Jürgen P. Kropp and Hans Joachim Schellnhuber (Eds.)

IN EXTREMIS
Extremes, Trends, and Correlations in Hydrology and Climate

SPIN; Version: January 30, 2017

Springer-Verlag
Berlin Heidelberg NewYork
Barcelona HongKong London
Milan Paris Singapore Tokyo

Contents

List of Contributors

Yosef Ashkenazy
Department of Solar Energy &
Environmental Physics,
Ben-Gurion University
Negev, PLZ, Israel
ashkena@bgu.ac.il

Katrin Bürger
Dostal UmweltConsulting
Environmental Consultant
Agency Freiburg
79102 Freiburg, Germany
kabu100@gmx.de

Armin Bunde
Institut für Theoretische Physik
Justus-Liebig-Universität Giessen
35392 Giessen, Germany
bunde@physik.uni-giessen.de

Peter Braun
Bayerisches Landesamt für
Umwelt
95030 Hof, Germany
braun.peter@lycos.de

Chandranath Chatterjee
Indian Institute of Technology
Agricultural and Food Engineering Department
Kharagpur-721302, West Bengal,
India
cchatnih@yahoo.com

Paul Dostal
Environmental Modelling Group
EMG
University Mainz
55099 Mainz, Germany
dostal@uni-mainz.de

Jan F. Eichner
Justus-Liebig-Universität Giessen
35392 Giessen, Germany
jan.f.eichner
@physik.uni-giessen.de

Shlomo Havlin
Minerva Center and Department
of Physics
Bar-Ilan University
Ramat-Gan 52900, Israel
havlin@ophir.ph.biu.ac.il

Florian Imbery
Meteorologisches Institut
Universität Freiburg
79085 Freiburg, Germany
florian.imbery
@meteo.uni-freiburg.de

Malaak Kallache
Integrierte Systemanalyse
Potsdam Institut für
Klimafolgenforschung
14412 Potsdam, Germany
malaak.kallache@pik-potsdam.de

Jan W. Kantelhardt
Fachbereich Physik und Zentrum
für Computational Nanoscience
Martin-Luther-Universität
06099 Halle, Germany
kantelhardt@physik.uni-halle.de

Eva Koscielny-Bunde
Institut für Theoretische Physik
Justus-Liebig-Universität Giessen
35392 Giessen, Germany
Eva.Koscielny-Bunde
@physik.uni-giessen.de

Jürgen P. Kropp
Integrierte Systemanalyse
Potsdam Institut für
Klimafolgenforschung
14412 Potsdam, Germany
kropp@pik-potsdam.de

Rakesh Kumar
National Institute of Hydrology
Surface Water Hydrology Division
Roorkee-247667, Uttaranchal,
India
rakesh@nih.ernet.in

Zbigniew W. Kundzewicz
Research Centre for Agricultural
and Forest Environment
Polish Academy of Sciences
60809 Poznań, Poland
and
Potsdam Institut für
Klimafolgenforschung
14412 Potsdam, Germany
Zkundze@man.poznan.pl
zbyszek@pik-potsdam.de

Jürgen Kurths
Institut für Physik
Universität Potsdam
14469 Potsdam, Germany
juergen@agnld.uni-potsdam.de

Holger Lange
Norsk institutt for skog og
landskap
1431 Ås, Norway
holger.lange@skogforsk.no

Gunnar Lischeid
Bayreuther Zentrum für Ökologie
und Ökosystemforschung
Universität Bayreuth
95440 Bayreuth, Germany
gunnar.lischeid
@bayceer.uni-bayreuth.de

Valerie N. Livina
School of Environmental Sciences,
University of East Anglia
Norwich, PLZ, United Kingdom
v.livina@uea.ac.uk

Miguel D. Mahecha
Biogeochemical Model-Data
Integration Group
Max-Planck Institut für Biogeo-
chemie
07701 Jena, Germany
miguel.mahecha@bgc-jena.mpg.de

Douglas Maraun
Institut für Physik
Universität Potsdam
14469 Potsdam, Germany
maraun@agnld.uni-potsdam.de

Manfred Mudelsee
Meteorologisches Institut
Universität Leipzig
04103 Leipzig, Germany
and
Climate Risk Analysis
06114 Halle, Germany
mudelsee
@climate-risk-analysis.com,

Jörg Neumann
Bayerisches Landesamt für
Umwelt
95030 Hof, Germany
joerg.neumann@lfu.bayern.de

Maciej Radziejewski
Faculty of Mathematics and
Computer Science
Adam Mickiewicz University
61614 Poznań, Poland,
and
Research Centre for Agricultural
and Forest Environment
Polish Academy of Sciences
60809 Poznań, Poland
maciejr@amu.edu.pl

Henning W. Rust
Institut für Physik
Universität Potsdam
14469 Potsdam, Germany
hrust@agnld.uni-potsdam.de
and
Potsdam Institut für
Klimafolgenforschung
14412 Potsdam, Germany
hrust@pik-potsdam.de

Diego Rybski
Institut für Theoretische Physik
Justus-Liebig-Universität Giessen
35392 Giessen, Germany
diego.rybski
@physik.uni-giessen.de

Bettina Schaefli
Institut für Geoökologie
Universität Potsdam
14476 Golm, Germany
schaefli@uni-potsdam.de

Hans-Joachim Schellnhuber
Potsdam Institut für
Klimafolgenforschung
14412 Potsdam, Germany
john@pik-potsdam.de

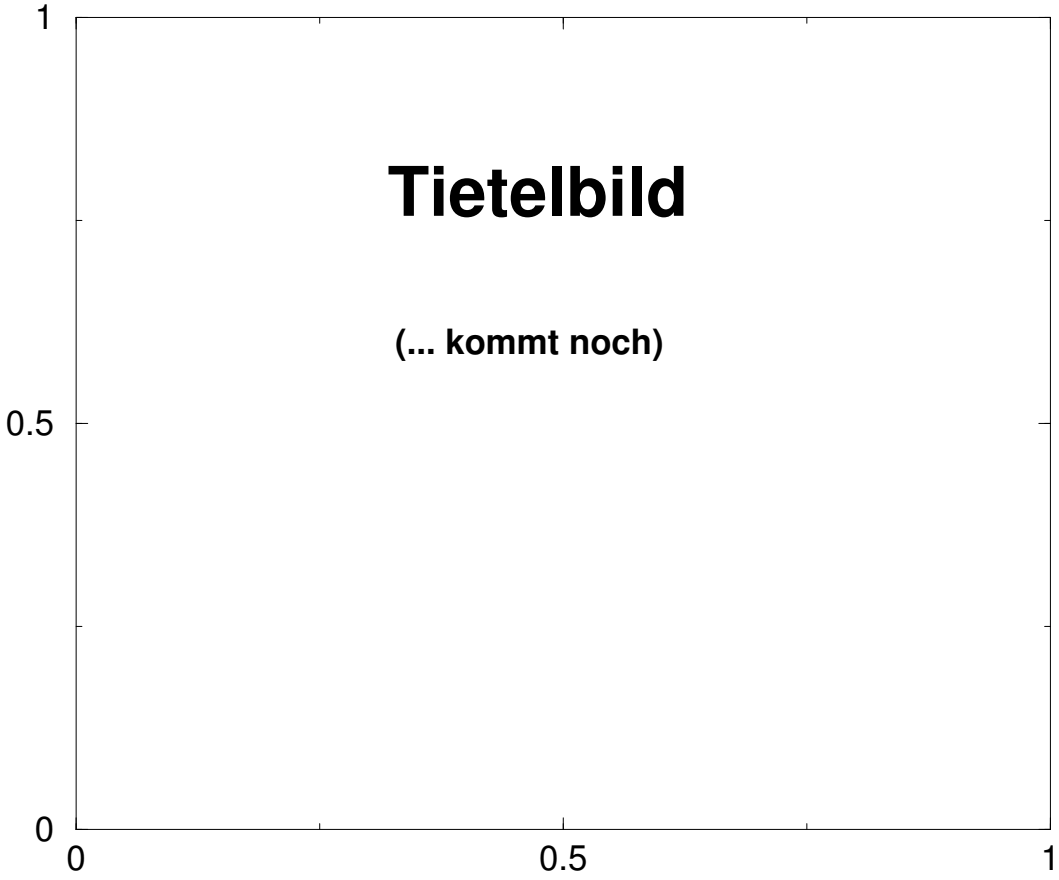
Christian-D. Schönwiese
Institut für Meteorologie
und Geophysik
Geothe Universität Frankfurt
60054 Frankfurt, Germany
schoenwiese
@meteor.uni-frankfurt.de

Jochen Seidel
Dept. of Hydrology and Geohy-
drology
University of Stuttgart
70569 Stuttgart, Germany
jochen.seidel
@iws.uni-stuttgart.de

Silke Trömel
Institut für Meteorologie
Universität Bonn
PLZ Bonn, Germany
silke.troemel@uni-bonn.de

Part I

Extremes and Trend Detection


Tietelbild
(... kommt noch)
1
0.5
0
0
0.5
1

1. Extreme Value and Trend Analysis based on Statistical Modelling of Precipitation Time Series

Silke Trömel and Christian-D. Schönwiese

Abstract Application of a generalized time series decomposition technique shows, that observed German monthly precipitation time series can be interpreted as a realization of a Gumbel distributed random variable with time dependent location parameter and time dependent scale parameter. The achieved complete analytical description of the series, that is the probability density function (PDF) for every time step of the observation period, allows probability assessments of extreme values for any threshold at any time. So, we found in the western part of Germany, that climate is getting more extreme in winter. Both the probability for exceeding the 95th percentile and the probability for falling under the 5th percentile are increasing. Contrary results are found in summer. The spread of the distribution is shrinking. But in the south, relatively high precipitation sums become more likely and relatively low precipitation sums become more unlikely in turn of the 20th century.

Furthermore, the decomposition technique provides the mean value of the Gumbel distributed random variable for every time step, too. So, an alternative approach for estimating trends in observational precipiation time series is achieved. On that way, the non-Gaussian characteristics can be taken into account and robust estimates can be provided. In constrast, application of the least squares estimator to non-Gaussian climate time series often leads to everestimated trends in the expected value.

◄ **Fig. 1.0.** Application of the generalized time series decomposition technique on modelled time series of the coupled global model ECHAM4/OPYC3 [1.12], 1990-2100, under SRES A2 scenario shows in January in the northern part of Europe strong increases in the probability for exceeding the 95th percentile (top) and less pronounced decreases in the probability for falling under the 5th percentile (bottom). In central Europe an area with increases in both kind of extremes can be seen.

1.1 Introduction

The analysis of climate variability as reflected in observational records is an important challenge in statistical climatology. In particular, it is important to estimate reliably trends in the mean value as well as changes in the probability of extreme values. In case of monthly or annual temperature data, often Gaussian distributions are adequate for their description. If any linear or non-linear trend of the time series average occurs, this may be described by a shift of a Gaussian probability density function (Pdf). Additionaly, some other time series components like the annual cycle for example may also vary and cause a shift to higher and lower values again during the observation period as discussed below [1.2]. Grieser et al. (2002) [1.2] consider temperature time series as a superposition of trends, annual cycle, episodic components, extreme events and noise. Thereby, Gaussian assumptions are used, which implies that residuals can not be distinguished from a realization of a Gaussian distributed random variable. However, this simple model of a shifting Gaussian distribution with constant variance over the observation period is not suitable to describe the variability of precipitation time series. We observe a skewed distribution as well as seasonal and sometimes long-term changes in the shape of the distribution. Moreover, changes in the spread of the distribution have to be considered, too. In summary, in addition to the expectation, also the variance and/or further moments may vary with time.

Evidently, wrong assumptions concerning the Pdf lead also to biased estimators of trends or probabilities of extreme values. In consequence, we present a generalized time series decomposition technique which allows any Pdf and any related parameter change in time. That is changes in the location, the scale and the shape parameter of a Pdf are allowed. In case of Gaussian distributions for example, the location parameter is realized by the average and the scale parameter is realized by the variance.

Here, we apply this technique on precipitation time series from a German station network focussed on extreme value and trend analysis. The detection of structured time series components like trends etc., called signals, is based on several parameters of a Gumbel or Weibull distribution, respectively, as described in the following. Section 1.2 provides the defintion of the components for the analysis of monthly climate time series and Sect. 1.3 presents a brief overview of the detection of the analytical functions reflecting the time dependence of the different distribution parameters. For all details see [1.10] and [1.11].

1.2 Components

The equation

$$S_{j,k}(t) = d_{j,k}t^k \cos\left(2\pi\frac{j}{12}t\right) + e_{j,k}t^k \sin\left(2\pi\frac{j}{12}t\right) \tag{1.1}$$

with wave number $j = 1, \ldots 6$ per year and $k = 0, 1, 2$ gives the basis functions to describe the seasonal component. Evidently, the maximum wave number 6 is tuned to the analysis of climate data with a monthly sampling rate. Besides fixed annual cycles, changes in amplitude and phase are allowed. For the amplitude linear and quadratic time dependence is considered. Superposition of two or three harmonics of the annual cycle with the same wave number j but different time dependence k in one time series yields

$$S_j(t) = A_j(t)\cos\left(2\pi\frac{j}{12}\left(t - t_j\right)\right) \tag{1.2}$$

with amplitude

$$A_j(t) = \sqrt{\sum_{k=0}^{2}\left(d_{j,k}^2 + e_{j,k}^2\right)t^{2k}} \tag{1.3}$$

and phase

$$t_j(t) = \frac{12}{2\pi} arctan\left(\frac{\sum_{k=0}^{2} d_{j,k}t^k}{\sum_{k=0}^{2} e_{j,k}t^k}\right). \tag{1.4}$$

In this way, the detection of linear, progressive and degressive shaped changes in phase and amplitude of the annual cycle is possible.

In addition, trends up to the order 5 are considered:

$$T_i(t) = g_i + h_i t^i \quad \text{with } i = 1, \ldots, 5. \tag{1.5}$$

To detect all significant structures but neglecting, on the other hand, all unsignificant structures, in a first step the detection of the seasonal component and the trend component of the parameters is performed simultanously within the modified stepwise regression precedure (see Sect. 1.3 and [1.9]).

In a second step we observe sometimes relatively low-frequency variations superposed on the components mentioned above. So we offer also polynomial equations up to the order 5:

$$V_l(t) = c_o + \sum_{i=1}^{l} c_i t^i. \tag{1.6}$$

In a third step a search for extreme events, which are independent from changes in the parameters of the distribution, is performed. According to Grieser et al. (2002) [1.2], we define extreme events as a relatively small number of extreme values which are unexpected within the scope of the fitted statistical model. In contrast to extreme events, extreme values are relatively high or low values which occur by chance. A positive trend in the mean value, for example, increases the probability of occurence of relatively high extreme values.

Within the strategy introduced detected extreme events are extracted and replaced by a random value distributed conform with the PDF and the two parameters at the given time (see Grieser et al., 2002 [1.2], for further details). The iterative procedure for the detection of trends, seasonal component and low-frequency variations is applied until no further extreme events are found. In the application of Gaussian assumptions and the least-squares method, the quadratic function which has to be minimized in order to fit any analytical functions is called distance function of the Gaussian distribution. Generally speaking, it depends on the robustness ([1.3]) of the applied distance function whether structured components are more or less influenced by extreme events.

Using an iterative procedure to find within these functions the best model equations of two parameters of a PDF leads to a very extensive procedure. Note that the model equation of one parameter influences the equation for the second parameter and vice versa. To reduce this effort, an additional restriction is introduced. For one of the parameters all functions mentioned above are offered:

$$P_g(t) = \sum S_{j,k}(t) + \sum T_i(t) + V_l(t). \tag{1.7}$$

The second parameter is assumed to be of minor relevance to describe the time series. Only the one cycle per year harmonic and one trend function can be chosen:

$$P_r(t) = S_{1,0} + T_i. \tag{1.8}$$

At the end of the time series decomposition procedure, a priori assumed residual distribution is tested. In case of a chosen Gumbel distribution, residuals should follow a Gumbel distribution $G(0, 1)$ with a location parameter equals 0 and a location parameter equals 1 after elimination of detected structured components in the parameters $P_g(t)$ and $P_r(t)$. Additional stationarity of the distribution points to a complete description of the time series within the PDF and its time-dependent parameters.

It depends on the characteristics of the time series under consideration, which PDF has to be chosen and which PDF parameter implies a larger number of degrees of freedom in comparison to others. Not before the end of the precedure, including the analysis of the statistical properties of the residuals, it can be decided whether the chosen model provides an adequate description of the time series.

1.3 The distance function and the model selection criterion

The basis of any time series decomposition technique are the distance function and a model selection criterion. The least-squares estimator broadly used in

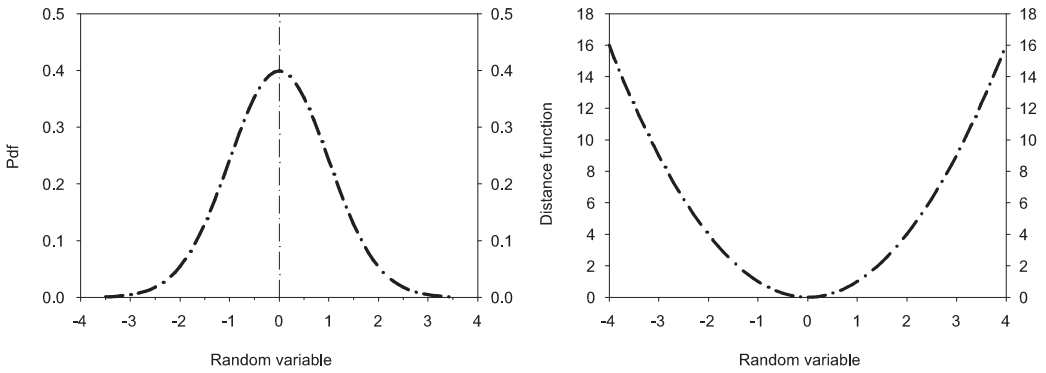


Fig. 1.1. The PDF (left) of the Gaussian distribution and its distance function (right) under the assumption of constant variance.

trend analyses is the maximum likelihood estimator under the assumption of Gaussian distributed residuals with constant variance. This quadratic function which has to be minimized is called distance function of the Gaussian distribution. So, consistently with the maximum likelihood principle another distance function, defined as the negative logarithm of the PDF, replaces the function of squared errors to be minimized, if we choose another distribution as basis of the decomposition procedure, for example the Gumbel or the Weibull distribution. Then time dependence for different distribution parameters can be allowed. With the exchange of the distance function, fitted basis functions describe changes in the location, scale or shape parameter of an appropriate PDF. As an example, the corresponding distance function of the Gumbel distribution is

$$\rho(x,t) = ln(b(t)) + exp\left(-\frac{x-a(t)}{b(t)}\right) + \frac{x-a(t)}{b(t)}. \tag{1.9}$$

Under the assumption of statistically independent random variables the coefficients in the model equations for estimating $a(t)$ and $b(t)$ are chosen by minimizing

$$\sum_t \rho\,(x,t) = min., \tag{1.10}$$

equivalent to the maximum of the loglikelihood function. In this work Powells method [1.6], p. 406, is used to minimize ρ in the multidimensional space.

Figure 1.1 shows the PDF of the Gaussian distribution and on the right hand-

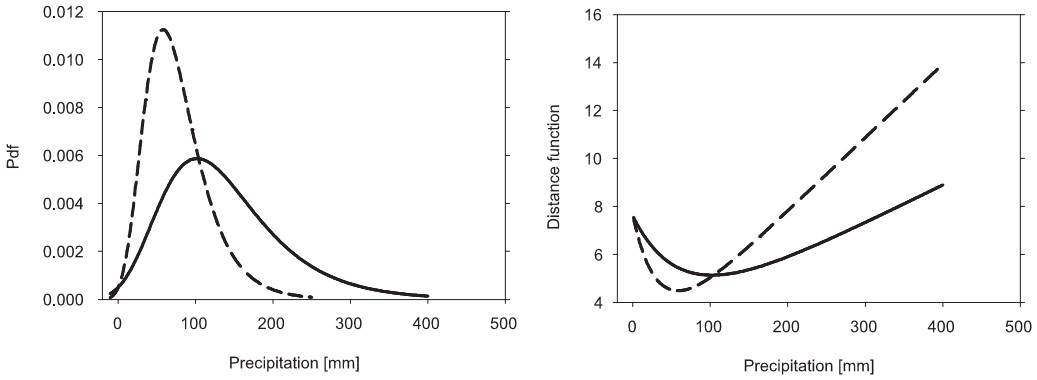


Fig. 1.2. The PDF (left) of the Gumbel distribution with two different location and scale parameters and the associated distance functions (right).

side its distance function, the quadratic function. The more deviant the points, the greater the weight. The influence increases very fast because the probability of occurrence of relatively high or relatively low values is very small in the Gaussian case. For comparison the Gumbel distribution for two different location and scale parameters and the associated distance functions are shown in Fig. 1.2. The tails are more prominent and consequently, if we take a look at the distance functions, the influence increases less rapidly. And one value in a given distance from the location estimator has more weight the smaller the scale parameter. So, structured components can be detected in different parameters and estimators of different parameters compete with each other. Finally, Fig. 1.3 shows the Weibull distribution with two different scale and shape parameters. And on the right-hand side the distance functions clearly show the dependence on the shape of the distribution. One important point is that the PDF chosen and the basis functions used to describe the signals are complementary. The other point is, that the basis used to describe one parameter of the PDF influences the basis necessary to describe the second parameter. That is why a dynamic procedure is necessary to estimate the coefficients of the basis functions of the different components of both parameters simultaneously. A flexible model selection criterion often used is the stepwise regression. Stepwise regression [1.9], p. 166, represents a dynamic model selection criterion in order to find the optimal regression equation. Within the generalized time series decomposition a modification to handle two distribution parameters is required.

The common iterative application of forward selection and backward elimination is used to determine the model equation of a first distribution parameter as, for example, the scale parameter $b(t)$. But the distance function used now,

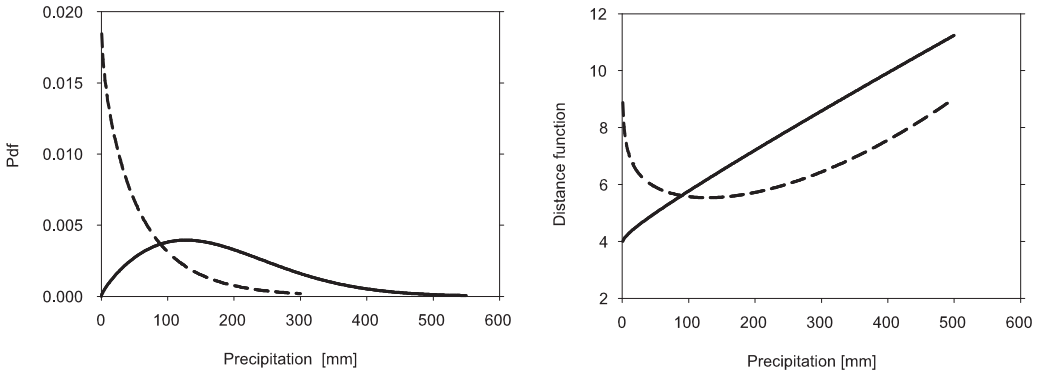


Fig. 1.3. The PDF (left) of the Weibull distribution with two different scale and shape parameters and the associated distance functions (right).

also depends on the selected model equation describing the second distribution parameter (see Eq. 1.9). Between these alternate parts redefinition of the second parameter ($a(t)$) is inserted, now taking into account the selected model of the first parameter ($b(t)$). In the modified version, the flexible strategy of stepwise regression is used twice. Now the parameters of the model equations influence themselves mutually. A regressor of the first parameter selected at an earlier stage can be superfluous because of a new entry candidate in the model equation of the first parameter or because of the actualisation of the second parameter and vice versa.

The common F-test statistic broadly used in regression analysis to decide whether a specific regressor contributes significantly to explained variance is sensitive to departures from the Gaussian distribution and, therefore, has to be replaced. A test statistic based on a likelihood ratio test seems to be more applicable, because likelihood values are computed anyhow minimizing the distance function (Eq. 1.10). Define $D(p)$ as the minimum value of (1.10) subject to the model containing p regressors and $D(q)$ as the minimum value of (1.10) subject to the model containing q regressors, with $p > q$, a direct generalization of the common F-test may be based upon the statistic

$$F_M = \left(D(p) - D(q)\right) / (p - q) \tag{1.11}$$

(Schrader and Hettmansperger, 1980) [1.8] with $p - q$ and $N - p$ degrees of freedom. Here N represents the number of data. If the density function of the residuals has not the form exp(-ρ), the model assumptions are not fulfilled and a correction term is necessary. For small sample sizes Schrader and Hettmansperger (1980) propose to compare F_M with a critical value from a central F distribution.

The modified stepwise regression including a modified F-statistic (Eq. 1.11) represents the basis of a generalized time series decomposition technique. In-

serting the corresponding distance function it can be used for time series analysis based on any data distribution and any statistical model. Thereby, the initial choice of the probability density function and the corresponding distance function should depend on the characteristics of the climate variable under consideration. The Gumbel distribution, the Lognormal distribution as well as the Gamma distribution are promising distributions already by reason of the skewness. However, the Weibull distribution seems to be a promising choice when changes in the shape of the distribution occur. Not until the final residual analysis at the end of the decomposition tests whether a priori assumed statistical model can be confirmed. If so, a complete analytical description of the time series is achieved.

Actually, the generalized time series decomposition technique in a deterministic and a statistical part is applied to 4 different models: Observed time series can be interpreted as

- a realization of a Gaussian-distributed random variable with time-dependent location (mean value) and additional time-dependent scale parameter (standard deviation),
- a realization of a Gumbel-distributed random variable with time-dependent location and scale parameter, it depends on the choice to offer the location or the scale parameter the greater pool of regressors if we talk about
  – the Gumbel model with special emphasize on location or
  – the Gumbel model with special emphasize on scale.
- a realization of a Weibull-distributed random variable with time-dependent scale and shape parameter.

1.4 Application to a German station network

1.4.1 General remarks

The generalized time series decomposition technique is applied now to monthly precipitation sums from a German station network of 132 time series covering 1901-2000. At least for the most part of the sample the decomposition technique shows, that observed time series can be interpreted as a realization of a Gumbel distributed random variable with time-dependent location parameter $a(t)$ and time-dependent scale parameter $b(t)$. So the decomposition is based on the probability density function (PDF) of the Gumbel distribution

$$f(x,t) = \frac{1}{b(t)} \left\{ \exp\left(-\frac{x-a(t)}{b(t)} \right) \exp\left[-e^{-(x-a(t))/b(t)} \right] \right\}. \tag{1.12}$$

Since the Gumbel model with emphasis on scale leads to a better description of the time series, we define the location parameter $a(t) = P_r(t)$ and the scale

parameter $b(t) = P_g(t)$ (see Eqs. 1.7 and 1.8). However, 7 out of 158400 (132·1200) monthly precipitation sums are extracted as extreme events which are unexpected within the scope of the fitted Gumbel model. As aforementioned the analysis of the remaining residuals represent an important part of the anaysis procedure. After elimination of the detected structures in the location and in the scale parameter, residuals should fulfill the condition of the a priori assumed statistical model. Concerning the German station network of 132 monthly precipitation sums, the residuals should be undistinguishable from the realization of a Gumbel distributed random variable G(0,1) with the location parameter 0 and the scale parameter 1. In fact the Kolmogorov-Smirnov test [1.6] rejects in 7 out of 132 cases this hypothesis with a probability larger than 90%. This is less (<10%) than may be expected by chance. Furthermore, again a Kolmorgorov-Smirnov test statistic is used in order to check the stationarity of the residuals. Now in 6 out of 132 cases stationarity is rejected with a probability larger than 90%. Consequently, the residuals confirm the model applied and a complete description of the observed time series on the basis of the Gumbel model with emphasis on the scale could be achieved.

The provided PDF for every time step of the observation period allows probability assessments of extreme values, i.e. the probability for exceeding a given precipitation sum (threshold) at any time. In this context it should be reemphasized that in contrast to analyses dealing with the familiar Generalized Extreme Value distribution [1.4] [1.1], the generalized time series decomposition technique presented in this issue requests the analytical description of the whole times series instead of adressing exclusively the maxima of the observations. Furthermore, with the PDF the mean value of the Gumbel distributed random variable can be given for every time step, too.

In Section 1.4.2 the time series decomposition of an examplary time series is discussed in more detail for a better understanding of the method introduced. Subsequently, results concerning changes in the probability of extreme values and the expected value of the entire station network are presented in Sect. 1.4.3 and Sect. 1.4.4.

1.4.2 Example: Eisenbach-Bubenbach

In case of the Gumbel model with emphasis on scale the greater pool of regressors is offered to the scale parameter $b(t)$ of the Gumbel distribution and the smaller one to the location parameter $a(t)$. Table 1.1 shows the significant functions detected in the location parameter $a(t)$ and the scale parameter $b(t)$ of the Gumbel distribution to describe the time series observed in Eisenbach-Bubenbach (47.97°N, 8.3°E). Phase angles are defined with reference to December 15th. Consequently, we observe a superposition of two harmonics with wave number one per year. The function $S_{1,0}$ has a constant amplitude with maximum in September but $S_{1,1}$ reveals linear time dependence in the ampli-

Table 1.1. Significant functions in the location parameter $a(t)$ and scale parameter $b(t)$ for the describtion of the monthly precipitation time series observed in Eisenbach-Bubenbach.

Parameter	Function	Amplitude [mm]	Phase [days]
$b(t)$	$S_{1,0}$	$7.04 \cdot \mathrm{t}^{\mathrm{k}}$	-82.73
	$S_{1,1}$	$0.148 \cdot 10^{-1} \cdot \mathrm{t}^{\mathrm{k}}$	44.87
	T_1	5.78	-
$a(t)$	$S_{1,0}$	$-8.89 \cdot \mathrm{t}^{\mathrm{k}}$	24.21

tude and has its maximum in January. So, the winter becomes more variable and the summer becomes less variable during the observation period. The seasonal component in the scale parameters shows a phase shift of 109 days from September to January. The resulting effects can also seen in Figs. 1.4 and 1.5. The former figure illustrates the decrease in the scale parameter in August. The functions given in Tab. 1.1 define the scale and the location parameter of the Gumbel distribution for every time step. No low-frequency variations ($V_l(t)$)

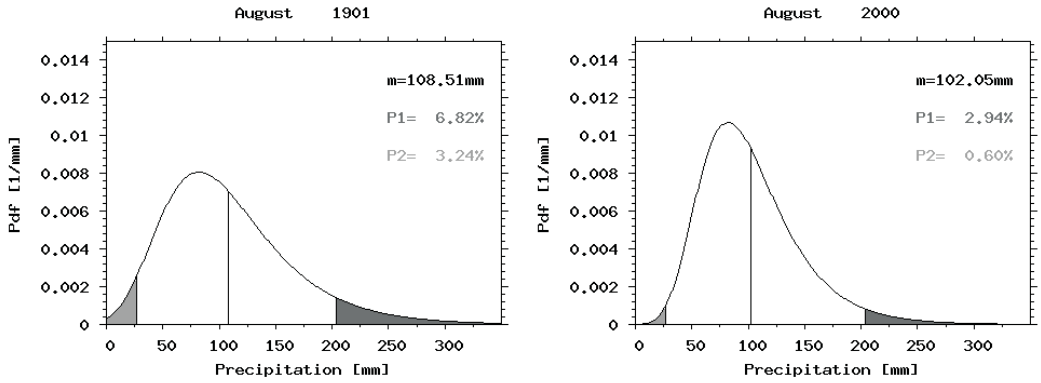


Fig. 1.4. On the basis of the entire time series estimated PDFs for two time steps: August 1901 (left) and August 2000 (right). Areas below the 5th percentile are marked in light grey, the areas above the 95th percentile in dark grey and the vertical line marks the expected value m. The respective value m, the probability P1 for exceeding the 95th percentile and the probability P2 for falling under the 5th percentile are given in the upper right corner respectively.

and no unexpected values defined as extreme events (see Sect. 1.2) within the time series decomposition technique can be detected in the time series. However, probability assessments of extreme values can be based on the analytical description provided. Figure 1.4 shows the estimated PDFs for the first and the last August in the observation period. The probability P1 for exceeding the 95th percentile, in this case 204 mm, and the probability P2 for falling under the 5th percentile, in the present case 27 mm, are also given for the two time steps. The corresponding areas within the PDF are marked in dark grey and light grey, respectively. First of all we see is a decrease in the scale parameter,

458 going along with a decrease in the probability of exceeding the 95th percentile
459 from 6.8% to 2.9% and we see a smaller decrease in the probability of falling
460 under the 5th percentile from 3.2 to 0.6%. The probability of both kinds of
461 extremes has decreased and in contrast to the Gaussian model the expected
462 value is also affected. The expected value decreases from 109 mm to 102 mm in
463 the observation period. It is worth to mention that the least squares estimator
464 is not able to describe those decreases or increases in variability and not the
influence on the expected value either. In Fig. 1.5 integration over the dark

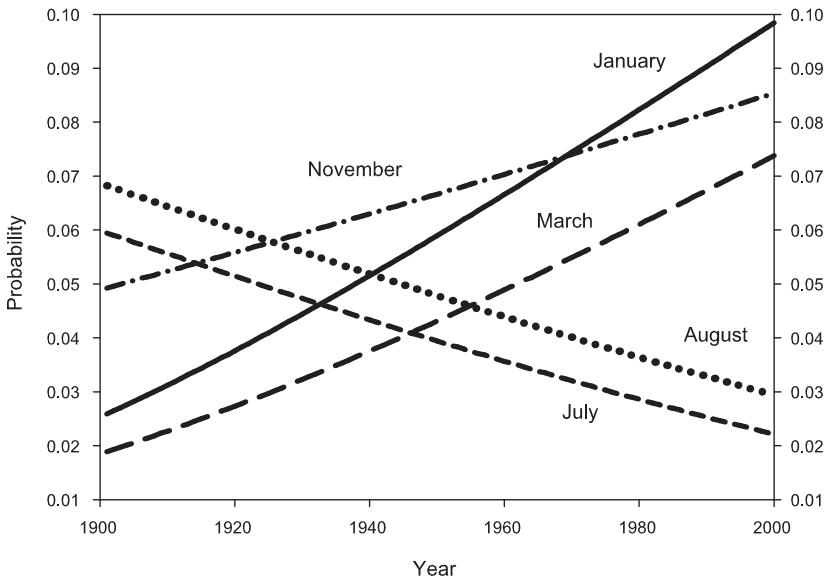


Fig. 1.5. Estimated probabilities for exceeding the 95th percentile between 1901 and 2000 in January, March, July, August and November at the station Eisenbach-Bubenbach.

465
466 grey area for every time step and different months has been done. Contrary
467 tendecies in the probability of exceeding the 95th percentile can be seen. At
468 the beginning the relatively high precipitation sums occured with higher prob-
469 ability in summer but at the end of the observation period the probability
470 for exceeding the threshold is highest in January. Both, the phase shift in the
471 seasonal component and the positive linear trend detected in the scale param-
472 eter describe an increasing variability in winter. The distribution is widening.
473 The tendencies in summer are contrary. Both kind of extremes become more
474 unlikely between 1901 and 2000.

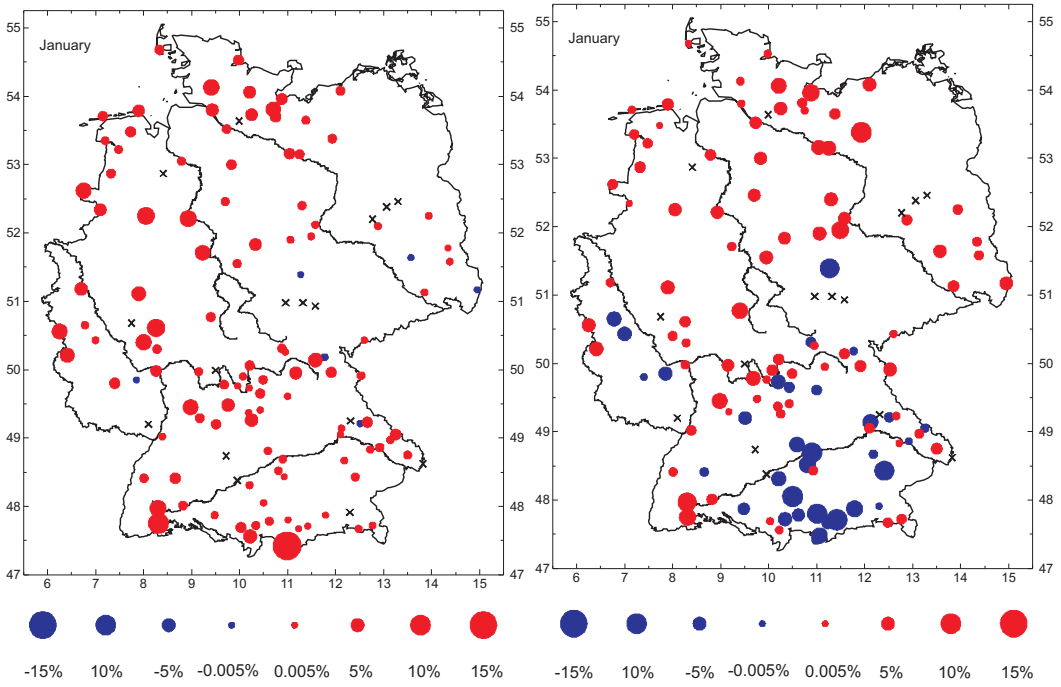


Fig. 1.6. Changes in the probability of occurence of a monthly precipitation sum greater than the 95th percentile (left) and in the probabiliy of occurence of a monthly precipitation sum smaller than the 5th percentile (right). Results are given for January in the observation period 1901-2000. Red dots indicate an increase and blue dots indicate an decrease in the probability of occurrence. The size of the dots is proportional to the magnitude in probability change.

1.4.3 Probability assessment of extreme values

If the time series decomposition in a statistisal and a deterministic part succeeded, the time-dependent PDF f(x, P_g(t), P_r(t)) as a complete analytical description is provided (see again Fig. 1.4).

In this model

$$p_G(x \geq x_s, t) = 1 - \int_{-\infty}^{x_s} \frac{1}{P_g(t)} \left\{ \exp\left(-\frac{x - P_r(t)}{P_g(t)} \right) \exp\left[-e^{-(x-P_r(t))/P_g(t)} \right] \right\} dx$$

$$= 1 - \exp\left\{ -\exp\left(-\frac{x_s - P_r(t)}{P_g(t)} \right) \right\} \tag{1.13}$$

gives the probabiliy for exceeding the threshold x_s at time t. Significant structured components in the location and the scale parameter cause changes in the probability of occurence of these high precipitation sums. The probability for falling under x_s is given by

$$p_G(x \leq x_s, t) = \int_{-\infty}^{x_s} \frac{1}{P_g(t)} \left\{ \exp\left(-\frac{x - P_r(t)}{P_g(t)} \right) \exp\left[-e^{-(x-P_r(t))/P_g(t)} \right] \right\} dx$$

$$= \exp\left\{ -\exp\left(-\frac{x_s - P_r(t)}{P_g(t)} \right) \right\}. \quad (1.14)$$

In Fig. 1.6 tendencies in the probability of a monthly precipitation total greater than the 95% percentile (left map) and tendencies in the probability of a monthly precipitation total smaller than the 5% percentile (right map) are shown. That is for every single time series the threshold is selected so that 95% (5%) of all monthly rainfall totals of the series are smaller than the respective threshold. The result is in January an increase in the probability of exceedance in the overwhelming majority of stations. In the right map the tendencies in the probability of a monthly precipitation total less than the

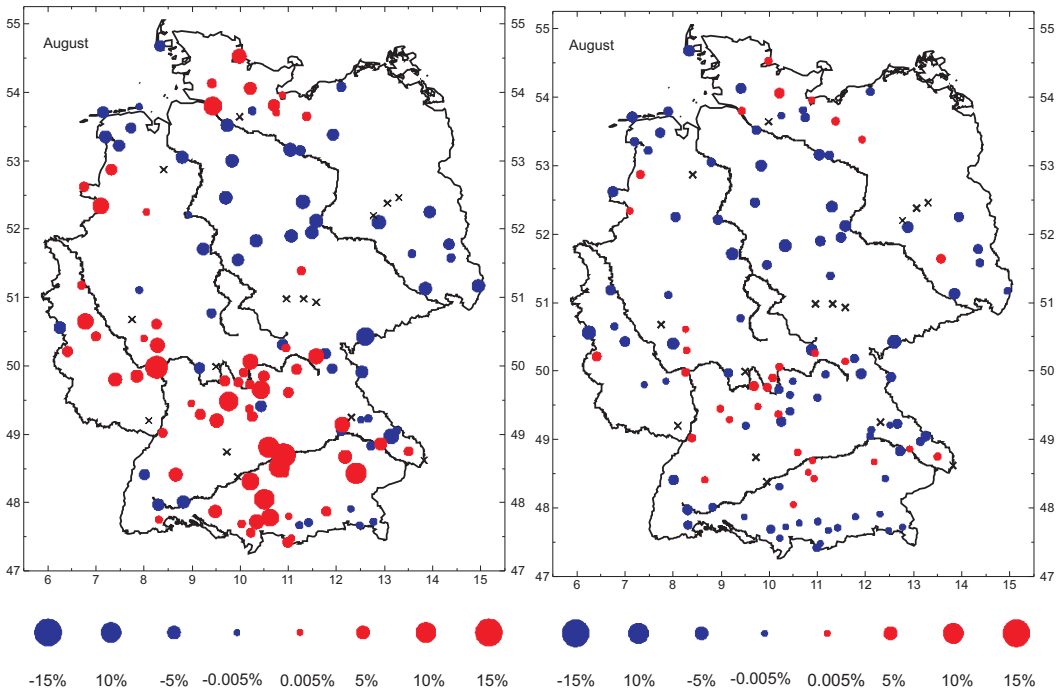


Fig. 1.7. Analogous maps to Fig. 1.6 with results for August.

5th percentile are seen. In January we see in the northern part of Germany increases in the probability of occurrence for these small precipitation totals, too. So relatively high and relatively small precipitation totals become more likely during the 20th century. The distribution is widening. However, in the south the distribution is shifting to higher values.

Figure 1.7 shows the respective results in August. In the northern part of Germany decreases in the probabiliy of relatively high precipitation sums are detected during the 20th century. But in the south we see several increases

in August as well as in January. In the right map it can be seen, that decreases in the probability of occurrence of relatively small precipitation totals are detected in the overwhelming majority of stations. Combined with the tendencies in the probability of exceedance of the 95th percentile we come to the conclusion that in summer in the northern part of Germany the extreme high and extreme low precipitation totals are getting more unlikely. In the south we observe again at several stations a shift of the distribution to higher values but results are not as uniform in that region in summer.

1.4.4 Changes in the expected value

As could already be seen in Fig. 1.4 the statistical modelling provides the expected value for every time step of the observation period. Consequently, the method introduced represents an alternative approach to estimate trends. The difference in the expected value between the January 2000 and the January in the year 1901 is defined as the trend in the expected value at every station in that month. With the use of an appropriate distance function relatively high precipitation sums do not get more weight than can be justified from a statistical point of view. Additionally, changes in different parameters of the distribution can be taken into account, because the expected value of a Gumbel distributed random variable is defined as

$$\mu(t) = a_G(t) + b_G(t)\gamma \approx a_G(t) + 0.57722 \cdot b_G(t) \tag{1.15}$$

with the location parameter $a_G(t)$, the scale parameter $b_G(t)$, and Eulers constant γ. Because of the skewness of the distribution a change in the scale parameter causes changes in the expected value, too. And the expected value of a Weibull distributed random variable depends on the location parameter a_W, the scale parameter b_W and the shape parameter c_W:

$$\mu(t) = a_W(t) + b_W(t)\Gamma\left(1 + \frac{1}{c_W}\right). \tag{1.16}$$

The Gamma-function Γ is definied as

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt. \tag{1.17}$$

A familiar nonparametric trend test would be the Mann-Kendall trend test [1.5]. Even though this test contains information about a increase or decrease in the time series considered, no information about the temporal evolution nor the amplitude of the trend is provided. In most cases, changes in the expected values of non-Gaussian climate time series, e.g. ordinary trend maps are estimated using the least squares method, too. Comparison of the trend

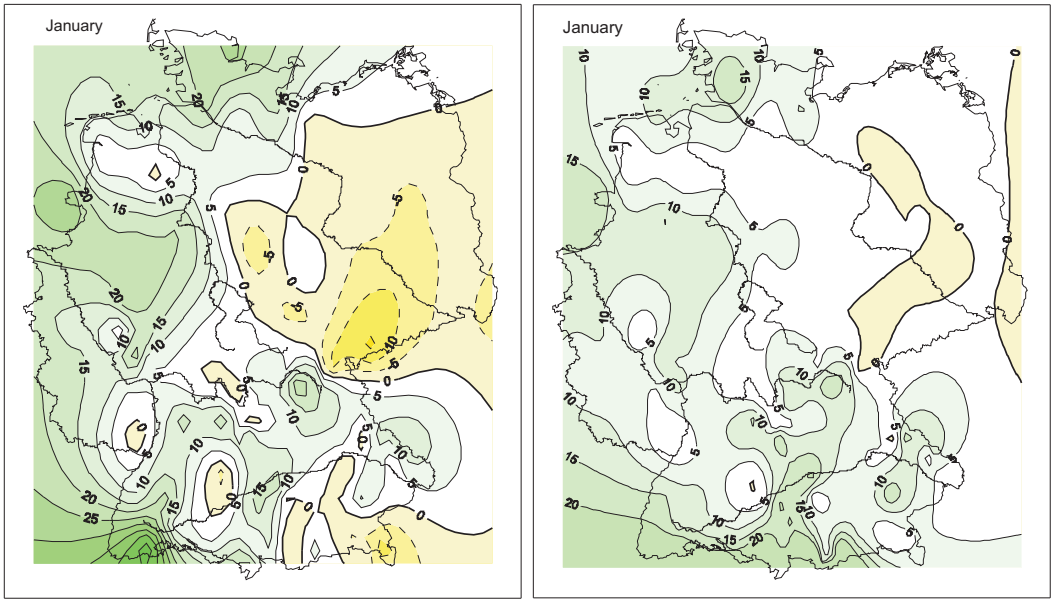


Fig. 1.8. Observed trends 1901-2000 in January estimated with the least-squares method (left) and on the basis of the statistical modelling of the time series using the distance function of the Gumbel distribution (right).

map on the basis of the least-squares-method with the trend map estimated on the basis of the method introduced (Fig. 1.8) shows in January similar spatial structures. We observe increases in the western and southern part of Germany, but decreases in the eastern part. However, the amplitudes are smaller using the distance function of the Gumbel distribution.

Another advantage of the method introduced is that trends of a particular month are estimated on the basis of the whole sample size. In order to demonstrate the large effect of the sample size on trend analysis results, Monte-Carlo simulations have been performed. 100 time series with a length of 100 or 1 000 time steps, respectively, of Gumbel-distributed variables with trends $\Delta\mu$ have been generated. Subsequently, we applied the least-squares estimator and tried to find the prescribed trend. In Tab. 1.2 the mean least-squares-estimator $\overline{\Delta\mu}_{KQ}$, its standard deviation $\sigma_{\Delta\mu}$, as well as the smallest and the greatest estimator, $\Delta\mu^{-}_{KQ}$ and $\Delta\mu^{+}_{KQ}$ are shown for eight different experiments using the longer generated time series (1 000 time steps). A positive bias can be observed. Table 1.3 shows the analogous results for the smaller sample size (100 time steps). The absolute value of the bias has clearly increased but the trend estimator shows a negative bias now. The estimators variance $\sigma_{\Delta\mu}$ has increased remarkably. It may be higher than the trend magnitude. Consequently, the greater sample size of precipitation totals taken into account within the statistical modelling approach (1 200 values instead of 100 in the present case) implies a smaller mean squared error of the maximum likelihood estimator given by the sum of the quadratic bias and the variance of the estimator.

Table 1.2. Trends of 100 generated Gumbel-distributed time series with 1 000 time steps and different linear changes in the location parameter (Δa) and constant or additional changing scale parameter (b). The resulting changes in the expected value $\Delta\mu$ (see again Eq. 1.15) are compared to mean least-squares estimator of the 100 time series $\overline{\Delta\mu}_{KQ}$, the standard deviation $\sigma_{\Delta\mu}$ of the estimator as well as the greatest and the smallest trend estimator ($\Delta\mu^{+}_{KQ}$ bzw. $\Delta\mu^{+}_{KQ}$).

	Δa=15 b=50	Δa=15 b=20	Δa=8 b=40	Δa=0 b=40	Δa=8 b=40+Δ8	Δa=-15 b=50	Δa=-8 b=40-Δ8	Δa=0 b=40+Δ10
$\Delta\mu$	15	15	8	0	12.6	-15	-12.6	5.77
$\overline{\Delta\mu}_{KQ}$	15.49	15.20	8.39	0.39	13.05	-14.51	-12.27	6.22
$\sigma_{\Delta\mu}$	6.97	2.79	5.57	5.57	6.16	6.70	5.02	6.31
$\Delta\mu^{-}_{KQ}$	0.78	9.31	-3.38	-11.38	0.08	-29.22	-22.83	-7.06
$\Delta\mu^{+}_{KQ}$	33.44	22.38	22.75	14.75	28.23	3.44	1.27	21.81

Table 1.3. Analogous to Tab. 1.2, but considering Gumbel-distributed time-series with 100 time steps.

	Δa=15 b=50	Δa=15 b=20	Δa=8 b=40	Δa=0 b=40	Δa=8 b=40+Δ8	Δa=-15 b=50	Δa=-8 b=40-Δ8	Δa=0 b=40+Δ10
$\Delta\mu$	15	15	8	0	12.6	-15	-12.6	5.77
$\overline{\Delta\mu}_{KQ}$	12.85	14.14	6.28	-1.72	10.78	-17.15	-14.21	3.90
$\sigma_{\Delta\mu}$	23.82	9.53	19.06	19.06	20.91	23.82	17.31	21.39
$\Delta\mu^{-}_{KQ}$	-43.84	-8.54	-39.07	-47.07	-38.96	-73.84	-55.18	-49.93
$\Delta\mu^{+}_{KQ}$	59.52	32.81	43.62	35.62	51.44	29.52	22.13	46.37

1.5 Conclusions

A generalized consistent decomposition procedure of precipitation time series into a statistical and a deterministic part is introduced. The basis functions allowed to describe the deterministic components only contain trends, annual cycle, episodic component and extreme events in order to restrict to physically explainable functions. Under the additional assumption that climate change is not restricted to the mean value the signal detection technique is applied to two instead of one parameter of a PDF, which can be chosen without any further restriction.

In particular, we show that a time series decomposition technique based on a Gumbel distribution, with flexible location and scale parameter, succeeds to describe monthly precipitation total time series from German stations completely. The model provides a full analytical description of the time series. On this basis probabilities of exceeding defined thresholds can be estimated reliably for every time step of the observation period. But the provided complete

analytical description can be used to calculate the expected value for every time step, too. So, statistical modelling represents an alternative approach for estimating broadly used trends in observational precipitation time series. In this way non-Gaussian characteristics can be taken into account. Additionally, changes in different parameters can be considered and the mean squared error of the trend estimator is smaller using the statistical modelling.

Application of the method to a German station network of 132 time series covering 1901-2000, shows in winter and summer at several stations in the southern part of Germany an increase in the probability of exceeding the 95th percentile and a decrease in the probability of falling under the 5th percentile. In the western part, we observe the same phenomenon in the summer months, but these changes go along with smaller magnitudes. However, climate is getting more extreme in that region in winter: Probability for both exceeding the 95th percentile and for falling under the 5th percentile is increasing. In the eastern part of Germany, increases in the probability of occurrence of relatively low precipitation in winter as well as decreases in both probabilities (>95th percentile, <5th percentile) in summer and autumn prevail.

Exemplary, the trend map on the basis of the familiar least squares method is compared to the trend map calculated on the basis of the Gumbel model. Both maps show positive trends in the western and southern part of Germany and negative trends in the eastern part. However, the robust method provides smaller amplitudes.

Acknowledgments. We gratefully acknowledge that this work is supported by the Federal German Ministery of Education and Research (BMBF) under grant No. 01LD0032 within the context of the German Climate Research Programme (DEKLIM, `http://www.deklim.de`).

References

1.1 **Coles**, S., *An introduction to statistical modeling of extreme values*. Springer Verlag, 2001.
1.2 **Grieser**, J., S.Trömel, C. D. Schönwiese, 2002: *Statistical time series decomposition into significant components and application to European temperature*. Theor. Appl. Climatol. 71, 171-183.
1.3 **Huber**, P. J., *Robust Statistics*. Wiley Series in Probabiliy and Mathematical Statistics, New York, 1981.
1.4 **Leadbetter**, M. R., G. Lindgren, H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, 1983.
1.5 **Mann**, H. B., 1945: *Nonparametric test agains trend*. Econometrica, 13, 245-259.
1.6 **Press**, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in Fortran 77*. Cambridge Univeritiy Press, 1992.
1.7 **Rinne**, H., *Taschenbuch der Statistik*. Verlag Harri Deutsch, Thun/Frankfurt, 1997.
1.8 **Schrader**, K. and T. Hettmansperger, *Robust Analysis of variance upon a likelihood ratio criterion*. Biometrika, 67, 93-101, 1980.
1.9 **Storch**, H. v. and F. W. Zwiers, *Statistical Analysis in Climate Research*. Cambridge University Press, 1999.

1.10 **Trömel**, S, *Statistische Modellierung monatlicher Niederschlagszeitreihen.* Report No. 2, Inst. Atm. Umwelt, Universität Frankfurt a.M., 2005.
1.11 **Trömel**, S., C. D. Schönwiese, 2005:*A generalized method of time series decomposition into significant components including probability assessments of extreme events and application to observational German precipitation data.* Met. Z. 14, 417-427.
1.12 World data center for climate, Hamburg. http://cera-www.dkrz.de/

peer-reviewed by 2 reviewers

Glossary

ACF	autocorrelation function
AP1824	Area Precipitation 1824
ASA	autocorrelation spectrum analysis
CDF	cumulative distribution function
CMDS	classical multidimensional scaling
CWT	continous wavelet transformation
DEM	Digital Elevation Model
DJF	December, January, February
DFA	detrended fluctuation analysis
DWD	Deutscher Wetterdienst (German Weather Service)
ECHAM	
ENSO	El Niño/Southern Oscillation
EOF	empirical orthogonal function
EXP	exponential distribution
FA	(standard) fluctuation analysis
FARIMA	fractionally integrated auto-regressive moving average process
FAR	fractionally integrated auto-regressive process
FMA	February, March, April
FFT	fast Fourier transform
FFM	Fourier Filtering Method
GEV	generalised extreme value distribution
GLO	generalised logistic distribution
GNO	generalised normal distribution
GPD	generalized Pareto distribution
GRDC	Global Runoff Data Centre
HadRM	Hadley Centre regional model
i.i.d.	independent and identically distributed
ISOMAP	isometric feature mapping
JJA	June, July, August

LARSIM	Large Area Runoff Simulation Model
LRD	long-range dependence
MC	Monte-Carlo simulations
MICE	modelling impact of climate change
MDS	multidimensional scaling
ML	maximum likelihood
MLE	maximum likelihood estimation
MRL	mean residual life plot
M-SSA	multichannel singular system analysis
$\mathcal{N}(\mu, \sigma^2)$	Gaussian/normal distribution with mean μ and variance σ^2.
N	Precipitation
NAO	North Atlantic Oscillation
NLDR	nonlinear dimensionality reduction
OND	October, November, December
PCA	principal component analysis
PC	principal component
PDF	probability density function
PE3	Pearson type-III distribution
POT	peak over threshold
PP	point process
PUB	prediction of ungauged basins
PWM	probability weighted moments
R	environment for statistical computing `www.r-project.org`
RC	reconstructed component
ROI	
SLP	see level pressure
SRD	short-range dependence
SSA	singular system analysis
t	air temperature
u	wind speed
UNF	uniform distribution
$W\mathcal{N}(\mu, \sigma^2)$	Gaussian/normal white noise
WAK	Wakeby distribution
WTMM	
d	Wind Direction
m a.s.l.	meter above sea level

Subject Index

Entries in this index are generally sorted with page number as they appear in the text. Page numbers that are marked in **bold face** indicate that the entry appears in a title or subheading. Page numbers that are marked in *italics* indicate that the entry appears in the caption of a table or figure.

D