Meteorologische Zeitschrift, Vol. **14**, No. 3, 1-11 (June 2005)

Article

A generalized method of time series decomposition into significant components including probability assessments of extreme events and application to observational German precipitation data

SILKE TRÖMEL* and CHRISTIAN-D. SCHÖNWIESE

Institute for Atmosphere and Environment, J. W. Goethe University, Frankfurt, Germany

(Manuscript received September 15, 2003; in revised form April 8, 2004; accepted May 5, 2005)

Abstract

The analysis of climate variability realized in time series of observational data needs adequate statistical methods. In particular, it is important to estimate reliably significant structured components like the annual cycle, trends, the episodic component and extreme events including variations of these components. In this issue estimators are called "reliable", if a priori assumed statistical assumptions are fulfilled. However, climate change concerns not only the mean value of meteorological variables, but all parameters of any related frequency distribution. In consequence, a generalized time series decomposition technique is presented allowing a free choice of the underlying probability density function (PDF). The signal (structured components like trends etc.) is detected in two instead of one parameter of a PDF, which can be chosen without any further restriction. So, the scale parameter of any PDF is no longer seen as a constant but rather affected by a deterministic process. The trend and seasonal component reflected in both parameters under consideration are estimated simulaneously in a modified stepwise regression. To deal also with superposed polynomial components and extreme events an iterative procedure is applied that converges to robust estimates of all the components. In particular, the method allows a consistent decomposition of precipitation time series into a statistical and a deterministic component. It arises, that in the special case of 132 time series of monthly precipitation totals 1901– 2000, from German stations, the interpretation as a realization of a Gumbel-distributed random variable with time-dependent scale and location parameter reveals a complete analytical description of the time series. In addition to the detection of the components mentioned above, now it is possible to quantify the probability of exceeding optional upper or lower thresholds, respectively, for any time step of the observation period.

Zusammenfassung

Die Analyse der Klimavariabilität, wie sie in Zeitreihen beobachteter Klimadaten erfasst wird, erfordert adäquate statistische Methoden. Insbesondere ist es wichtig, die darin enthaltenen signifikanten strukturierten Komponenten wie Jahresgang, Trends, episodische Komponente und Extremereignisse sowie deren Änderung verlässlich zu schätzen, das bedeutet, die statistischen Annahmen der Schätzer sollten erfüllt sein. Klimaänderungen müssen sich jedoch nicht nur auf das Mittel meteorologisch relevanter Variablen beschränken, sondern können sich in allen Parametern der betreffenden Häufigkeitsverteilung widerspiegeln. Daher wird hier eine verallgemeinerte Methode der Zeitreihenzerlegung vorgestellt, welche eine freie Wahl der zugrundeliegenden Wahrscheinlichkeitsdichtefunktion erlaubt. Das Signal (strukturierte Komponenten wie Trends usw.) wird nun in zwei statt nur einem Parameter einer ohne Einschränkungen wählbaren Wahrscheinlichkeitsdichtefunktion detektiert. So wird beispielsweise der Skalenparameter nicht mehr als stochastische Konstante des Prozesses, sondern vielmehr als ebenso determinisch veränderliche Variable betrachtet. Die Trend- und Saisonkomponente beider Parameter werden simultan in einer zu diesem Zweck modifizierten schrittweisen Reggression geschätzt. Um aber auch noch überlagerte polynomiale Schwankungen und Extremereignisse handhaben zu können, wird ein iteratives Verfahren angewandt, welches zu robusten Schätzern aller Komponenten konvergiert. Die Methode erlaubt insbesondere eine konsistente Zerlegung monatlicher Niederschlagsreihen in einen statistischen und deterministischen Anteil. Dabei zeigt sich, dass in dem hier betrachteten speziellen Fall von 132 Reihen monatlicher Niederschlagssummen 1901–2000 deutscher Stationen die Interpretation als Realisation einer Gumbel-verteilten Zufallsvariablen mit variablem Lage- und Streuparameter ein adäquates Modell zur vollständigen analytischen Beschreibung der Zeitreihen liefert. Somit sind neben der Erfassung der oben genannten Komponenten auch Über- und Unterschreitungswahrscheinlichkeiten beliebig wählbarer Schwellenwerte für jeden Zeitpunkt des Beobachtungszeitraums angebbar.

*Corresponding author: Silke Trömel, Institut für Atmosphäre und Umwelt, J. W. Goethe Universität, Postfach 11 19 32, 60054 Frankfurt a. M., Germany, e-mail: s.troemel@meteor.uni-frankfurt.de

0941-2948/2005/0039 $ 4.95

DOI: 10.1127/0941-2948/2005/0039

1 Introduction

The analysis of climate variability as reflected in observational records is an important challenge in statistical climatology. Especially during the last decades, the influence of anthropogenic forcing on climate has become of major public and scientific interest. But climate change is not restricted to the mean value of meteorological variables like temperature or precipitation. Increasing atmospheric greenhouse-gas concentrations may also lead to increased variability (SCHÄR et al., 2004) which, in turn, is important with respect to the probability of extremes (KATZ and BROWN, 1992). Anyway, the aim of the paper presented here is the detection and description of significant components and their variations in observed climate time series especially with regard to non-Gaussian precipitation time series without attribution to forcings.

Various methods exist for the analysis of time series. Fourier transformation **(?)fehlt Reference Fourier** or wavelet decomposition techniques (MALLAT, 1997), for example, yield a complete decomposition of time series using orthonormal basis functions. But the functions offered may not be of physical relevance. In this study we decompose times series into signal and noise, using only a limited basis which can be interpreted from a climatological point of view. Therefore we accept non-orthogonal basis functions. The signal is defined as a sum of constant or significantly changing time series components as annual cycle, trends, extreme events, episodic and harmonic components.

However, only a suitable analysis technique based on an adequate statistical model leads to reliable statements about significant changes in the components of the time series. GRIESER et al. (2002) decompose surface air temperature into a structured component and stationary Gaussian distributed noise. Superposition of detected components give the time-dependent average of the assumed Gaussian probability density function.

However, using the limited basis for the signal, an application of this strategy to non-Gaussian climate variables leads to insufficient or even wrong statements about changes in climate variability.

As an example, Figure 1 shows the observed spring precipitation totals 1900–1999 at the German station Kleinwaabs. Application of the least squares method detects a positive trend of 27.4 mm per century (so-called trend-to-noise-ratio tr/s=0.613; s=standard deviation), caused by the sensitivity with respect to extreme values. So, a shift of the assumed Gaussian distribution to higher values, implying simultaneously a decrease of low precipitation totals, is "detected". On the contrary, we observe the highest and the lowest precipitation totals near the end of the time interval considered. So, the trend "detection"is a misinterpretation.

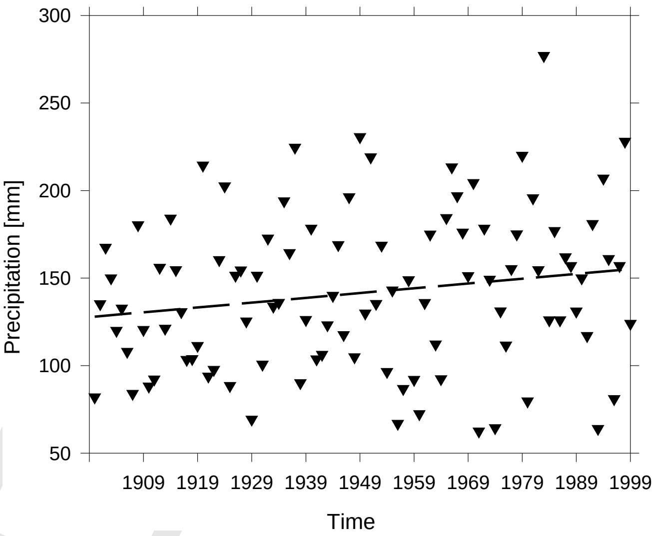

Figure 1: Spring precipitation totals (March to May, triangles) at the station of Kleinwaabs (54.53°N, 9.75°E) and the positive linear trend (line) the least squares estimator suggests.

In the following, in Chapter 2 a generalized time series decomposition technique is introduced where, on the basis of any frequency distribution two free parameters instead of one are used. As an example of this strategy, on the basis of a Gumbel distribution, these parameters are a time-dependent location and a time-dependent scale parameter. Now, changes in spring precipitation totals at station Kleinwaabs, presented in Figure 1, can be conceived as a linear positive trend in the scale parameter of a Gumbel distribution. The scale parameter is no longer seen as a constant of the underlying stochastic process, but as an additional variable to describe changes of structured components in observed time series.

Chapter 3 presents some results of this strategy applied to time series of monthly precipitation totals, 1901–2000, observed at 132 stations in Germany. Finally, Chapter 4 contains a brief discussion of the results and an outlook.

2 Method

As mentioned before,GRIESER et al. (2002) 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. Here, a flexible strategy is introduced to obtain robust estimates of all the components under consideration. The definition of the components and the iterative procedure to detect them remain similar as described in GRIESER et al. (2002). The differences are listed below.

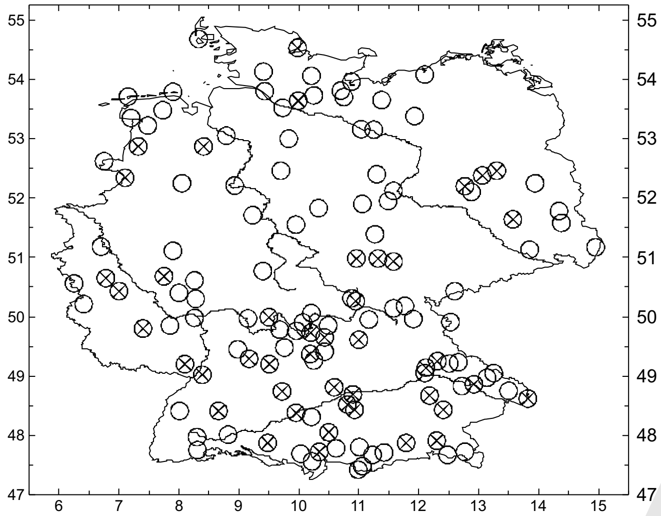

Figure 2: The stations spatial distribution of the considered 132 precipitation time series (circles). Stations with no significant changes in the annual cycle of the scale parameter are signed by crosses.

2.1 Basis functions

The equation

$$S_{j,k}(t) = d_{j,k}t^k \cos\left(2\pi\frac{j}{12}t\right) + e_{j,k}t^k \sin\left(2\pi\frac{j}{12}t\right) \tag{2.1}$$

with wave number $j = 1,\ldots 6$ per year and $k = 0,1,2$ gives the basis functions to describe the seasonal component. 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 wavenumber j but different time dependence k in one time series yields

$$S_j(t) = A_j(t)\cos\left(2\pi\frac{j}{12}(t-t_j)\right) \tag{2.2}$$

with amplitude

$$A_j(t) = \sqrt{\sum_{k=0}^{2}\left(d_{j,k}^2 + e_{j,k}^2\right)t^{2k}} \tag{2.3}$$

and phase

$$t_j(t) = \frac{12}{2\pi} arctan\left(\frac{\sum\limits_{k=0}^{2} d_{j,k}t^k}{\sum\limits_{k=0}^{2} e_{j,k}t^k}\right). \tag{2.4}$$

On that way, 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:

$$T_i(t) = g_i + h_i t^i \quad \text{with } i = 1,\ldots,5. \tag{2.5}$$

Thus, we are able to detect linear, progressive and degressive trends.

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 Section 2.2).

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:

$$V_l(t) = c_o + \sum_{i=1}^{l} c_i t^i. \tag{2.6}$$

Using an iterative procedure to find within these functions the best model equations of two instead of one parameter of a probability density function 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:

$$p_1(t) = \sum S_{j,k}(t) + \sum T_i(t) + V_l(t). \tag{2.7}$$

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

$$p_2(t) = S_{1,0} + T_i. \tag{2.8}$$

If in the second step a polynomial variation $V_l(t)$ is detected in $p_1(t)$, the estimated coefficients in model equation of $p_2(t)$ are new estimated.

Which PDF to choose and which of the parameters gets the larger number of degrees of freedom depends on the characterics of the time series. Not until the end of the decomposition the statisical properties of the residuals can confirm the chosen model as adequate description of the time series (see Section 3.5). If we assume time-dependent parameters, and the decomposition is based on the probability density function of the Gumbel distribution

$$f(x,t) = \frac{1}{b(t)}\left\{\exp\left(-\frac{x-a(t)}{b(t)}\right)\exp\left[-e^{-(x-a(t))/b(t)}\right]\right\}, \tag{2.9}$$

and the scale parameter $b(t) = p_1(t)$ and the location parameter $a(t) = p_2(t)$ we call it the Gumbel model with emphasis on the scale. We call the statistical model the Gumbel model with emphasis on location if the location parameter $a(t) = p_1(t)$ and the scale parameter $b(t) = p_2(t)$.

Table 1: Amplitudes and phase angles in days relative to 15th December for significant functions of the annual cycle at the stations Osnabrück, Glückstadt and Rostock. See Trömel (2004) for detailed information of all 132 stations.

Station	Parameter	Function	Amplitude $[mm/t^k]$	Phase angle $[days]$
Osnabrück (52.25^oN, 8.05^oE)	b(t)	$S_{1,0}$	-6.48	87.3
		$S_{1,1}$	$0.75 \cdot 10^{-2}$	60.13
	a(t)	$S_{1,0}$	-6.94	86.00
Glückstadt (53.80^oN 9.43^oE)	b(t)	$S_{1,0}$	-6.88	79.9
		$S_{2,1}$	$0.267 \cdot 10^{-2}$	37.83
	a(t)	$S_{1,0}$	-12.32	79.00
Rostock (54.08^oN 12.1^oE)	b(t)	$S_{1,0}$	-7.59	82.1
		$S_{1,1}$	$-0.416 \cdot 10^{-2}$	-52.04
		$S_{2,1}$	$0.349 \cdot 10^{-2}$	34.89
		$S_{4,2}$	$0.264 \cdot 10^{-5}$	15.21
	a(t)	$S_{1,0}$	-10.49	66.20

Table 2: Changes in amplitudes $\Delta A_j = A_j(1200) - A_j(1)$ and changes in phase angles $\Delta t_j = t_j(1200) - t_j(1)$ for significant functions of the seasonal component at the stations Osnabrück, Glückstadt and Rostock.

Station	Parameter	wavenumber j	$\Delta A_j[mm]$	$\Delta t_j[days]$
Osnabrück	b(t)	1	-0.62	-57.54
Glückstadt	b(t)	2	3.85	–
Rostock	b(t)	1	-1.91	-51.82
		2	5.02	–
		4	5.48	–

In a third step a search for extreme events, defined as unexpected values on the basis of the model, is performed. 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. for further details). If no further extreme events are found the iterative procedure for the detection of trends, seasonal cycle and polynomial equation is applied again. It depends on the robustness (HUBER, 1981) of the distance function whether structured components are more or less influenced by extreme events. The procedure is terminated if no further extreme events can be found.

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 G(0,1) after elimination of $p_1(t)$ and $p_2(t)$. Additional stationarity of the distribution points to a complete description of the time series within the PDF and its time-dependent parameters.

Nevertheless, the results of a statistical test can only show whether the model found is one possible choice. To decide between several possible models, the residual distribution and the stationarity condition contain not necessarily the information wanted. On the one hand, instationarity or rejection of the assumed residual distribution points to further undetected structured components or a bad choice of the model. On the other hand, several models can provide stationarity and confirmation of the residual distribution, but results, for example probability of exceedance of upper thresholds, may differ significantly. However, regarding a time series $x(t_1), x(t_2) \dots x(t_n)$ the likelihood

$$f(x(t_1), p_1(t_1), p_2(t_1)) \cdot f(x(t_2), p_1(t_2), p_2(t_2)) \cdot \dots f(x(t_n), p_1(t_n), p_2(t_n)) \quad (2.10)$$

shows whether statistical modelling of a times series on the basis of the PDF f and the estimated parameters p_1 and p_2 leads to a better description of the series than the PDF $\tilde{f}$ with the parameters $\tilde{p_1}$ and $\tilde{p_2}$.

2.2 The model selection criterion

The basis of any time series decomposition technique are the distance function and a model selection criterion. Consistently with the maximum likelihood principle the distance function, defined as the negative logarithm of the PDF, replaces the function of squared errors to be minimized. But besides the systematic approach to search for estimators, further advantages of the maximum likelihood principle are resulting consistent and asymptotically normal estimators (STORCH and

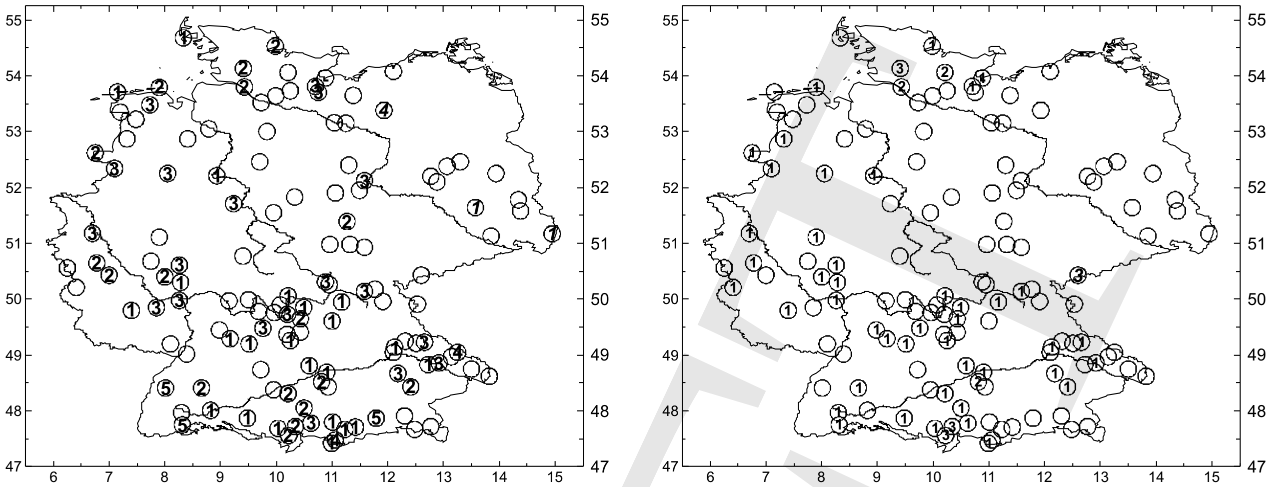

Figure 3: Spatial distribution of detected trends in the location (left) and in the scale parameter (right). The numbers give the order of the detected trend. Negative trends are given by italic numbers.

Table 3: Number of detected trends in the locatin and scale parameter of the Gumbel distribution.

	Scale parameter				Location parameter						
	pos.			neg.	pos.					neg.	
Order	1	2	3	3	1	2	3	4	5	1	4
Number	44	3	3	1	23	17	19	2	3	2	1

ZWIERS, 1999, p. 89). 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

$$\rho(x,t) = ln(b(t)) + exp\left(-\frac{x-a(t)}{b(t)}\right) + \frac{x-a(t)}{b(t)}. \tag{2.11}$$

The coefficients in the model equations for estimating $a(t)$ and $b(t)$ are chosen by minimizing

$$\sum_t \rho\,(x,t) = min., \tag{2.12}$$

equivalent to the maximum of the loglikelihood function. Powells method (**?**, p. 406) is used to minimize ρ in the multidimensional space.

Stepwise regression (STORCH and ZWIERS, 1999, 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, also depends on the selected model equation describing the second distribution parameter (see Eq. 2.11). Between these alternate parts redefinition of the second ($a(t)$) parameter 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. 2.12). Define $D(p)$ as the minimum value of (2.12) subject to the model containing p regressors and $D(q)$ as the minimum value of (2.12) subject to the model containing q regressors, with p>q, a direct generalization of the common F-test may be based upon the statistic

$$F_M = \left(D(p) - D(q)\right) / (p-q) \tag{2.13}$$

(SCHRADER and HETTMANSPERGER, 1980) 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.

The modified stepwise regression including a modified F-statistic (Eq. 2.13) represents the basis of a generalized time series decomposition technique. Inserting

Table 4: Number of detected trends in the locatin and scale parameter of the Gumbel distribution.

Station	Parameter	Order of the trend	Amplitude [mm]
Osnabrück	b(t)	1	6.85
	a(t)	3	7.09
Glückstadt	b(t)	2	6.5
	a(t)	2	7.78
Rostock	b(t)	–	–
	a(t)	–	–

the corresponding distance function it can be used for time series analysis based on any data distribution and any statistical model. Thereby, the choice of the probability density function and the corresponding distance function should depend on the characteristics of the climate variable under consideration and can be verified in the residual analysis at the end of the decomposition.

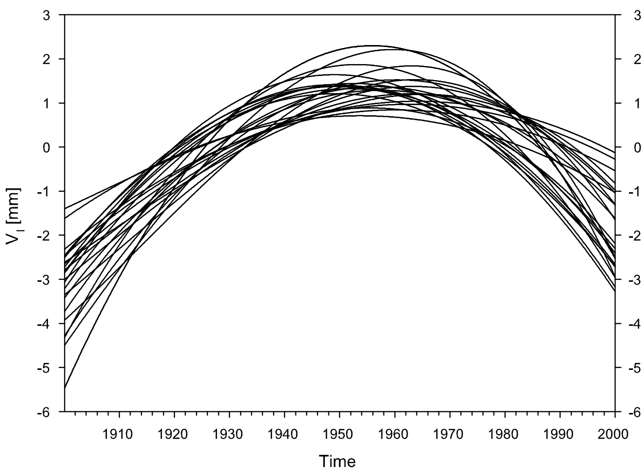

Figure 4: Development of all 24 detected low-frequency variations in the scale parameter.

3 Application

In the following subsections the results of the analysis of time series of monthly precipitation totals, 1901–2000, from 132 stations within Germany are presented. The time series are most likely homogeneous up to 1990 (SCHÖNWIESE and RAPP, 1997, p. 16) and updated without further homogeneity tests. Most likely homogeneous means here, that at least 3 out of 5 homogeneity tests decided to homogeneity instead of non-homogeneity. The spatial distribution of the stations can be seen in Figure 2.

To obtain a best as possible description of the time series several statistical models are considered: Interpretation as a Gaussian distributed random variable with time-dependent mean and variance (emphasis on location), as a Weibull distributed random variable with time-dependent scale and shape parameter (emphasis on scale) and interpretation as a Gumbel distributed random variable with time-dependent location and scale parameter, performed once with emphasis on location and once with emphasis on scale. Contrary to time series of monthly mean temperature the location parameter plays a minor role for the description of observed precipitation sums. In the following subsections results of a statistical modelling on the basis of the introduced Gumbel model with special emphasis on the scale are presented.

3.1 Seasonal component

The interpretation of observed monthly precipitation totals as a realization of a Gumbel distributed random variable with emphasis on scale offers the constant one cycle per year harmonic $S_{1,0}$ to the location parameter and seasonal variations $S_{j,k}$ with wave numbers $j = 1, \dots, 6$ with constant as well as linear and quadratic time dependence of the amplitude ($k = 0, .., 2$) to explain changes in the scale parameter. In doing this, in nearly all cases the one cycle per year harmonic $S_{1,0}$ contributes significantly to the scale and the location parameter. Only one of 132 time series shows no significant seasonal component in the location and one other time series shows no significant seasonal component in the scale parameter.

In 44 cases a constant annual cycle during the observation period is found. As can be seen in Figure 2, these stations in question are especially found in Western and Southern Germany. But the majority of stations show changes in the annual cycle. Changes in the amplitudes ($k \neq 0$) of the one cycle per year harmonic and the other harmonics are found in 88 time series. 54 of them also reveal changes in the phase angle of the seasonal component. As an example, quantitative results of 3 stations are listet in Table 1. Amplitudes and phase angles are given for significant functions $S_{j,k}$, where phase angles are given in days after (positive sign) the 15th december. All these stations reveal changes in the amplitude of the seasonal component in the scale parameter (see Table 2). In Osnabrück and Rostock the functions $S_{1,0}$ and $S_{1,1}$ are significant. Superposition of functions $S_{j,k}$ with the same wave number provide changes in the phase angle of the harmonic.

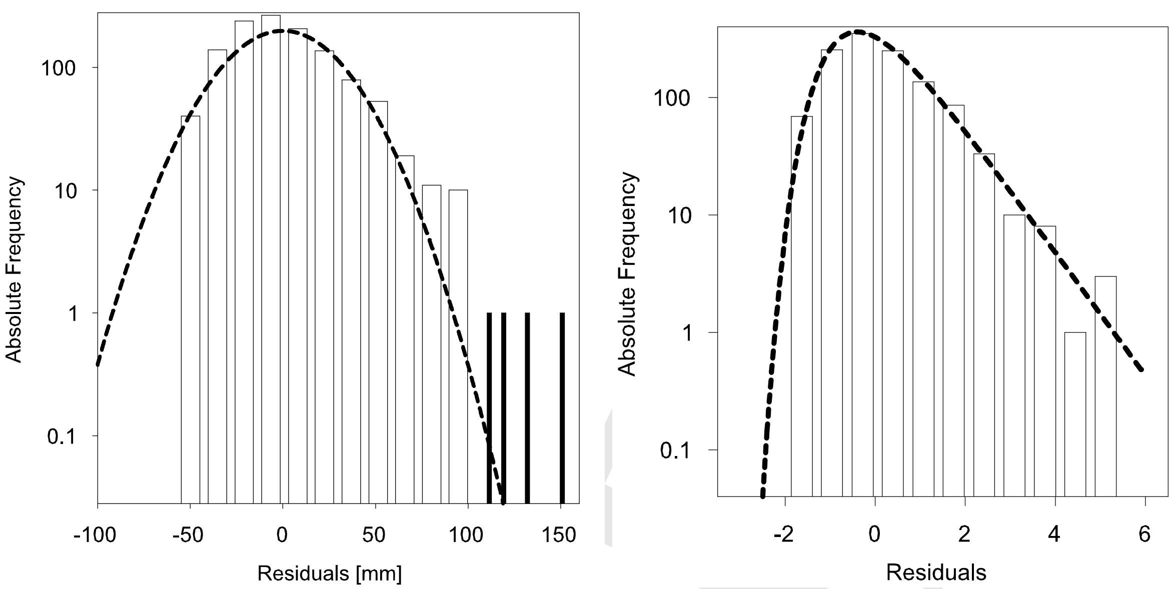

Figure 5: Empirical frequency distribution (bars) of the remaining residuals at the station Alsfeld (50.77°N, 9.40°E) after a time series analysis based on Gaussian distribution with constant variance together with the fitted Gaussian distribution (dashed line) and detected extreme events (black lines) are shown in the left plot. The right plot shows the respective residuals after a time series analysis based on the Gumbel distribution together with the fitted theoretical distribution. All plots are given in logarithmic scale to let deviations become more visible.

3.2 Trends

In case of trend analysis, we allow trends up to the order 5 for the model equations of the scale and the location parameter. Numbers encircled in Figure 3 specify the order of the detected trend in the location (left plot) and the scale parameter (right plot). Concerning both parameters the overwhelming majority of trends are detected in western and southern Germany. The eastern part of Germany reveals a small number of negative trends in the location parameter. We detect 51 trends concerning the scale parameter. 44 of them are positive and linear. But only 3 positive trends are of order 2 and 3 positive trends are of order 3. In the scale parameter only one negative trend of order 3 is detected.

Alltogether 67 out of 132 time series include a trend in the location parameter of the Gumbel distribution. The overwhelming majoritiy of 64 are positive ones. As can be seen in Table 3, trends of the order 1 to 3 dominate. Comparing the left and right plot of Figure 3 we see a superposition of positive trends in the location and the scale parameter for some stations, indicating a clear increase of relatively wet months. The effect for relatively dry months depends on the magnitudes of the competing trends. Positive trends in the location parameter and positive ones in the scale parameter reinforce the occurence of relatively wet months. But a shift to higher values and a spread of the distribution have contrary effects on the occurence of relatively dry months.

Quantitative results are again presented for the stations Osnabrück, Glückstadt and Rostock. Table 4 shows order and amplitude of detected trends for both, the scale and the location parameter. We detect no trends in the time series data observed in Rostock. Note that the expected value $\mu(t)$ at time t of a Gumbel-distributed random variable with time-dependent parameters is given by

$$\mu(t) = a(t) + \gamma b(t) \tag{3.1}$$

with Eulers constant $\gamma = 0.57722$, indicating that changes in the scale and the location parameter reveal changes in the mean value. So, in case of a sucessfull decomposition, the full analytical description of monthly precipitation series provides the expected value for every time step of the observation period. This result can be used to take into account the skewness of the distribution and changes in different parameters to estimate trends in the mean value.

3.3 Episodic component

Within the chosen Gumbel model with emphasis on the scale, low-frequency variations are only considered with respect to the scale parameter. 24 low-frequency variations are detected, 15 of them are polynomial equations of order 2 and 9 of them are of order 3. In Figure 4 we see, that they describe a similar time history: An increase in variability until approximately 1960 and a subsequent less pronounced decrease are detected.

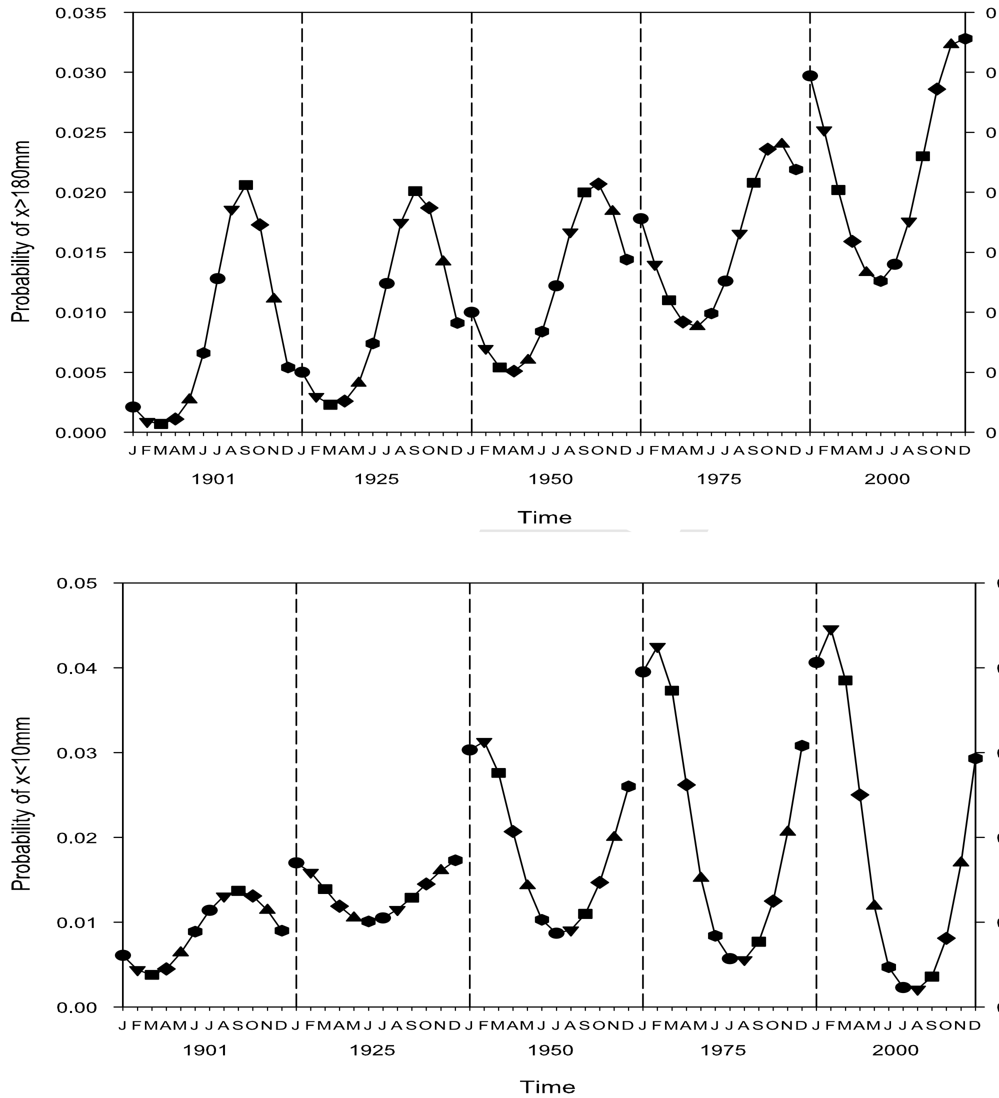

Figure 6: Annual cycles of the probability of monthly precipitation totals more than 180 mm (upper plot) and less than 10 mm (lower plot) at Osnabrück (52.25°N, 8.05°E) in the years 1901, 1925, 1950, 1975 and 2000.

In the time series observed at Osnabrück, Glückstadt and Rostock (chosen examples in Section 3.1 and 3.2) no significant episodic components are detected.

3.4 Extreme events

In contrast to extreme values (see section 4), according to GRIESER et al. (2002), we define extreme events as a relatively small number of extreme values which are unexpected within the scope of the fitted statistical model; see also SCHÖNWIESE et al. (2003). Obviously, this definition strongly depends on the basis functions offered for the description of the signal as well as on the choice of the probability density function the decomposition is based on. So, for example, proceeding from a Gaussian model to a model based on the Gumbel distribution with time-dependent location and scale parameter, the analysis of precipitation time series may reveal less extreme events. One reason of this effect is the larger tail of the Gumbel distribution. In Figure 5 a case study of an unjustified separation of extreme events, based on the application of a Gaussian model, is illustrated. After the elimination of the significant structured components, four unexpected values are detected at the station Alsfeld (50.77N 9.40E) on the basis of an Gaussian model fitting (left plot). However, the Kolmogorov Smirnov-test rejects the Gaussian distribution for the remaining data either (Si=99.96%). So an application of a time series analysis based on the Gumbel distribution is performed alternatively which provides a more complete description of the series. Actually, no extreme events are detected now (right plot), which differ significantly from a Gumbel distribution.

Altogether, using a Gumbel distribution, the analysis of the 132 time series under consideration reveals only 7 extreme events. All these extreme data are wet months,

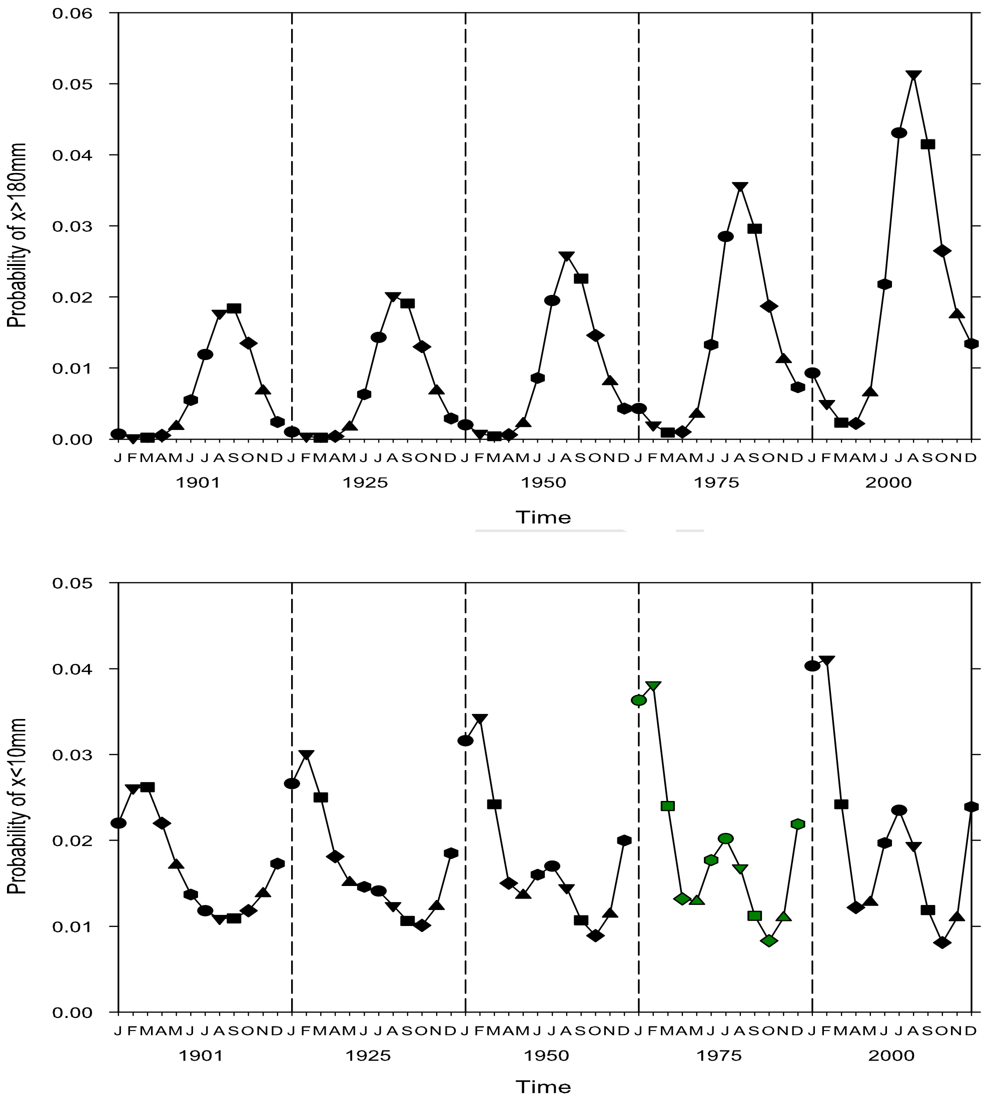

Figure 7: Annual cycles of the probability of monthly precipitation totals more than 180 mm (upper plot) and less than 10 mm (lower plot) at Glückstadt (53.80°N 9.43°E) in the years 1901, 1925, 1950, 1975 and 2000.

3 of them occur in July, 2 in February, one in March and one in January.

3.5 Residuals

The analysis of the remaining residuals after the time series decomposition represents an important part of this analysis 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. In case of the precipitation analysis presented in this paper, 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. The Kolmogorov-Smirnov test (see again PRESS et al., 1992) 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. Consequently, residuals can be seen as G(0,1)-distributed. In addition, the stationarity of the residuals is checked. Thereby, the sampling time is divided into two subintervals to compare the corresponding distributions using again a Kolmogorov-Smirnov test statistic. It arises that in 6 cases stationarity is rejected with a probability larger than 90%. In consequence, a complete description of the observed time series on the basis of the Gumbel distribution with emphasis on the scale could be achieved.

4 Probability of exceeding specific thresholds

In case of a successful time series decomposition, the analysis described above provides the related probability density functions f(x,t), which is the probability, that the precipitation total is between x and x+dx at the time t.

The integration of these density functions quantifies the probability of exceeding a defined threshold at any given time. In case of the chosen Gumbel model (see again Eq. 2.9) with emphasis on scale we yield for every time series the equation

$$\begin{aligned} p(x > x_s, t) &= 1 - \int_{-\infty}^{x_s} \frac{1}{p_1(t)} \left\{ \exp\left(-\frac{x - p_2(t)}{p_1(t)} \right) \exp\left[-e^{-(x-p_2(t))/p_1(t)} \right] dx \right\} \\ &= 1 - \exp\left\{ -\exp\left(-\frac{x_s - p_2(t)}{p_1(t)} \right) \right\}. \end{aligned} \tag{4.2}$$

So, variations in the probabilities of exceedance deduced from the PDF reflect detected structures in the estimated location and scale parameter $p_2(t)$ and $p_1(t)$. As an example, Figure 6 shows for the years 1901, 1925, 1950, 1975 and 2000 and the months January to December the probability of a monthly precipitation total larger than 180mm (upper plot) and lower than 10mm (lower plot) calculated for the station Osnabrück. Changes in the probabilities are caused by the one cycle per year harmonic and a positive trend of the order 3 in the location parameter and a detected seasonal component such as a positive linear trend in the scale parameter (see again Tables 1, 2 and 4). But the seasonal component concerning the scale parameter changes during the observation period. Changes in amplitude and phase angle of the annual cycle are detected. The consequence is a more pronounced increase in the probability of a precipitation total larger than 180 mm in winter compared to summer. While in July the probability of exceedance remains nearly constant, we observe in January an increase of the probability amounting of about 3 %. At the beginning of the 20th century, the maximum for the probability of exceedance was in September, at the end of the century the highest probability is observed in December. Taking also into account changes in the probability of a monthly precipitation total less than 10 mm, we see a simultaneous increase of both probabilites in winter, indicating an inrease in variability and a tendency to more extremes in winter. In contrast to that, in summer and autumn we observe small changes in the probability for wet months and a decrease in the probability of monthly precipitation totals less than 10 mm.

The same procedure of probability computations is applied on the time series of Glückstadt, see Figure 7. We detect the one cycle per year harmonic $S_{1,0}$ in the location parameter and in the scale parameter, characterized by nearly the same phase angle (see again Tables 1, 2 and 4). In addition, changes in the shape (amplitude) of the seasonal component concerning the scale parameter and positive trends of the order 2 in both parameters are detected. A maximum for exceeding the 180 mm level remains to be characteristic for stations in southern Germany in summer. Taking into account also the time history of the probability of a monthly precipitation total less than 10 mm, we see a simultaneous increase in both probabilities especially in July, indicating an increase of relatively dry and relatively wet months.

5 Conclusions and outlook

The aim of the contribution presented in this paper is the introduction of a generalized consistent decomposition procedure of precipitation time series into a statistical and a deterministic part. 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. We see, that the scale parameter plays the dominant role for explaining the deterministic components of the time series, contrary to traditional interpretations as a constant of the underlying process. Taking into account trends, constant or changing annual cycle and a polynomial component for the scale parameter of the Gumbel distribution, a full analytical description of the time series can be achieved.

In this paper, the authors retrict the application of the method to German observation data. But the application of the decomposition technique on the basis of the Weibull distribution provides also a statistical modelling of European and partly extra-European precipitation data. The decomposition technique based on the Weibull distribution allows to extract structured components also in the shape parameter of the distribution. Successful statistical modelling of a sufficient data base allows an insight into spatio-temporal changes in the probability of extremes (relatively wet or relatively dry months) during the 20th century e.g.

Additionaly, estimating changes in the expected value $\mu(t)$ on the basis of the statistical modelling gives the possibility to take into account changes in the location, the scale and shape parameter of the distribution, because Eq. 3.1 (see again Section 3.2) gives the expected value at a given time in case of the interpretation of the time series as a realization of a Gumbel-distributed random variable and

$$\mu(t) = a_W(t) + b_W(t)\Gamma\left(1 + \frac{1}{c_W}\right). \tag{5.1}$$

with the Gamma-function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{5.2}$$

gives the expected value at a given time if the Weibull modell is applicated. Here $a_W(t)$, $b_W(t)$ and $c_W(t)$ denote the location, the scale and the shape parameter of the Weibull distribution. Consequently, the method introduced provides trend estimators, which take into account the non-Gaussian characteristics of observed precipitation time series and will be subject of a further publication.

Acknowledgements

We gratefully acknowledge that this work is supported by the Federal German Minister of Education and Research (BMBF) under grant No. 01 LD 0032 within the context of the German Climate Research Programme (DEKLIM, www.deklim.de).

References

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.

HUBER, P. J., 1981: Robust Statistics. – Wiley Series in Probabiliy and Mathematical Statistics. New York. **Pages??**

KATZ, R., W.B.G. BROWN, 1992: Extreme events in a changing climate: Variability is more important than averages. Climate Change **21**, 289–302.

MALLAT, S. , 1997: A wavelet tour to signal processing. – Academic Press. **Pages??**

PRESS, W. H. , S. A. TEUKOLSKY, W. T. VETTERLING, B. P. FLANNERY, 1992: Numerical Recipes in Fortran 77. – Cambridge Univerity Press. **Pages ??**

SCHÖNWIESE, C.-D., J. RAPP, 1997: Climate trend atlas of Europe based on observations 1891–1990. – Dordrecht, Kluwer Academic Publishers, 228 pp.

SCHÄR, C. , P.L. VIDALE, D. LÜTHI, C. FREI C. HÄBERLI, M.A. LINIGER, C. APPENZELLER, 2004: The role of increasing temperature variability in European summer heatwaves. – Nature **427**, 332–336.

SCHLITTGEN, R.,B.H.J. STREITBERG, 1999: Zeitreihenanalyse. – Oldenbourg. **Pages ??**

SCHÖNWIESE, C.-D., J. GRIESER, S. TRÖMEL, 2003: Secular change of extreme monthly precipitation in Europe. – Theor. Appl. Climatol.**75**, 245–250.

SCHRADER, K., T. HETTMANSPERGER, 1980: Robust analysis of variance upon a likelihood ratio citerion. – Biometrika 67, 93–101.

VON STORCH, H., F. W. ZWIERS, 1999: Statistical Analysis in Climate Research. – Cambridge University Press. Pages ??

TRÖMEL, S, 2004: Statistische Modellierung monatlicher Niederschlagszeitreihen. – Dissertation, Report No. 2, Institute for Atmosphere and Environment, Johann Wolfgang Goethe University. **Pages ??**