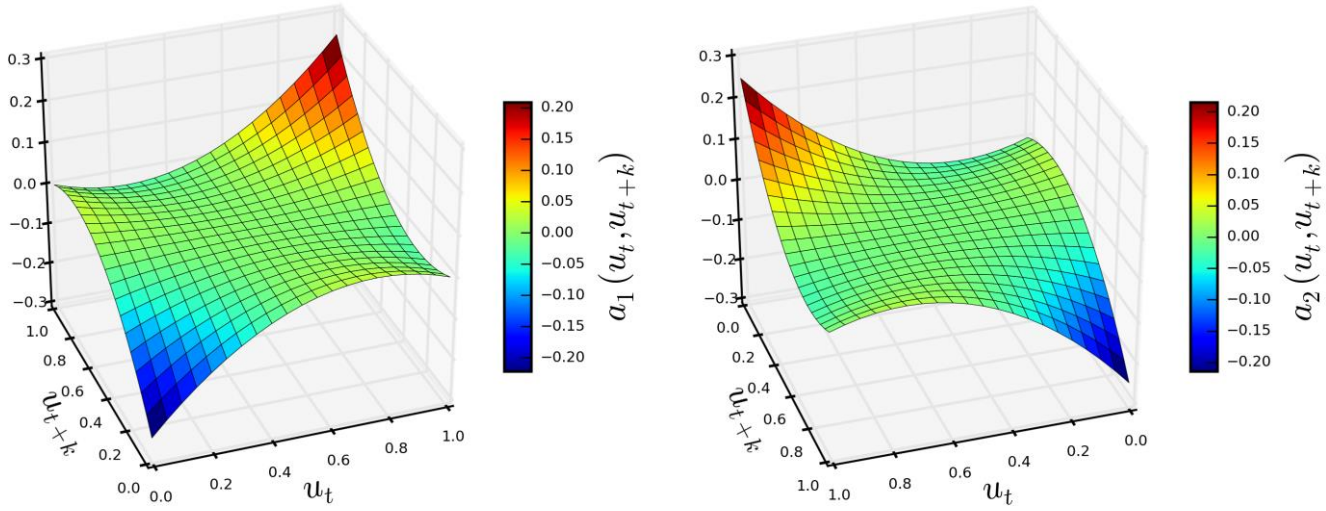



**Fig. 2.** Visualization of $a_1(u_t, u_{t+k}) = (u_t - 0.5)(u_{t+k} - 0.5)((u_t - 0.5) + (u_{t+k} - 0.5))$ (left) and $a_2(u_t, u_{t+k}) = (u_t - 0.5)(u_{t+k} - 0.5)((u_t - 0.5) - (u_{t+k} - 0.5))$ (right) which displays the contribution of single realization of $(U_t, U_{t+k})$ to asymmetry functions $A_1(k) = E\left[(U_t - 0.5)(U_{t+k} - 0.5)\left((U_t - 0.5) + (U_{t+k} - 0.5)\right)\right]$ and $A_2(k) = E[-(U_t - 0.5)(U_{t+k} - 0.5)((U_t - 0.5) - (U_{t+k} - 0.5))]$

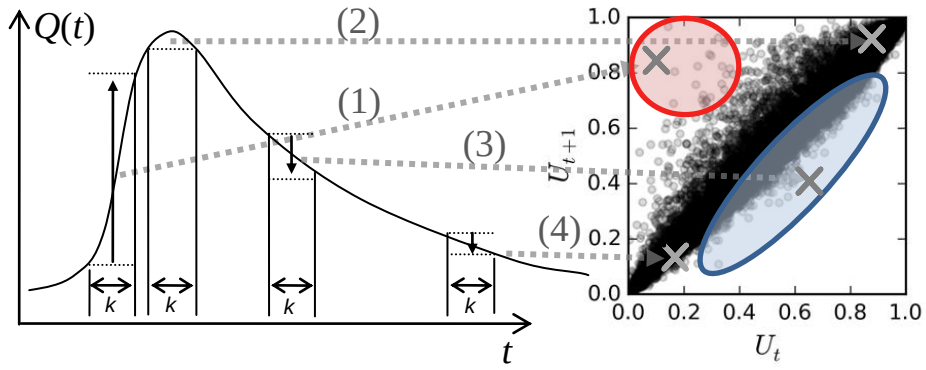



**Fig. 3.** Sketch of the transformation of the values from sample hydrograph (left) to the points on scatterplot of ranks (right): empirical copula calculated from two values separated by time lag $k$ = 1 (days) in a discharge time series of Andernach where rank correlation is 0.9870, $A_1(k=1) = -0.0002398$ and $A_2(k=1) = -0.00011037$. The possible combinations of high and low values, which has large impacts on asymmetry, are numbered: (1) low to high, (2) high to high, (3) high to low, (4) low to low. Negative contribution to $A_2$ is drawn with red circle and positive contribution with blue oval.

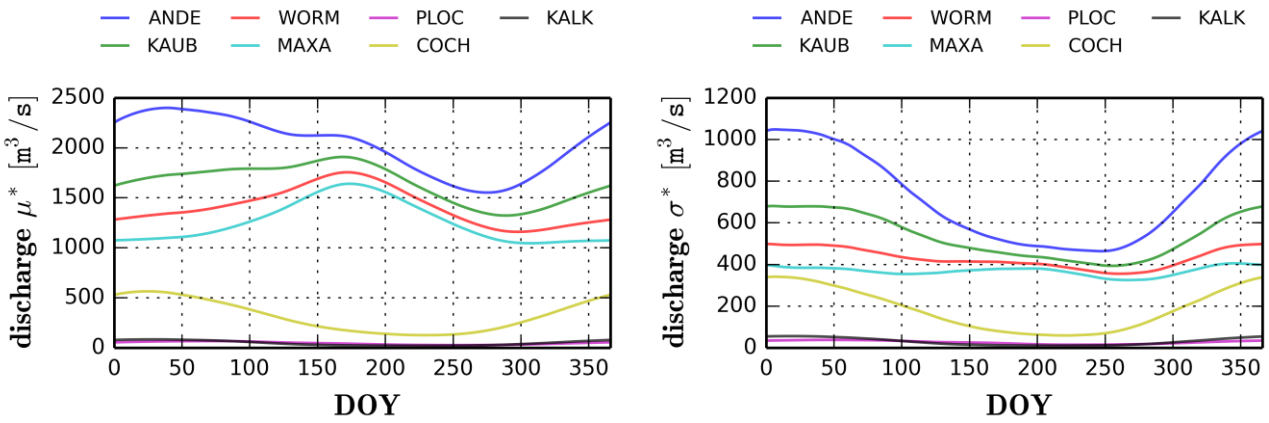



**Fig. 4.** Annual cycles of mean discharge measured at seven sites in the Rhine basin after smoothing (left) and annual cycle of standard deviation after smoothing (right)

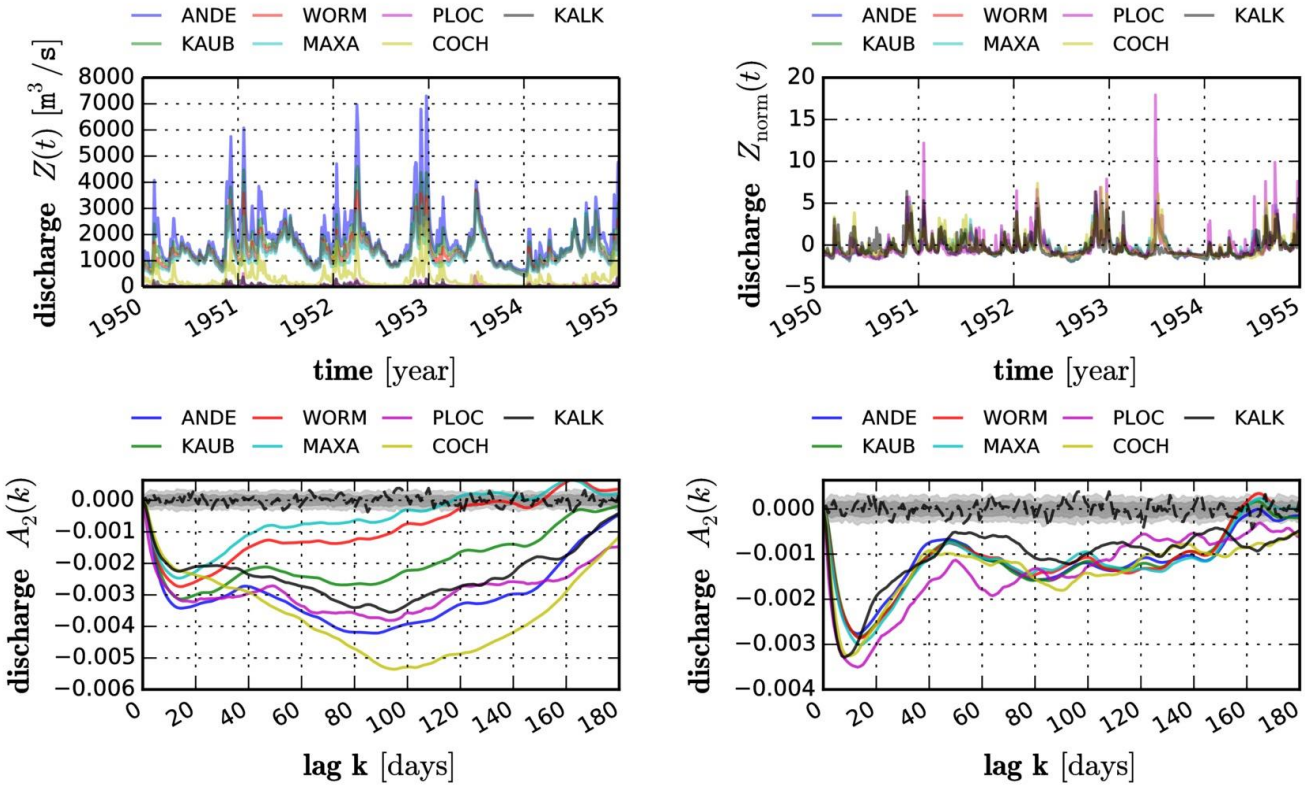
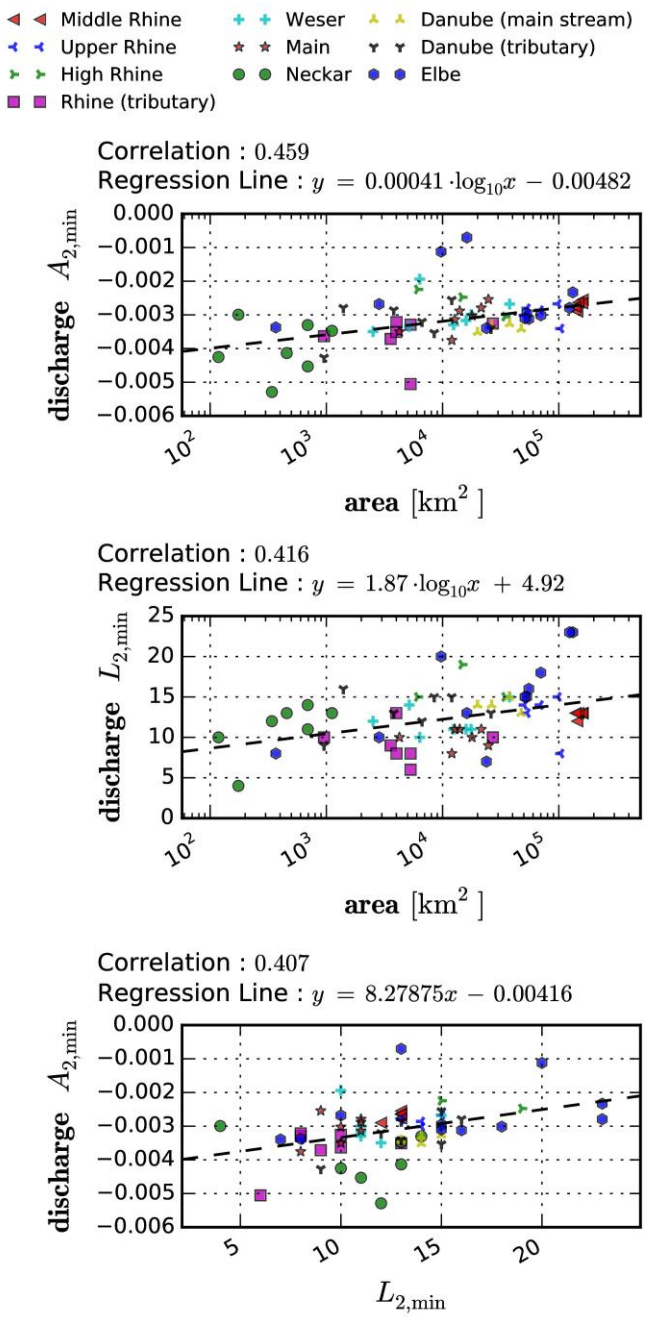



**Fig. 5.** Discharge time series measured at seven sites in the Rhine basin between 1950 and 1955 before applying normalization (upper left) and after applying normalization (upper right). $A_2(k)$ calculated for entire time series before applying normalization (bottom left) and after applying normalization (bottom right) with 90% confidence intervals (grey) calculated for 100 realizations of Gaussian process (dashed line is $A_2(k)$ calculated for one of the realization of Gaussian process ).

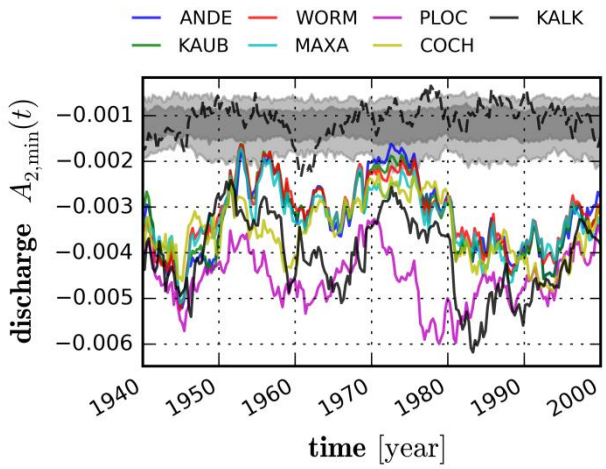



662

663 **Fig. 6.** Relation between asymmetry of discharge data and catchment characteristics: $A_{2,\min}$ of discharge

664 and catchment area (top), $L_{2,\min}$ of discharge and catchment area (middle), $A_{2,\min}$ of discharge and $L_{2,\min}$ of

665 discharge (bottom)



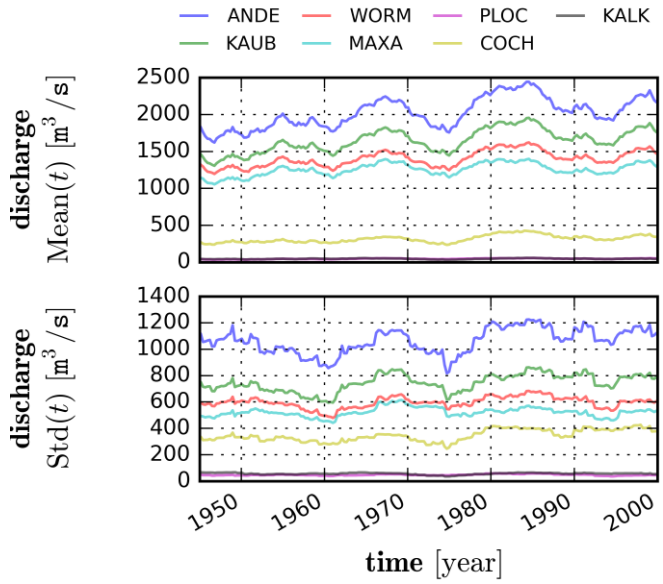

666

667 **Fig. 7.** Temporal change of asymmetry2 : $A_{2,\min}(t)$ for 7 discharge records and, for comparison, confidence

668 intervals calculated from the Gaussian process (90% confidence interval with grey color and 60%

669 confidence interval with dark grey color) and one of its realizations (dashed line)

670




671

672 **Fig. 8.** Moving average and standard deviation of the 7 daily discharge records for the window size *w* =

673 3000 (days)

674

675

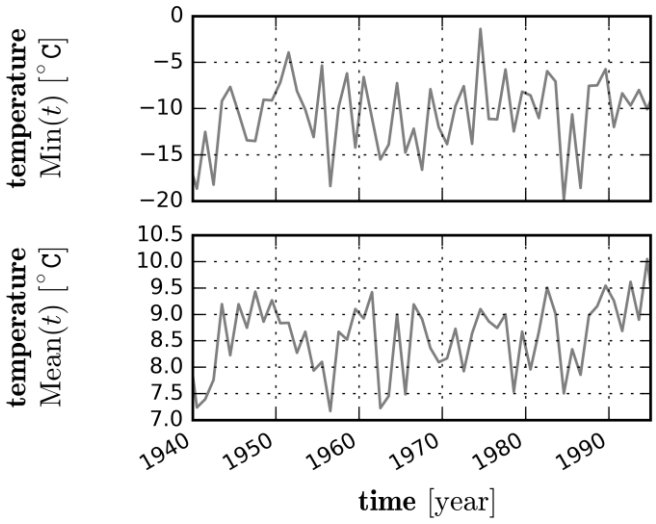



**Fig. 9.** Annual minimum (upper panel) and mean of aggregated daily temperature (lower pannel) in the Baden-Württemberg state of Germany

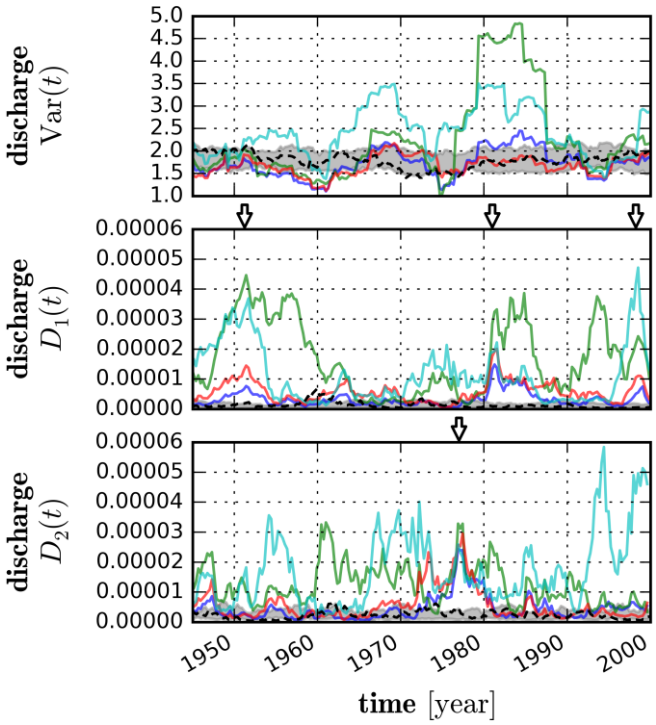



**Fig. 10.** Copula distances of discharge time series in moving time window: variance (top), distance type1 (middle) and distance type2 (bottom), each panel containing the 80% confidence interval of Gaussian process and one of its realization (dashed line). The arrows point 1947, 1982, 2000 and 1977 in which the

clear signals of anomalies are detected for all four discharge time series: Andernach (ANDE), Cochem(COCH), Maxau(MAXA) and Plochingen(PLOC)).

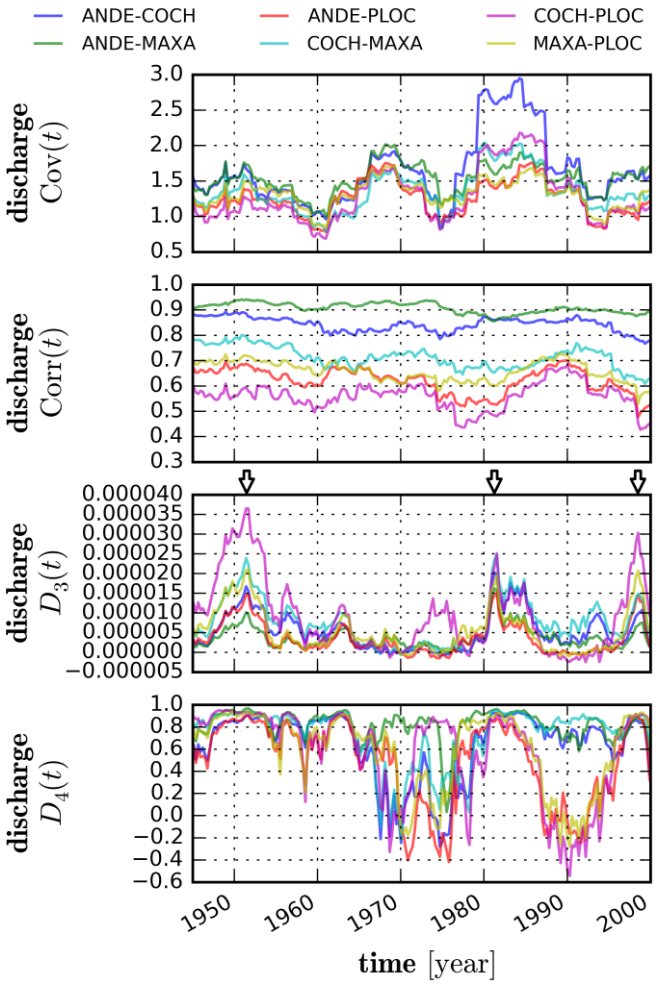
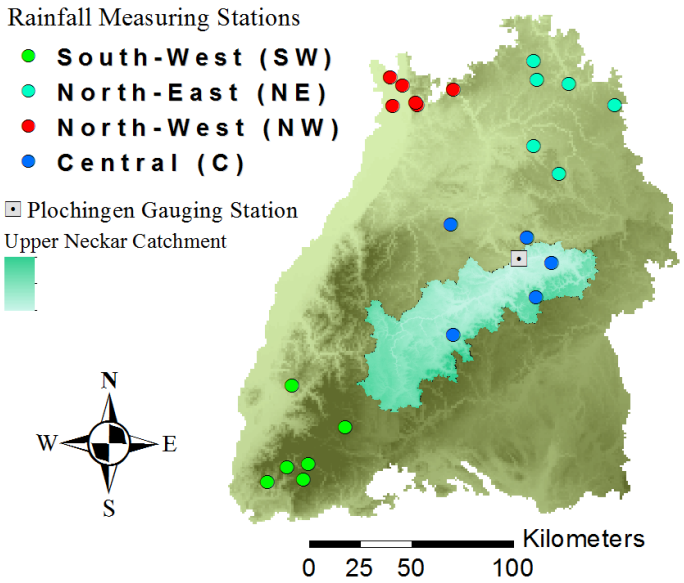



**Fig. 11.** Copula distances of discharge time series in moving time window: covariance (top), correlation (second), copula distance type3 (third) and copula distance type4 (bottom). The arrows point 1947, 1982 and 2000 in which the clear signals of anomalies are detected for the comparisons between 4 discharge time series: Andernach(ANDE), Cochem(COCH), Maxau(MAXA), Plochingen(PLOC)

692




693

694 **Fig. 12.** Locations of the precipitation gauge stations within Baden-Württemberg (Germany) indicated by

695 coloured circles. Upper Neckar catchment is identified by the light green area and the location of the

696 gauging station is indicated by a square

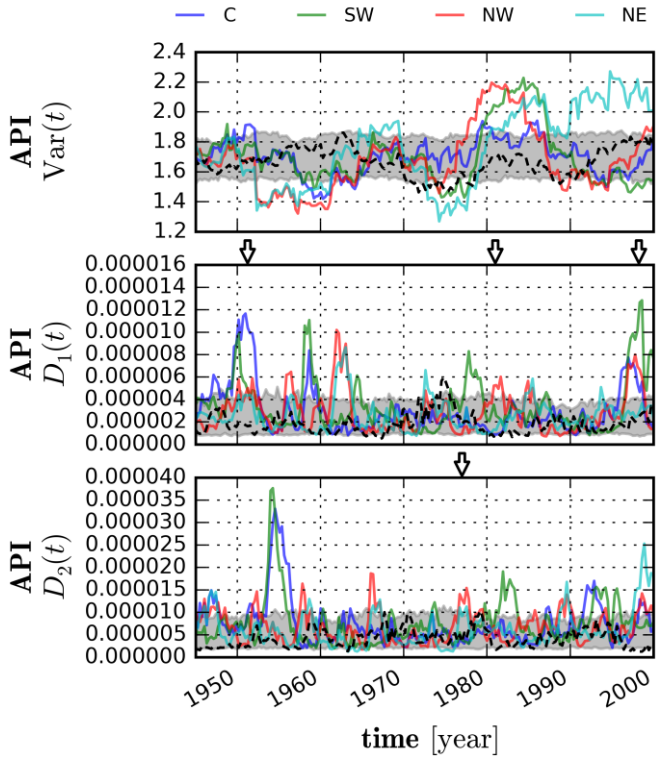



**Fig. 13.** Copula distances of API time series in moving time window: variance (top), copula distance type1 (middle) and copula distance type2 (bottom) where 'C' denotes central, 'SW' denotes southwest, 'NW' denotes northwest and 'NE' denotes northeast part of the Baden-Württemberg State of Germany, each containing 80% confidence interval of Gaussian process and one of its realization (dashed line). The arrows indicate the years in which anomalies are detected in the previous analysis (Fig. 10)

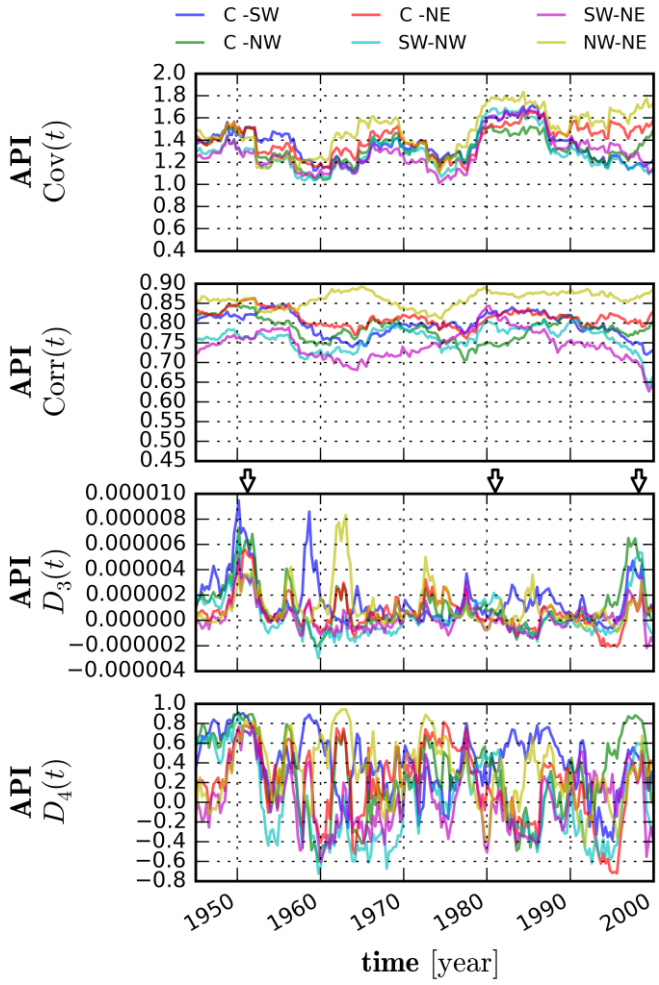



**Fig. 14.** Copula distances of API time series in moving time window: covariance (top), correlation (second), copula distance type3 (third) and copula distance type4 (bottom). The arrows indicate the years in which anomalies are detected in the previous analysis (Fig. 11)

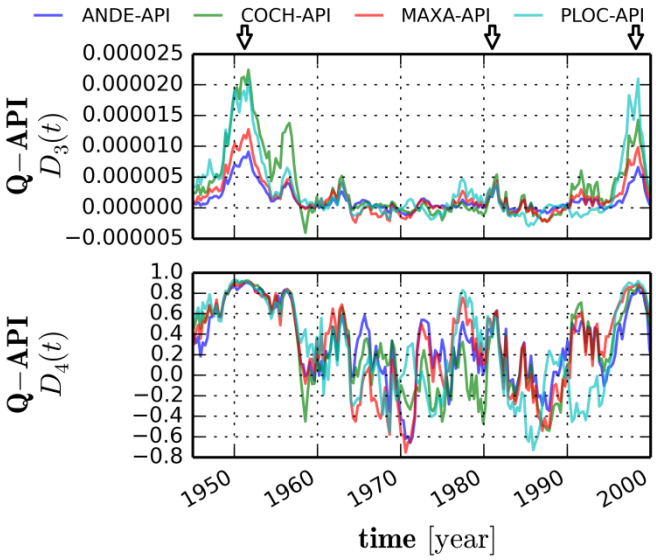



**Fig. 15.** Copula distance type3 (top) and type4 (bottom) between 4 discharge and 1 API time series which is aggregated for all the daily precipitations depicted in Fig. 12. The arrows indicate the years in which anomalies are detected in the previous analysis (Fig. 11)

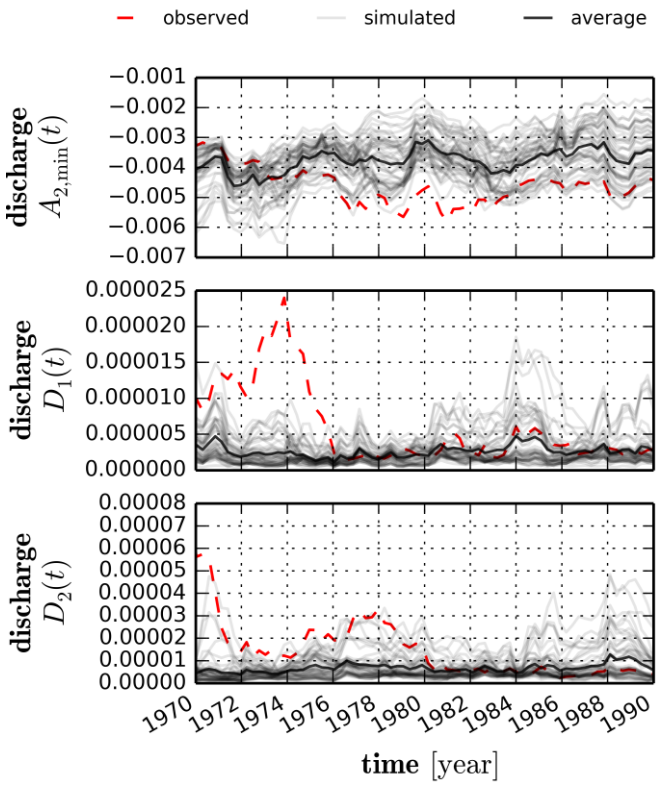



**Fig. 16.** Copula asymmetry and copula distances for 30 simulated and one observed discharge time series at Plochingen between 1965 and 2000: $A_{2,\min}$ for the time lag $k = 2$ days (top), copula distance type1 (middle), copula distance type2 (bottom)