Meteorological extremes and their variability on different time scales

Dissertation

Zur Erlangung des Doktorgrades der Naturwissenschaften im
Department Geowissenschaften der Universität Hamburg

vorgelegt von

Frank Sienz

aus

Kirchheim/Teck

Hamburg
2010

Als Dissertation angenommen vom Department Geowissenschaften der Universität Hamburg

Auf Grund der Gutachten von Prof. Dr. K. Fraedrich
und Dr. R. Blender

Hamburg, den 27. Oktober 2010
(Datum der mündlichen Prüfung)

Prof. Dr. Jürgen Oßenbrügge
Leiter des Department Geowissenschaften

Abstract

The analysis of extremes is of increasing importance in meteorological science. This is mainly due to the associated damages caused by the events and the expected changes going along with changing climate conditions. Extremes are of relevance in different fields and can be classified by the time scale of the generating processes. This circumstance complicates the analysis, beside their inherent rareness. These complications have to be overcome with adequate methods and the following are applied: Standardized Precipitation Index (SPI), extreme value statistics and extreme event return time analysis.

This study contributes to the description of extreme events. The more technical oriented part discusses shortcomings of the applied methods and resolves them partly. The impact of climate variability and climate change on frequency and intensity of the extremes is analyzed for Icelandic precipitation and North Atlantic cyclones. Further, extreme event return time properties are investigated in long-term memory processes near $1/f$.

Contents

1 Introduction

1.1 Extremes, their characteristics and definition

1.1.1 Severe meteorological events

One objective of weather forecasts is to give early warnings of upcoming severe weather events. This is of outstanding importance for shipping or air traffic, to name a few. The knowledge of future weather states comes along with the ability to take preventive measures, reducing damage costs and helping to save lives. In addition to short-term prediction, a climatological description of severe events helps for future planning. The long-term strategy for dikes is one example, where from observed sea levels a dike height has to be estimated to prevent from floods in the remote future.

The fundamental meteorological quantities, such as temperature, precipitation and wind speed are involved, if severe meteorological events are considered. However, a unifying property is often the associated exceptional damage caused by those events. Further, this events are unusual in terms of their rare occurrence. Hence, the individual characteristics could be quite different as the following examples demonstrate.

Venezuela flood, December 1999

The rainfall from 14 to 16 December 1999 in Venezuela caused one of the largest historical rainfall-induced debris flows ever documented worldwide. Extensive flooding came along with landslides and avalanches of mud, rocks and trees and caused between 20000 to 50000 fatalities.

The unusualness of this event is deduced from annual maximum daily rainfall amounts from 1951 to 1999 (Figure 1.1). The outstanding high value is achieved at 15 December 1999, reaching about $410mm/dy$. This is more than twice as large as any of the previously recorded annual maxima from 1951 to 1999. For comparison, the highest recorded daily precipitation amount measured at Dresden station was only slightly above $150mm/dy$. This value was recorded during the Elbe flood at 12 August 2002.
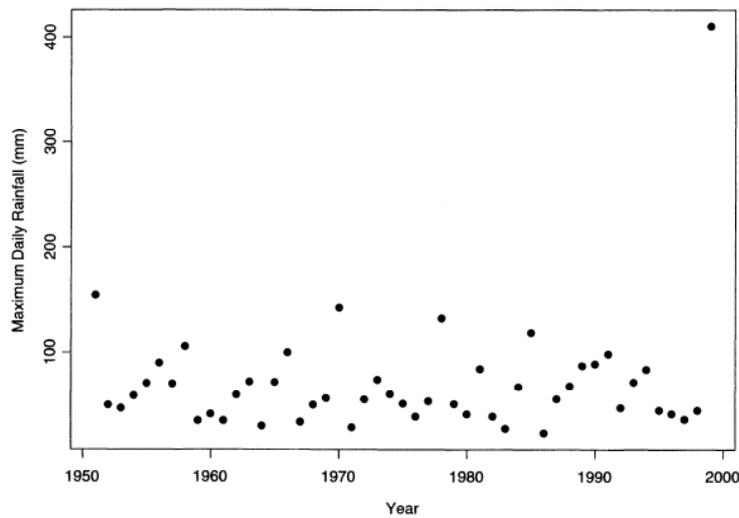
Figure 1.1: Annual maximum daily rainfall reported at Maiquetia international airport, Venezuela (Figure from Coles and Pericchi (2003)).

The 15 December event was embedded in a series of days with high precipitation amounts. However, it is this single peak value which is of interest. The reason is the outstanding high deviation from the other yearly maxima, which is easily referred just by visual inspection. Leading to the question: is it possible to forecast such an unusual high event on the basis of observational data ?

European heat wave, summer 2003

One of the most sever natural catastrophes in Europe is the heat wave in the year 2003. In the first half of August exceptional high and long lasting temperatures caused between 40 000 and 70 000 fatalities (World Health Organization (2004)).

The heat wave development is exemplified with observed daily mean data at the French station Besançon[1]. Positive temperature anomalies persisted throughout August and exceeded the 99% quantile at several successive days at the beginning of summer (Figure 1.2 a)). At the 7 August, the highest daily temperature was observed, ever recorded during the 1931 to 2004 period. High June and July temperatures preceded the August heat wave. Whereas the monthly mean July temperature reached a moderately high magnitude, as well as in August the June mean temperature broke any previous temperature record (Schär et al., 2004). This together leads to the highest observed summer temperature ever recorded, considerably outstanding the center of the mean summer temperature distribution (Figure 1.2 b)).

---

[1]Data from the European Climate Assessment & Dataset project (Klein Tank et al., 2002)
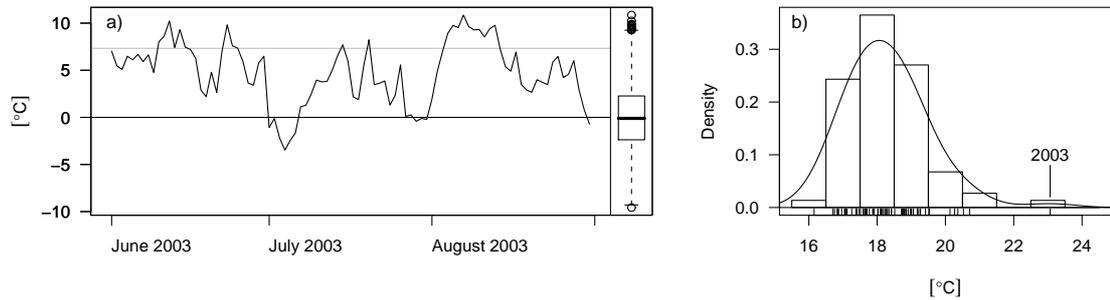
Figure 1.2: Observed summer temperatures at Besançon station (France): a) Year 2003 daily temperature anomalies, in respect to the climatological day (reference period 1931-2004) and box plot consisting of all anomalies on the right. The gray line gives the 99% quantile. b) Histogram of summer mean temperatures, as well as kernel density estimate (black line).

In contrast to the previous example the highest occurred daily value is not distinct different from previous maxima (box plot, Figure 1.2 a)). This points out, that a single peak value is of minor importance for the severeness of the heat wave in 2003. The important property however, is the persistence of high temperature anomalies on a large number of consecutive days.

Iberian Peninsula drought, 2005

The year 2005 was one of the driest years in Spain and Portugal since the beginning of the observational record. Drought conditions lead to rationing of water, losses in agriculture, enhanced forest fire risk and reduced electricity production through direct (hydroelectric power plants) or indirect effects (for example water cooling systems in nuclear power plants).

Figure 1.3 shows daily accumulated precipitation anomalies for three selected Spanish stations in years of major droughts. Since the Spanish water supply highly depends on precipitation stored mainly in autumn and winter, the lines start with the onset of the hydrological year in October. Despite differences between the single stations, a common property of all major drought events is the accumulation of precipitation deficits over a long time period of 9 months. From October 2004 to June 2005 the accumulated precipitation percentage of the normal 1961 to 1990 period was smaller than 40% in Reliquias and Granada and smaller than 60% in Barcelona (García-Herrera et al., 2006). This lead to further consequences for the subsequent year as the Spanish water reservoirs could not recover and reached only 47.6% of the total fill level in January 2006 (Spanish Ministry of the Environment).
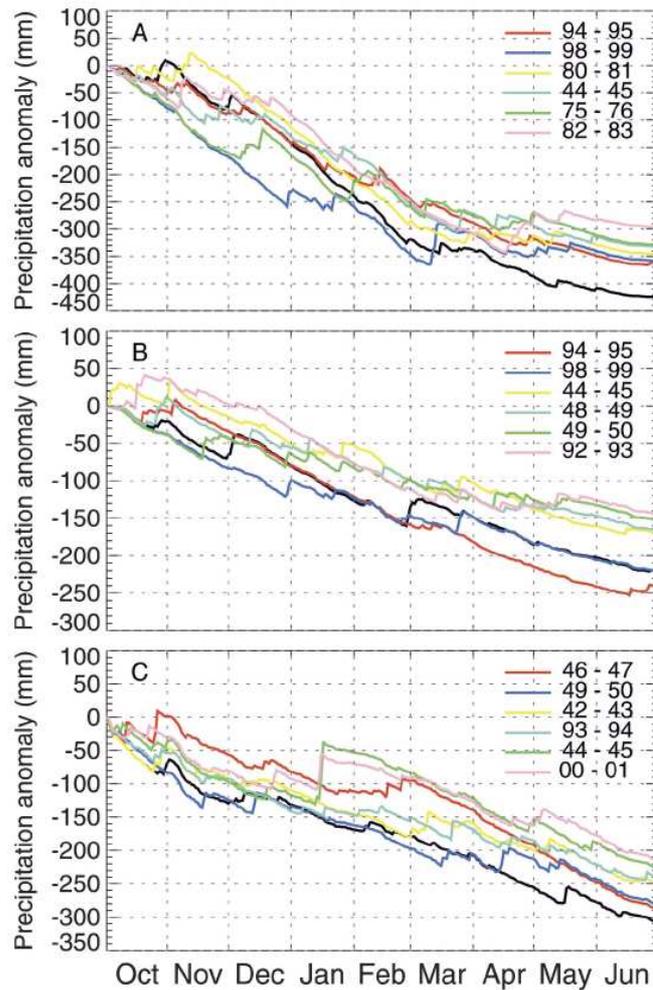
Figure 1.3: Development of major droughts in Spain recorded at a) Reliquias, b) Granada and c) Barcelona. Shown are daily accumulated precipitation anomalies in respect to climatology (1961-90). The black lines give anomalies from October 2004 to June 2005 (Figure from García-Herrera et al. (2006)).

Accumulated precipitation deficits are the main physical process for drought development. The accumulation time span is subject to change, leading to an integrated timescale of droughts. With this example the timescale of severe events is further extended up to several months.

1.1.2   Definition of extremes

So far, severe meteorological events have been described and their importance has been demonstrated, but a definition of what has to be considered as an extreme event is missing. The common property of all presented examples was

the unusualness of the events. Unusualness can be quantified through the concept of probability.

Thinking of an observed time series, an extreme event is surely the highest value that occurred during the observational period. This is why extreme value theory starts with considering the maximum, M_n of an independent and identical distributed (iid) random sample, $X_1, X_2, \ldots, X_n$:

$$M_n = \max\{X_1, X_2, \ldots, X_n\} \tag{1.1}$$

It is obvious that, the larger the observation time period (the sample size, n) the higher maximum values are expected. This extends naturally to the question whether a distribution function exists for M_n by taking the limit $n \to \infty$. The answer is given by the Fisher-Tippett Theorem (also know as Extremal Types Theorem; Fisher and Tippett (1928)), which demonstrates that convergence of a limit process with suitable normalized maxima, z, against a distribution function $H(z)$ is achieved and $H(z)$ belongs to one of the following distribution families (details given in Coles (2003)):

- Type I: Gumbel distribution ($\xi = 0$)

- Type II: Fréchet distribution ($\xi > 0$)

- Type III: Weibull distribution[2] ($\xi < 0$)

Note, that these are the only possible limits. The three types can be unified into one distribution function, the generalized extreme value distribution (GEV), where the shape parameter, ξ, determines the type at hand.

Return level plots are the usual way to visualize GEV. The return levels (values) are plotted against the return period on logarithmic scale (time). This simplifies interpretation, because the expected level can be deduced, which is reached or exceeded during given time period. Further, it guides the decision for the extremal type at hand. Type Gumbel shows up as a straight line and type Fréchet (Weibull) as a concave (convex) curve (Figure 1.4).

Extremes belong to the upper tail of the distribution functions and the upper tail behavior is completely determined by the GEV. This holds for a large class of distributions, because the condition required for the random sample is only the iid assumption. Extensions exist for serial or spatially dependant variables. Making extreme value theory to the method of choice for extremes in meteorological applications.

1.1.3   Simulation example

The advantage of extreme value statistics is demonstrated and it is shown that simple criteria may give wrong findings for the extremal behavior. Two artificial

---

[2]In the extreme value context it is the reversed Weibull distribution.
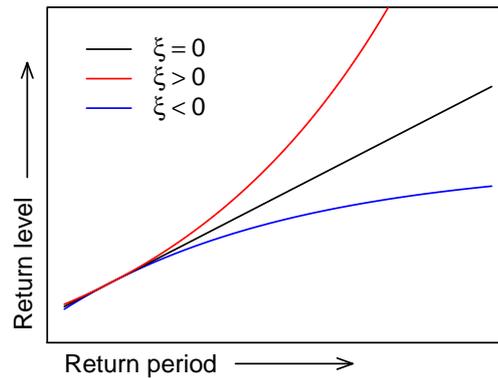
Figure 1.4: Return level plot illustrating the three different extremal types: Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$).

data sets with known outcome are created. The aim is to show differences in the extremes of two samples. In this way the example is an analog for climate change experiments, where differences between present and future climate states are of interest.

The random samples are Weibull distributed. This is a realistic example, because observed wind speeds are often thought of as following a Weibull distribution. The parameters of the distributions are given as follows: for sample one (sample two) the shape parameter is set to 0.6 (0.9) and the scale parameter is equal to 1.5 (5). The sample size is chosen to correspond to 150 years of daily observational data. This is sufficiently large ($n > 50000$) to minimize sample size effects.

Some typical measures, sample mean and standard deviation, as well as higher quantiles are calculated for the two samples. Quantities like these are often chosen to characterize or to draw conclusions about the extremes. The widely used extreme indices[3] use either absolute or quantile based thresholds and therefrom derived quantities, like for example the number of tropical nights, warm spell duration index, precipitation due to very wet days and many more. Comparing the two samples yields higher values throughout for sample two (Table 1.1). This indicates that higher values are more probable and that extreme values occur more often in sample two. By contrast, the sample maximum is lower in sample two, hinting at problems with the above interpretation (Table 1.1). However, the maximum is just a single value, so one is tempted to attribute this discrepancy to sample size effects.

The previous finding is contradicted by the estimated GEV. The resulting return level plot yields higher return values for return periods above the intersection of the two lines (around 1 year) for sample one (Figure 1.5 a)). The difference

---

[3]For example, the set of descriptive indices of extremes defined by the "Expert Team on Climate Change Detection and Indices" (Klein Tank et al., 2009)

	$\hat{\mu}$	$\hat{\sigma}$	$Q95$	$Q99$	max
sample one (blue)	2.3	4.0	9.3	19.1	96.5
sample two (red)	5.3	5.8	16.9	27.3	66.7

Table 1.1: Estimated mean ($\hat{\mu}$), standard deviation ($\hat{\sigma}$), 95% quantile ($Q95$), 99% quantile ($Q99$) and sample maximum (max) for the example of two Weibull distributed random samples.

between the two curves enlarges for increasing return periods. Therefore the extreme values are lower in sample two, in contrast to the previously used measures. As the example uses simulated data, that is random values from specified Weibull distributions, the outcome is known. The true densities cross near a value of 60, attributing higher probabilities for values above the intersection for sample one (Figure 1.5 b)). This confirms that sample two extremes are reduced, as achieved by the estimated GEV, where even the intersection occurs almost at the same value.

The discrepancy between the two outcomes results therefrom, that the quantities in Table 1.1 do not represent the upper tail and are mainly determined from the range below the intersection. There and below reside most values (according to Figure 1.5 b)), leading to the higher sample two quantiles. To conclude, for the example given higher quantiles have to be chosen to capture the upper tail and to

(a) Return level plot
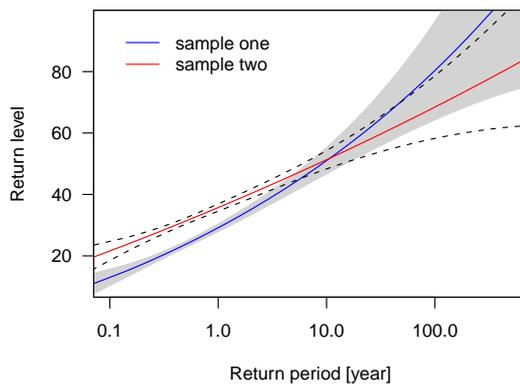
(b) True density
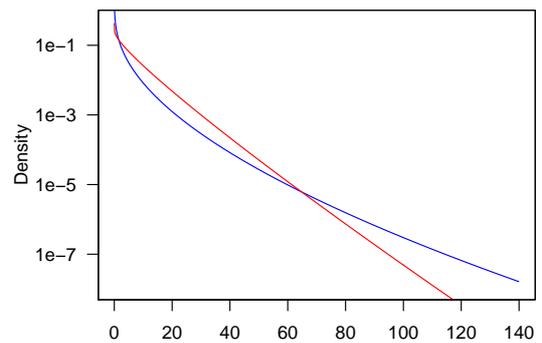
Figure 1.5: Example with two Weibull distributed random samples, with specified Weibull parameters. Shown are: a) return level plot for estimated extreme value distributions, together with 95% confidence intervals (gray shaded and dashed lines) and b) true density functions with logarithmic vertical axis for sample one (blue lines) and sample two (red lines).

reproduce the correct extremal behavior. The example is also a contribution to the discussion about the impact of changing mean and variance on the extremes. Comparing these different factors it was found that a change of both can lead to increased extremes, with the variance even more important for the upper tails (Katz and Brown, 1992). However, here increasing mean and variance comes along with reduced extremes. The reason is the effect of the shape parameter determining the Weibull distribution. This constrain was already pointed out by Katz and Brown (1992) and described more precisely by the hazard function.

In summarizing, simple measures, like indices of extremes, are not recommended for extreme value analysis. The example demonstrates, that they give neither sufficient nor necessary conditions for the extremal behavior. However, simple measures can be meaningful to describe moderate extremes. An example is the density increase below the intersection in the simulation study. They gain further importance in cases where extreme value statistics is not applicable (see later discussion in Chapter 1.2).

1.1.4   Outliers and extreme values

Meteorological time series are sometimes outlier adjusted. For example:without any additional knowledge about the December 1999 rainfall event in Venezuela (Chapter 1.1.1), one is tempted to consider the unusual high value as an outlier. For outlier detection bulk formulas or statistical tests are used, like for example a multiple of the standard deviation, measures based on the interquartile range or Grubbs' test. While not explicitly shown, one can expect that most of the criteria or tests will interpret this event as an outlier. In this example however, one surely knows that the observed value is correct and not the effect of some kind of measurement error. Thus the Venezuela rainfall event deals also as an warning not to falsely interpret unusual high values as outliers.

Long ago a critical assessment on outlier removal was given by R. A. Fisher, who stated: *"... the rejection of observations is too crude to be defended: and unless there are other reasons for rejection than the mere divergences from the majority, it would be more philosophical to accept these extremes, not as gross errors, but as indication that the distribution of errors is not normal"* (Fisher, 1922), pointing out the importance of the distributional assumption. In fact most outlier criteria and tests are based on the assumption of normal distributed values.

Returning to the Venezuelan time series the outstanding high event is also a challenge for extreme value analysis. The problem of attributing a realistic probability was addressed in several publications (Coles and Pericchi, 2003; Coles et al., 2003). The extremes were analyzed, using different extreme value models and estimation technics. Without going into details here, the return level curve of the seasonal model is coming close to the empirical estimated return period of the 12 December event (Figure 1.6). All models were estimated predictive, that
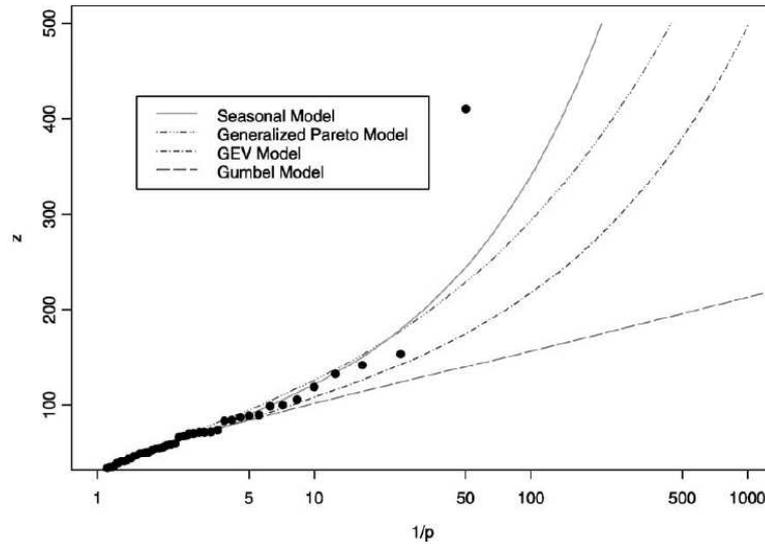
Figure 1.6: Return level plot for different predictive extreme value models fitted to daily precipitation observed at Maiquetia international airport, Venezuela (Figure from Coles et al. (2003); return level z in mm/dy and return period $1/p$ in years).

is the highest value is excluded in the estimation process. The seasonal model leads to a return period of about 130 years and the calculated probability that an event as large as $410mm/dy$ will occur in a 49 years period is 30% (Coles et al., 2003). This demonstrates the ability to account for realistic probabilities for such an unusual high event, if extreme value statistics is correctly applied and contradicts any speculations on an outlier.

1.2 Peak values and long lasting events - time series aspects

As shown by the presented examples, the spectrum of severe meteorological events is large. This does not only result from different meteorological quantities. Most striking is the involved time structure, ranging from single peak values up to long lasting events. An important characteristic for events on longer time scales are aggregation effects, as the drought example demonstrated. In this context single values and even single extreme values are not meaningful. It is the accumulation of many negative or positive deviations over a certain time period that matters and constitutes the extreme event.

On the other hand, the framework given by extreme value statistics is the method of choice for the analysis of extremes. If peak values are of interest

standard methods, like block-maxima or the peaks over threshold approach can
be applied to calculate extreme value distributions. Care has to be taken in
the case of meteorological time series, as many of them are serial dependant (or
exhibit memory) with nonvanishing autocorrelation for higher lags. One source
therefore are often the thermal inertia of oceans and land surfaces. However, this
property violates the basic iid requirement. To overcome this, extensions were
developed and modifications exists, so that standard extreme value statistics for
iid data is still applicable (overview and discussion in Dress (2008)):

- Clusters of exceedances: detecting independent clusters of exceedances and
  using the cluster maxima for further analysis

- Nonparametric: applying standard extreme value statistics to all exceedances,
  together with adequate adjustment of confidence intervals

- Semiparametric: using the residuals of a previously fitted time series model

Each of the methods has their pros and cons depending on the situation at
hand. If clusters could be physically defined or the autocorrelation function is
rapidly vanishing for higher lags, the cluster of exceedances approach is sufficient.

From another point of view, serial dependance is a necessary condition for ac-
cumulation processes. Because the probability that a positive (negative) anomaly
follows a positive (negative) one is increased. In contrast to iid samples, where
positive and negative anomalies in the future time step are equal likely indepen-
dent of the past. If the total effect of successive high events is of interest the time
series needs preprocessing. This might be done in a way analogous to the cluster
approach. But, instead of cluster maxima, cluster sums can be used for extreme
value analysis. Although, this is appropriate on short time scales, constrains oc-
cur for processes developing on long time periods. Sources for long lasting events
are often associated with non stationary behavior. Examples include persistent
weather regimes, the El Niño/Southern Oscillation phenomenon or variability up
to decades and more (like the meridional overturning circulation in the North
Atlantic). For events lasting a month or longer the sample size is drastically
reduced, if the data is preprocessed in the same manner. So, even though ex-
treme value statistics is the preferable method for the analysis of extremes, it is
impractical for events on long time scales.

1.3   Outline

The problems mentioned above make it necessary to use different methods, ad-
equate for the problem at hand. The standardized precipitation index (SPI) is
subject of Chapter 2 and 3. The SPI is related to accumulated deficit precipi-
tation and can be constructed for different time scales (McKee et al., 1993). In

Chapter 2 a local study is accomplished, relating Icelandic SPI dry and wet extremes to a distinct dipole pressure field. It is further demonstrated, how the SPI can be used for climate change assessment. The basis for the SPI construction is the probability distribution for precipitation. The Gamma distribution is the usual standard and is reconsidered in Chapter 3. This is done with a comparative analysis, using other admissible distributions. In Chapter 4 North Atlantic cyclone extremes and their potential change under greenhouse gas warming are analyzed. Extreme value statistics is applied, focusing on return level and parameter changes. Another way of analyzing extremes is to look at the time distances of successive threshold crossings. Extreme event return time distributions and their memory properties in long-term memory processes are analyzed in Chapter 5.

2 Extreme dry and wet events in Iceland: Observations, simulations and scenarios

Monthly extremes of dryness and wetness in Iceland are analyzed based on the standardized precipitation index (SPI). The analysis is performed for observations and four sets of coupled atmosphere-ocean climate model simulations (ECHAM5/MPI-OM) to link water cycle extremes in Iceland with regional atmospheric flow patterns and to estimate and evaluate future changes. The following results are obtained: (i) SPI extremes are linked with a Europe-Greenland Index (EGI) describing south-westerly flow anomalies by a dipole and the related geopotential height differences. The good agreement between the observed statistics and transient 20th century simulations encourages analysis of future climate projections. (ii) Comparison of the 21st century A1B-scenario with the pre-industrial climate reveals significant and large differences: While extremes of dryness hardly change, extremely wet conditions increase in winter and spring. As there is no flow intensification and cyclone density decreases, the cause maybe found air moisture raising in a warmer climate.

2.1 Introduction

Iceland's climate and climate variability are commonly associated with storm-tracks and cyclone path regimes (Schneidereit et al. (2007)) depending on teleconnections affecting the North Atlantic sector like, for example, the North Atlantic Oscillation (NAO) or El Nino - Southern Oscillation (ENSO); for comprehensive reviews see SFB-512 (2005) and Fraedrich (1994). Here, however, the indicator of climate variability to be analyzed is not the meridional surface pressure dipole between Iceland and the Azores, but extremes of the water cycle in Iceland located at the northern NAO-pole. Water cycle extremes affect land (and sea) and are characterized by periods of dryness and wetness, both of which occur in arid and in humid climates. Notwithstanding the considerable ecological and economical impact, the corresponding atmospheric flow patterns are also of interest, in par-

ticular, when climate simulations are to be verified and future climate scenarios need to be evaluated.

Extremes are rare by definition and difficult to estimate. Different methods can be used to describe extremes. Nonparametric methods are based on percentiles or quantiles and are often used in meteorological science for the calculation of extreme indices (for example: Klein Tank and Können (2003)). Parametric methods, however, enable calculations of extreme value distributions (Coles (2003)). In the following the standardized precipitation index (SPI) is applied, which has been proposed to obtain a quantitive meaning for an otherwise loose definition (Bordi et al. (2006)), in order to monitor dryness and wetness (McKee et al. (1993)). Although for the index calculation parametric methods are used, the SPI is a nonparametric method, because estimated distributions are only used for the standardization. Therefore, the SPI may be regarded as an extension of nonparametric methods.

The goals of this chapter are twofold: (a) linking extreme dryness and wetness with atmospheric flow patterns and (b) estimating future changes in the occurrence of the extremes. To achieve these aims, our analysis is structured as follows: The SPI and the observed and simulated data sets to be analyzed are introduced (Section 2.2 and 2.3). Observed dryness and wetness as extremes of the SPI and the associated observed large scale atmospheric flow are identified and compared with transient 20th century model simulation (Section 2.4). Based on this comparison dryness and wetness (and the related atmospheric flow) statistics are evaluated for future climate scenarios (Section 2.5). A brief summary concludes the analyses (Section 2.6). In addition, the relation between SPI extremes to cyclone track densities are investigated in the Appendix.

2.2   Data and model simulations

Iceland precipitation is represented by single time series in terms of monthly means of area (about 10^5km^2) averages to make analyses of higher resolution observations and low resolution simulations comparable.

Observations

Observed Iceland precipitation values are taken from a precipitation climatology at 0.5° resolution (VASCLIMO, Variability Analysis of Surface Climate Observations, Beck et al. (2005)). The VASCLIMO data set is very similar to the CRU TS 2.1 data set (CRU, Climate Research Unit, Mitchell and Jones (2005)): for area averaged Island precipitation on a monthly basis, time series show nearly identical distributional properties with the same variability (correlations larger than 0.9). The observed atmospheric flow is presented by the $500hPa$ geopotential height analysis (ERA40, Simmons and Gibson (2000)). Both precip-

itation and geopotential height data sets are analyzed for the overlapping period
1958 to 2000.

Simulations

The coupled atmosphere-ocean climate model is ECHAM5/MPI-OM (Roeck-
ner et al. (2003), Marsland et al. (2003)); the atmosphere is in T63 spectral
resolution (about 1.8°) with 31 vertical levels. Thus Iceland is represented by 6
grid boxes. No flux correction is applied. The following model simulations are
analyzed:

1. Pre-industrial control experiment (CTL) with constant greenhouse gas con-
   centrations as observed in 1860 (500 years integration).

2. Transient 20th century simulation (1860 to 2000) to compare with observa-
   tions (three ensemble members).

3. Stabilization run (20C) with greenhouse gases fixed at present day climate
   of the year 2000 (three 100 year ensemble members).

4. Stabilization run (A1B) with greenhouse gases fixed at 2100 following the
   A1B scenario (Nakićenović et al. (2000)), which represents the climate af-
   ter an intermediate greenhouse gas increase (three 100 year ensemble mem-
   bers).

Trends

Precipitation trends from 1951 to 2000 are determined for each calendar
month for both observations and the 20th century simulation. Although there is
evidence of local trends in observed Iceland precipitation (Beck et al. (2005)), the
area averaging removes them. Only two months show a significant trend (95%
confidence level, according to Mann-Kendall-Test). Trends are not removed in
the following, because calculations with and without them did not change the
results.

2.3   Standardized Precipitation Index (SPI)

The SPI was introduced by McKee et al. (1993) to classify and monitor dry-
ness and wetness. The calculation of the SPI is based on an "equal probability
transformation": monthly precipitation is transformed to a standard normal dis-
tribution to yield SPI values by preserving probabilities (details are given in the
Appendix of Bordi and Sutera (2001)). The standardization ensures, that the SPI
gives a uniform measure for dryness and wetness in different climate regimes or

under seasonal dependence. The SPI can be constructed for different timescales characterizing meteorological, hydrological and agricultural periods of dryness and wetness. Here we choose the monthly time scale characterizing the meteorologicaly relevant period, that is, wet and dry spells and its extremes. The SPI classification is shown in Table 2.1.

Distribution

The transformation depends on the assumed statistical distribution of monthly precipitation. A false distribution type leads to systematic errors, which are most severe at the upper and lower tails and, therefore, for the extreme values. The gamma distribution, as used by Bordi et al. (2006) to describe precipitation in Sicily, has lead to adequate results. However, this distribution does not hold for all months for the Iceland precipitation time series; for some months better fits can be achieved by the Weibull distribution. To still use a single unifying distribution type the "Generalized Gamma Distribution" is applied instead:

$$f(x) = \frac{d}{\Gamma(k)} b^{-dk} x^{(dk-1)} \exp\left(-\left(\frac{x}{b}\right)^d\right) \tag{2.1}$$

with scale parameter b, two shape parameters d and k, and the gamma function $\Gamma(y)$. This version includes as special cases the gamma distribution (by setting $d = 1$) and the Weibull distribution (with $k = 1$). The two shape parameters make maximum likelihood estimation difficult and lead to convergence problems. Therefore, parameter estimation is performed by using the reparameterized version of the generalized gamma distribution (Lawless (1982)).

Changing climate

The transformation can also be used to identify possible dryness and wetness changes in terms of SPI values. The solid lines and arrows illustrate the trans-

SPI intervals	SPI classes	P
SPI ≥ 2	W3: extremely wet	2.3
$2 >$ SPI ≥ 1.5	W2: severely wet	4.4
$1.5 >$ SPI ≥ 1	W1: moderately wet	9.2
$1 >$ SPI > -1	W0, D0: normal	68.2
$-1 \geq$ SPI > -1.5	D1: moderately dry	9.2
$-1.5 \geq$ SPI > -2	D2: severely dry	4.4
SPI ≤ -2	D3: extremely dry	2.3

Table 2.1: Standardized Precipitation Index (SPI) classes and corresponding event probability, P in %.

formation to SPI (Figure 2.1). For an example value $x = 20$, from a sample (X), which is gamma (Γ_1) distributed (solid curve), the transformation leads to a SPI value of approximately 1. Transforming the whole sample, X, results in standard normal distributed SPI values, with preserved probabilities, but only if the samples are indeed drawn from a Γ_1-distribution. The same transformation of a $y = x$ from another sample (Y), which is Γ_2 distributed, will lead to the same SPI value. But the "preserved" probability is now the probability of the Γ_2 distribution. Transforming the whole sample Y will not give the standard normal distribution. The probability difference between x from Γ_1 and y from Γ_2 occurs also after the transformation (difference between the two horizontal arrows). So, possible future precipitation changes are also detected and expressed in terms of the SPI by this transformation property.

2.4 SPI extremes and atmospheric flow: present day climate

For 1958 to 2000 the observed monthly extreme SPI classes for Iceland are associated with monthly mean northern hemisphere $500hPa$ geopotential height anomaly composites to identify atmospheric flow fields related to extreme dryness and wetness. This provides the background for comparison with climate simulations.
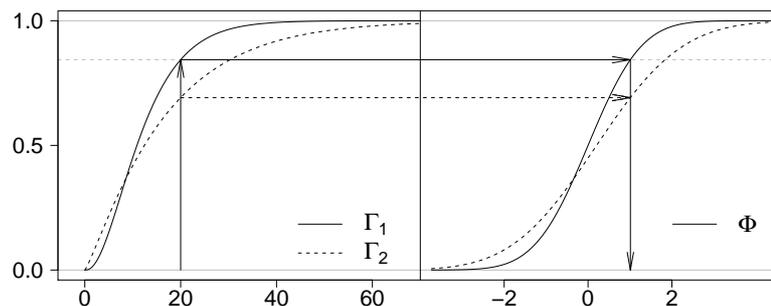
Figure 2.1: The transformation to SPI is illustrated by solid lines and arrows: a gamma distribution (Γ_1, left) is transformed to the standard normal distribution (Φ, right). Another gamma distribution (Γ_2, dashed lines, left) is transformed such that the probability differences remain the same on the both sides. Note that the resulting distribution (dashed lines, right) does not have the properties of the standard normal distribution (shown are cumulative distribution functions).

2.4.1   Observations

SPI

In a first step, precipitation is transformed to SPI. The time period is short and the extreme events are rare. According to Table 2.1 one expects approximately 11 extreme wet and 11 dry months in 43 years. Extreme and severe SPI classes are combined to increase the sample size. Then, composite maps of the $500hPa$ geopotential height anomalies (of extreme wet and dry months) are constructed as averages over the deviations from the mean over all months of the remaining SPI classes.

Circulation

In Figure 2.2 the resulting geopotential height anomalies are composited for severe and extreme wet (a) and dry (b) conditions and are averaged for the whole year. In the severe and extreme wet case a negative anomaly center is found near Greenland and a positive anomaly center near Europe between England and Scandinavia, which are steering centers of the flow. The signs of the anomaly centers are reversed in the case of severe and extreme dry conditions, but their locations remain the same. These patterns are nearly consistent throughout
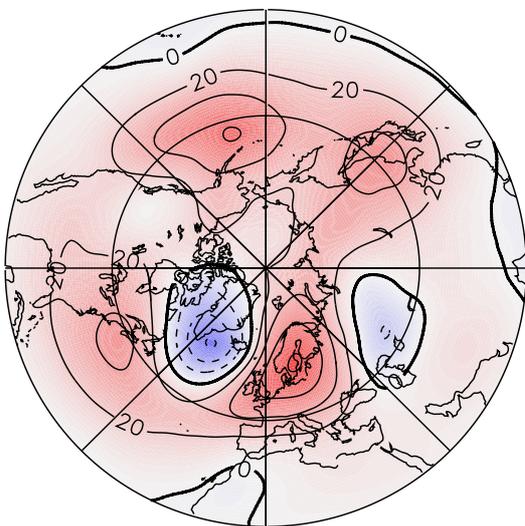
(a) Severe and extreme wet conditions   **(b)** Severe and extreme dry conditions
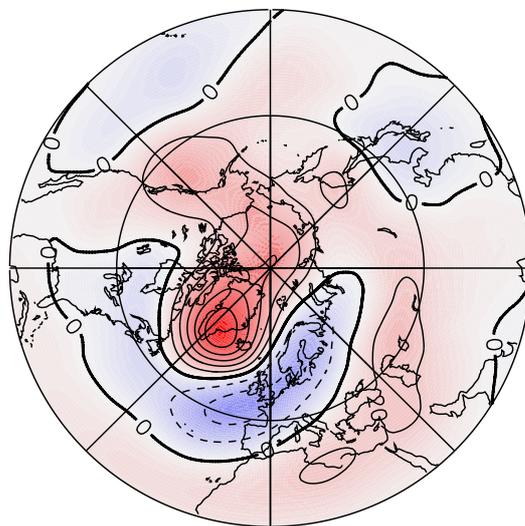
Figure 2.2: Composite maps of ERA40 geopotential height anomalies $[gpm]$ corresponding to observed severe and extreme a) wet (SPI ≥ 1.5) and b) dry (SPI ≤ -1.5) conditions in Iceland. Continuous lines show positive and dashed lines negative anomalies. The contour intervals are $20gpm$.

the seasons, with slightly varying magnitude and location of anomaly centers (not shown). The anomalies are highest during autumn and winter, in spring and summer the anomalies are weaker. Around Greenland no deviations of the location occur throughout the seasons, whereas in the wet cases the positive anomaly center is shifted northward in spring. In the dry cases the negative anomaly center is split in two poles in autumn.

Cassou et al. (2004) found four climate regimes in the winter $500hPa$ geopotential height field by cluster analysis, capturing nonlinearities of North-Atlantic weather regimes. One of the regimes (the west-east dipole between Greenland and Scandinavia, their Figure 3 d)) is similar to the wet extreme anomaly field (Figure 2.2 a)). However, the dry extreme anomaly field (Figure 2.2 b)) has no clear equivalent: for NAO negative (Cassou et al. (2004) with Figure 3 a)), the positive poles over South-Greenland coincide, while the negative pole (Figure 2.2 b)) is displaced north-eastward. That is, extremes can partly be associated with weather regimes, associated with cluster analysis. Therefore, NAO relations with Iceland precipitation (Hanna et al. (2004)) and, possibly, cyclone activity (Serreze et al. (1997)) could be extended in view of these results for both present day and changing climates.

In summarizing, the anomalous flow patterns associated with severe and extreme wet SPI classes (Figure 2.2 a)) are related to an anomalous northward atmospheric flow (of warm and wet air masses) from the Atlantic region. This flow is reversed in severe and extreme dry conditions, where Iceland is under the influence of dry polar air masses. Thus, a Europe-Greenland geopotential height difference is suggested as a flow index, EGI (used in Section 2.5), representing the pressure gradient, which characterizes the anomalous circulation.

2.4.2   Climate model: transient simulation

Verifying climate models by comparing observations with the transient 20th century simulations (1860 to 2000) is prerequisite for evaluating the representativeness of future scenarios. Thus, the respective ensemble simulations are subjected to the same SPI and circulation pattern analysis. It should be noted that the ensemble members are not simulations of the present climate, but of scenarios starting from the pre-industrial climate forced with observed greenhouse gases until the year 2000.

SPI

Since SPI is calculated on a monthly basis, the comparison of the observed and model simulated precipitation is made for each calender month. A two-sample Kolmogorov-Smirnov test is performed testing the null hypothesis that precipitation data are drawn from the same continuous distribution (applied to each individual ensemble member). Test results are shown in Table 2.2 in terms of

Month	1	2	3
1	**0.717**	**0.272**	**0.179**
2	**0.998**	**0.549**	**0.396**
3	**0.179**	0.001	**0.068**
4	0	0.006	0.039
5	**0.717**	0.022	**0.272**
6	**0.396**	**0.272**	**0.549**
7	**0.549**	**0.179**	**0.717**
8	**0.869**	**0.396**	**0.272**
9	0.039	0.001	**0.068**
10	**0.967**	**0.869**	**0.967**
11	0.039	**0.396**	**0.272**
12	0.006	**0.272**	0

Table 2.2: Two sample Kolmogorov-Smirnov test: estimated p-values are shown for each ensemble member (1 to 3). The null hypothesis is rejected at 5% level for $p < 0.05$.

probability, p, of null hypothesis rejection: In six months (January February, June to August, October) the null hypothesis is accepted for each member; in three months (March, May and November) there is one ensemble member for which the null hypothesis is rejected, so that there is still some agreement between observation and model simulations. In September and December one ensemble member fulfills the test, while no agreement is found for April. In summarizing, no single season shows a systematic departure. Best agreement exists in summer and worst in spring, possibly due to ENSO influencing the flow and cyclone track patterns (Fraedrich (1994)). Though not perfect, the agreement between observed and model simulated precipitation can be considered as good.

Circulation

Now the flow fields associated with severe and extreme SPI classes are determined from model simulations to analyze and compare wetness and dryness related circulation patterns. All ensemble members are analyzed together and the results are presented in Figure 2.3, where significant anomaly differences are shaded according to a T-test (95% confidence). Nearly the same anomaly patterns are found in observations and the transient simulations of the present day climate (Figures 2.2 and 2.3). Comparison shows the following results: For severe and extreme wetness the model area of significant positive geopotential height anomalies is shifted southward with the European center near England while negative anomalies show the same strength and location. For severe and extreme dryness
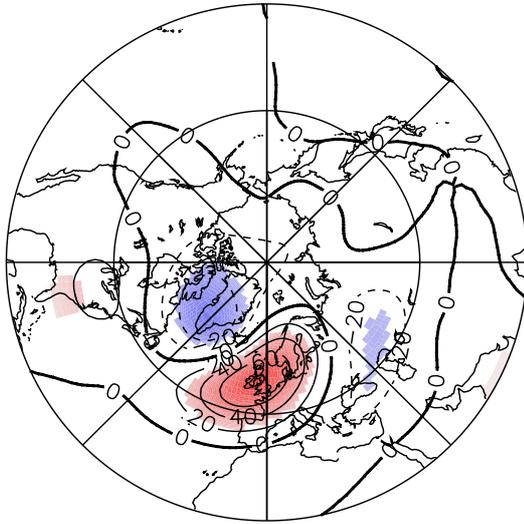
(a) Severe and extreme wet conditions   **(b)** Severe and extreme dry conditions
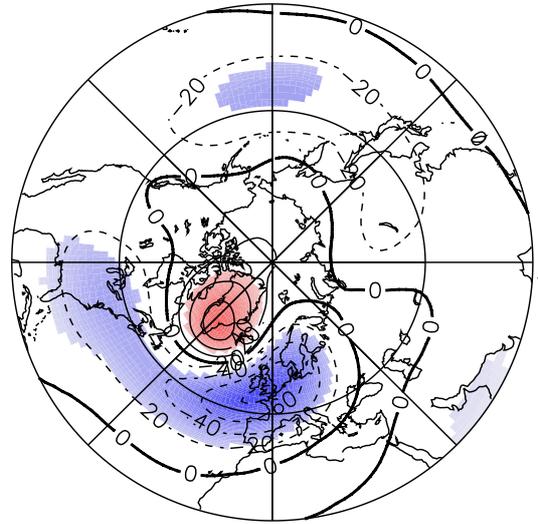
Figure 2.3: Composite maps of geopotential height anomalies ($[gpm]$) of the transient 20th century simulation corresponding to observed severe and extreme a) wet (SPI ≥ 1.5) and b) dry (SPI ≤ -1.5) conditions in Iceland. Contour lines see Figure 2.2, significant areas shaded (95% confidence level).

composites show larger deviations: positive anomalies are much stronger in the observed geopotential heights, while negative ones reach higher values in simulations. The strengths of the Greenland and Europe anomalies are asymmetric in the observations and almost symmetric in the simulations, while the centers' locations remain the same (Figure 2.3). Some of these differences may result from the short time period of the observations and, therefore, represent random fluctuations, or from model bias underestimating the tropical influence on mid-latitude weather and climate. However, the time limitation in the data series does not influence the present analysis. Only if differences were much larger than found, the model bias would have reached a magnitude that credible results of the projected future may affected. In summarizing, the patterns of observed and simulated flow fields agree well and the statistical test highlights the important areas steering the flow. Furthermore, they are consistent with the previously given physical explanation.

2.5 Climate change scenarios

The ensemble members of two stabilization runs are analyzed representing present day (20C) and the scenario (A1B) climates and both sets are compared with the pre-industrial (CTL) climate.

SPI

The SPI classes (Figure 2.4) show almost no differences between present day and pre-industrial frequencies. However, the scenario climate compared with the pre-industrial shows a strong increase of severe and extreme wet conditions. The frequency of the severe wet class is more than doubled, while in the extreme wet class it is more than trebled. In all dry classes the frequencies are reduced, except for the extreme dry class, which remains unchanged. Differences in frequency between extreme SPI classes of model simulations are now evaluated on a seasonal basis (Figure 2.5). The present day climate shows only small changes in wet and dry extremes compared with the pre-industrial simulation. The scenario climate reveals small (large) changes for the extreme dry (wet) SPI classes; note that, in the scenario climate, the increase in the wet extremes is strongest in summer and autumn; a considerable increase occurs in winter, while it is lowest for spring. Small differences between the single ensemble members are found, which lie in the range of random variations. All members show the same seasonal response.

Circulation

The SPI extremes in both pre-industrial and scenario simulations are also associated with the previously found atmospheric circulation anomalies and their steering centers (not shown). Relating the frequency increase of extreme wet SPI classes to anomalous flow patterns is analyzed by the joint density distribution of the respective indices, SPI and EGI (introduced in Sections 2.3 and 2.4.1). Sig-
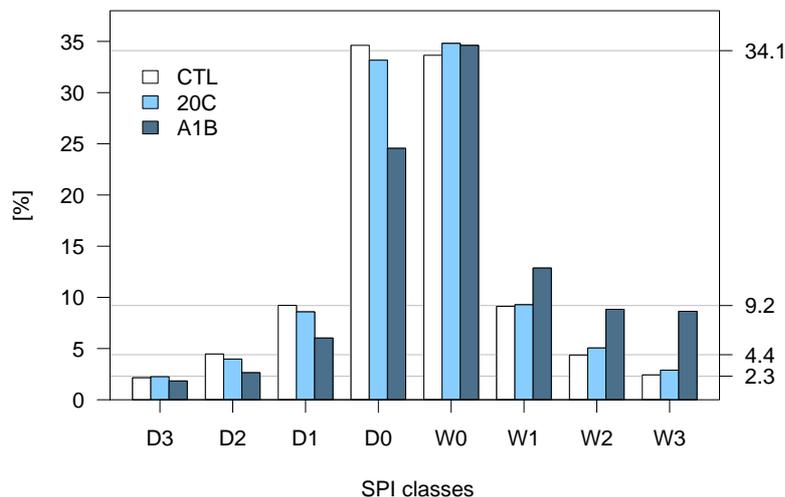
Figure 2.4: Frequencies of SPI classes (see Table 2.1): pre-industrial (CTL), present day (20C) and scenario climate (A1B); SPI event probabilities in % on right vertical axis.
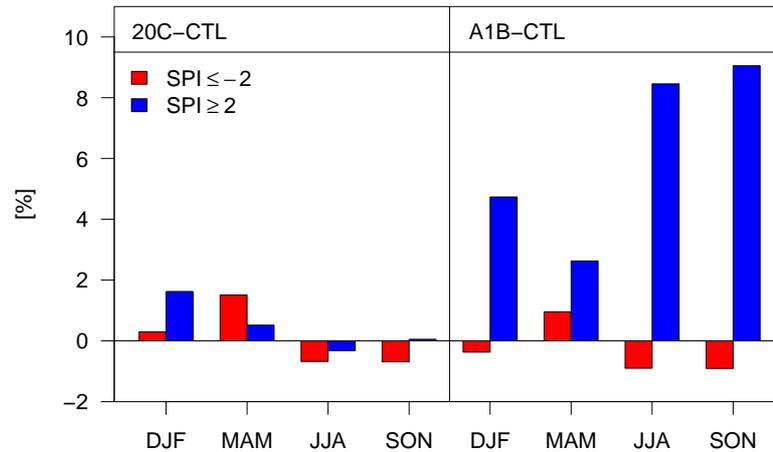
Figure 2.5: Differences of extreme SPI class frequencies (per season): present day (20C) minus pre-industrial (CTL) climates (left), scenario (A1B) minus pre-industrial (CTL) climates (right).

nificant positive and negative geopotential height anomalies define the steering centers of both wet or dry patterns (see Figure 2.3). The "wet/dry" intersections yield a dipole, whose area averaged geopotential height anomalies define the Europe-Greenland Index (EGI), that is, a difference of area averaged geopotentials between the Europe and Greenland "wet/dry" intersections. Extreme wet patterns correspond to positive EGI values with enhanced southwesterly flow and extreme dry conditions in Iceland to negative ones, with reduced southwesterly flow. This Europe-Greenland circulation index, EGI, is now related to the Iceland SPI for each season (Figure 2.6). The following results are noted:

EGI-SPI correlations

In the pre-industrial (observed) climate SPI and EGI are highly correlated and the correlations range from 0.64 (0.6) in summer to 0.72 (0.69) in winter. This shows that the EGI is not only relevant for the extreme but also for the other SPI classes. The correlations remain almost unaltered in the scenario climate.

EGI-SPI joint density (pre-industrial)

The joint density estimates also show that high (low) SPI values are associated with high (low) EGI values (Figure 2.6, left column). EGI values show larger variability in winter than in summer (see also discussion in Section 2.2), as it is obvious from density tails. In summarizing, these statistics support the introduction of EGI representing the atmospheric flow relation with SPI and its extremes (Section 2.4.1).
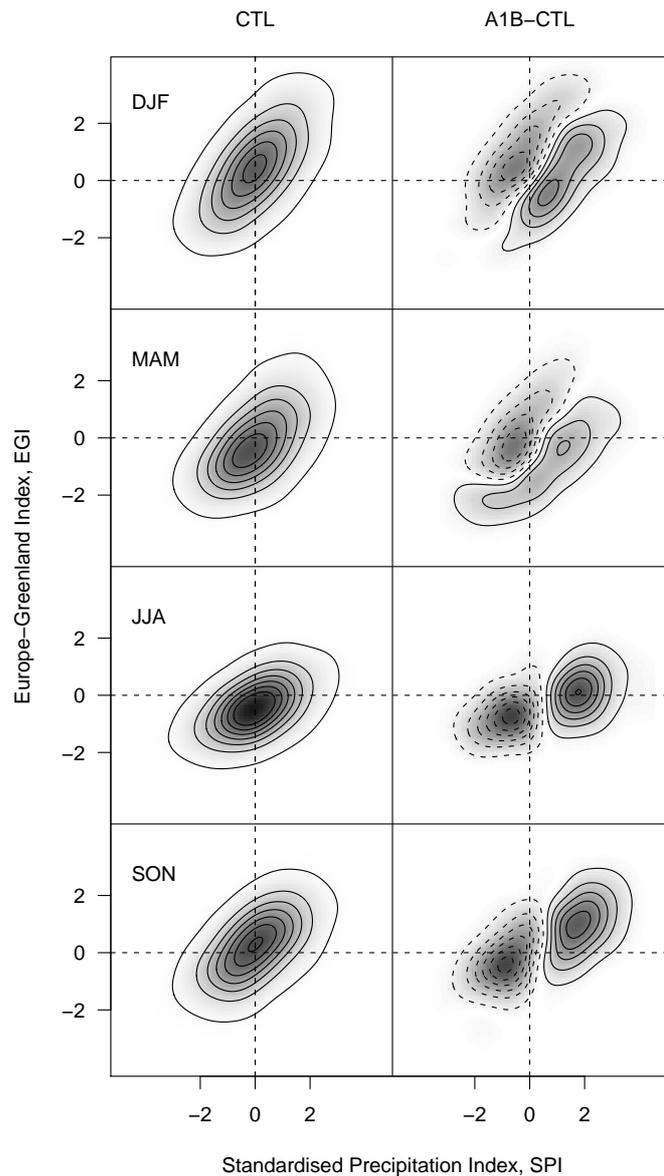
Figure 2.6: Seasonal joint density estimates of standardized precipitation index (SPI axis) and Europe-Greenland geopotential height difference (EGI axis in [100gpm]): pre-industrial climate (CTL, left) and density differences between scenario and pre-industrial climate (A1B-CTL, right). Density contours start from 0.005 with contour intervals 0.025; negative (positive) density differences are dashed (solid) starting from -0.001 (0.001) with contour intervals 0.001. Higher values are shaded darker.

EGI-SPI joint density difference (A1B and CTL)

The two dimensional densities (EGI, SPI) are calculated for the scenario climate to determine the density differences between the scenario and pre-industrial

climate (Figure 2.6, right column). A shift along the SPI axis to higher values is obvious in all seasons. Note that the positive contour lines reach highest values (SPI ≈ 2) in summer and autumn, which is due to more extreme wet events (than in the other seasons, see also Figure 2.5). The shift along the EGI axis, however, is less pronounced. The change of the densities to higher EGI values is only found in summer and autumn, i.e. the wet pattern (Figures 2.2 a) or 2.3 a)) occurs more often and higher differences between the poles are getting more probable. This is reversed in winter and spring, where the density is reduced for high positive EGI values. In these seasons the occurrence of the dry pattern (Figures 2.2 b) or 2.3 b)) has an increased probability. Note, that all individual ensemble members show the same response, as in Figure 2.6, with nearly no differences.

In summarizing, we conclude that (i) the increase of extreme wet events (Figure 2.5) in winter and spring (of the A1B scenario climate) is *not* a result of an intensification of the anomaly flow pattern. That is, because the density of high EGI values decreases while the density of high SPI values increases. As the relation between EGI and SPI is still present in the scenario climate, the noted SPI increase in the scenario must result from another mechanism as, for example, moisture rising in a warmer climate. (ii) The seasonal variability of extreme wet SPI frequency changes (A1B-scenario minus CTL-control, Figure 2.5) are related to EGI (or flow pattern) density changes (A1B-CTL, 2.6). That is, the largest extreme wet SPI frequency increase occurs in summer and autumn, which corresponds to the increasing density of positive EGI or, to higher EGI values occurring more often (see anomaly pattern in Figure 2.3 a)). In addition, extreme wet SPI frequencies rising less in spring than in winter (Figure 2.5) is consistent with a stronger decrease of density of positive EGI.

2.6   Summary and conclusions

Extremes of dryness and wetness in Iceland are analyzed in terms of the standardized precipitation index (SPI). Observed geopotential height anomalies composited about extreme SPI events show a dipole like structure representing steering centers over Europe's west coast and Greenland's South. The Europe-Greenland Index (EGI) of geopotential height differences reverses sign from wet to dry events. The analysis of observations is compared with transient 20th century (ensemble) simulations followed by the same analysis of the future A1B-scenario. Some results are summarized:

1. Agreement between observed and simulated transient 20th century precipitation (SPI) statistics can be considered as good and the observed SPI-circulation relation is also found in the simulation.

2. No significant differences appear in the frequencies of the monthly SPI classes (per year) and monthly SPI extremes (per season) between present

day climate (20th century stabilization runs) and pre-industrial (CTL) sim-
ulation.

3. Significant and large differences occur in the frequencies of the severe and
   the extreme wet SPI classes between scenario (A1B) climate (21st cen-
   tury stabilization runs) and the pre-industrial (CTL) climate (stabilization
   runs).

4. Extreme wet SPI frequencies increase in winter and spring of the scenario
   climate. This is not a result of intensifying flow anomalies but may be
   related to other mechanisms as, for example, moisture rising in a warmer
   climate.

5. But, the seasonally differing responses in scenario frequencies (of extreme
   wet SPI classes) are consistent with the changing anomalous flow pattern.

These analyses add information to Iceland climate studies introducing
monthly precipitation extremes in terms of dryness and wetness and their pos-
sible link to an atmospheric circulation pattern steered by geopotential height
anomalies over Greenland and Europe. New are the methods introduced for an-
alyzing extremes under changing climate conditions and linking this with the
associated atmospheric flow pattern. However, to aid decision making based
on future scenarios requires test and analysis of other climate change scenarios
and different climate models. Extension to other regions governed by different
circulation regimes and climates will be presented in due course.

2.A Cyclone density analysis

This appendix presents results of cyclone tracking analysis applied to spells of
wetness and dryness in Iceland. The tracking algorithm is applied as described
in detail by Schneidereit et al. (2007) (see also references therein) with two ex-
ceptions: the minimum cyclone lifetime is 48 hours and their minimum travel
distance is 1000 km. The winter season is considered. Cyclone densities for
wet extremes (Figure 2.7 a)) show enhanced magnitude near $60°$ North with
a southwest-northeast orientation along the Greenland trough anomaly and its
largest gradient. The pattern is similar to the winter mean density. For dry
conditions (Figure 2.7 b)) cyclone activity is enhanced and more zonally oriented
along $50°$ North.

In addition, cyclone density change is presented for scenario (A1B) minus
transient 20th century simulation. For extreme wet conditions (SPI $\geq$ 1.5) the
density is strongly reduced in Norwegian-Barents Sea (up to -10%, Figure 2.7
c)), consistent with the reduced EGI (Figure 2.6, top right panel). That is, the
increasing SPI change in Iceland (Section 2.5) cannot be related to an increasing
number of cyclones.

(a) Transient 20th century: severe and extreme wet conditions
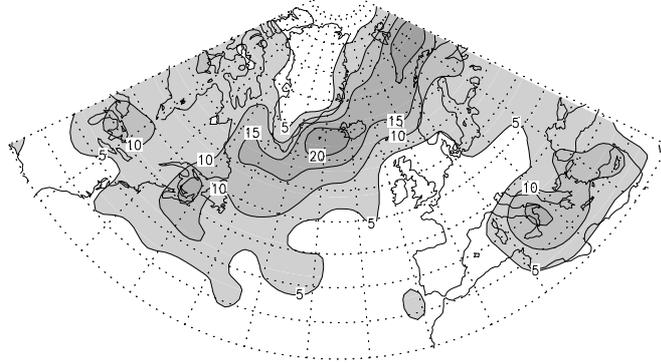
(b) Transient 20th century: severe and extreme dry conditions
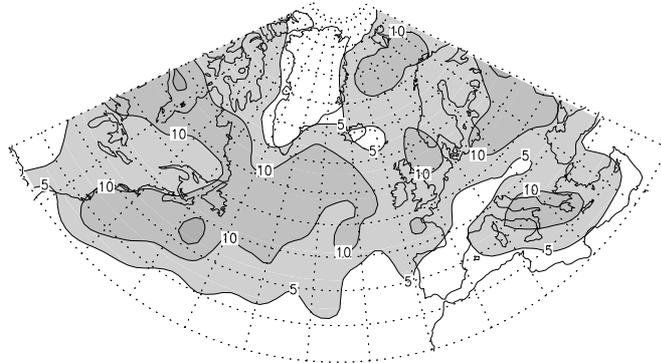
(c) Scenario minus transient 20th century: cyclone density change
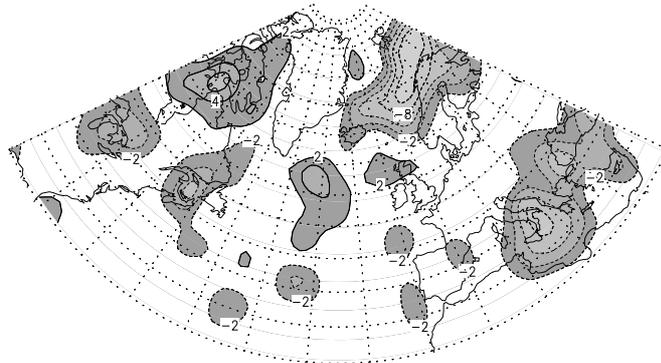
Figure 2.7: Cyclone densities of the transient 20th century simulation in winter (DJF, 1950- 2000) corresponding to observed severe and extreme a) wet (SPI ≥ 1.5) and b) dry (SPI ≤ -1.5) conditions in Iceland. The contour intervals are 5%. c) Mean cyclone density change between scenario (A1B) and transient 20th century simulation climate for severe and extreme wet conditions. Continuous lines show positive and dashed lines negative differences, with contour intervals of 2%.

3 Implications for drought monitoring and projection - SPI bias

The applicability of the Gamma distribution (GD) for monthly precipitation sums is reconsidered. The motivation for this study is the observation that the GD fails to represent precipitation in considerable areas of global observed and simulated data. This misinterpretation may lead to erroneous Standardized Precipitation Index (SPI), model evaluations and climate change assessments. In this study, the GD is compared to the Weibull (WD), Burr Type III (BD), exponentiated Weibull (EWD) and generalized Gamma (GGD) distribution. These distributions extend the GD in terms of possible shapes (skewness and kurtosis) and the behavior for large arguments. The comparison is based on the Akaike information criterion, which maximizes information entropy, and reveals a trade-off between deviation and the numbers of parameters used. We use monthly sums of precipitation given by England-Wales, CRU and ECHAM5 data sets for 12 individual months of the year. Weibull type distributions give distinctly improved fits, while the worst results are obtained for the GD. This is obtained by a global assessment of observed and simulated data throughout the whole year.

3.1 Introduction

The Standardized Precipitation Index (SPI) is widely applied for the description of extreme dryness or wetness. An increasing number of publications uses the SPI to diagnose observed precipitation deficits or excesses and analyze its variability. The SPI is further applied as a monitoring tool, which is able to give the actual state of meteorological, agricultural and hydrological droughts. The World Meteorological Organization (WMO) recommends the SPI to characterize meteorological droughts by all meteorological and hydrological services (WMO press release No. 872, December 2009). Newer applications use the SPI for diagnosing future drought occurrences in climate change scenarios (Sienz et al., 2007; Burke and Brown, 2008).

One reason for the wide appliance of the SPI is its simplicity compared to other drought indicators, such as the Palmer drought severity index (PDSI; Palmer (1965)). Only precipitation is needed as input quantity, contrary to the PDSI,

where in addition temperature and local available water content of soil are required. Although regarded as a simple measure, the SPI has its restrictions concerning the sample size and in arid environments. Wu et al. (2005) presented a critical assessment of sample size impacts. Furthermore, months without precipitation create a lower bound in the SPI, leading to problems with drought indication (Wu et al., 2007).

For the calculation of the SPI the probability distribution of precipitation is of importance. This has been demonstrated by Guttman (1999), who concluded that "the SPI should not be used widely until a single probability distribution is accepted as a standard". Guttman (1999) compared different distributions with a regional drought model and proposed the three parameter gamma distribution as standard. The Gamma distribution with either two or three parameters is now widely applied in hydrological and climatological science. However, several authors pointed out that the GD can lead to problems and does not fulfill goodness of fit criteria (Lloyd-Hughes and Saunders, 2002; Sienz et al., 2007).

Apart from applying the distribution of precipitation for SPI calculation, the distribution itself is of interest and there is a long history of applying and comparing different kinds of distribution functions (Groisman et al., 1999; Mielke and Johnson, 1974). The knowledge about the underlying distribution is of importance, because every probabilistic property of precipitation is derivable therefrom.

It is worth mentioning that nearly each SPI analysis could comparably be done in terms of the estimated distributions. However, some additional effort is needed because of the missing standardization. It is the standardization, which makes the SPI to the preferred method in analyses, where relative deviations from a climatological mean state are of interest. But the SPI is useless in applications, where direct precipitation properties should be described. Here, the distribution itself gains in importance, as for example precipitation climatology, climate model evaluations or analyzing climate change by comparing distributional properties for present and future climates.

In this chapter we investigate the GD as the standard distribution for monthly precipitation. One of the main findings is that the GD describes precipitation not adequate in many parts of the world, leading to biases in the description of precipitation. In cases where the GD is the wrong assumption, the error will propagate also into the calculation of the SPI. The consequence is a biased SPI, mainly affecting the SPI extremes, leading to underestimation, respectively overestimation, of extreme dryness or wetness.

A comparative method is used to demonstrate that SPI biases arise from wrong distributional assumptions. Therefore, four further distributions, the Weibull (WD), Burr Type III (BD), exponentiated Weibull (EWD) and the generalized Gamma distribution (GGD) are compared to GD. The comparison uses the Akaike information criterion (AIC), which quantifies the information gain or loss by the chosen statistical model (the distribution). This is supported by a simulation study where the performance of different model selection criteria are

analysed. Further, the outcome serves as guide for the interpretation of the later results.

The actual analysis of precipitation data uses multiple data sets, ranging from single observed time series up to precipitation simulated by an actual coupled climate model. This avoids, at least to some kind, that the results are bounded to local areas or the usage of a particular data set.

The chapter is structured as follows: In Section 3.2 applied methods and used data sets are introduced. A simulation study provides a basis for the further analysis is given in Section 3.3. Subsequently precipitation data sets are investigated (Section 3.4), including observed precipitation (Sections 3.4.1 and 3.4.2) as well as climate model output (Section 3.4.3). The results are discussed and an outlook is given in Section 3.5.

3.2   Methods and Data

3.2.1   Standardized Precipitation Index (SPI)

The SPI was introduced by McKee et al. (1993) to classify and monitor dryness and wetness. The calculation of the SPI is based on an equal probability transformation: monthly precipitation is transformed to a standard normal distribution[1] to yield SPI values by preserving probabilities. The standardization ensures, that the SPI gives a uniform measure in different climate regimes or under seasonal dependance. The SPI definition is given in Table 3.1. The SPI can be constructed for timescales ranging from months to years and therewith enables the description of meteorological, agricultural and hydrological drought. The following analysis is restricted to the monthly time scale.

---

[1]given by mean, $\mu = 0$ and standard deviation, $\sigma = 1$

SPI intervals	SPI classes	P[%]
$SPI \geq 2$	W3: extremely wet	2.3
$2 > SPI \geq 1.5$	W2: severely wet	4.4
$1.5 > SPI \geq 1$	W1: moderately wet	9.2
$1 > SPI > -1$	N0: normal	68.2
$-1 \geq SPI > -1.5$	D1: moderately dry	9.2
$-1.5 \geq SPI > -2$	D2: severely dry	4.4
$SPI \leq -2$	D3: extremely dry	2.3

Table 3.1: Standardized Precipitation Index (SPI) classes and corresponding event probability (P).

The transformation depends on the assumed statistical distribution of precipitation. Distributions are calculated for each month separately. As consequence seasonal dependance vanishes, but on the other hand the sample size is reduced. A false distribution type leads to systematic errors, which are most severe at the upper and lower tails and, therefore, for the extreme values. This observation was one motivation for the present analysis.

3.2.2 Distribution functions

The observed monthly precipitation sums are described by asymmetric and positively skewed distribution functions. In the present analysis the Gamma distribution (GD) is compared to four other admissible distributions: the Weibull (WD), Burr Type III (BD), exponentiated Weibull (EWD) and generalized Gamma distribution distribution (GGD). These distributions have two or three parameters (dimensions) and reveal flexibility that increases for the higher dimensional distributions.

For all distribution functions used in this study the minimal set of parameters includes scale (σ) and shape parameter (γ). The three parameter distributions, which include partly the lower dimensional ones as subset, are extended by an additional shape parameter (α). Below main properties of these types are summarized.

(i) The Gamma distribution (GD) is the recommended distribution for SPI calculations (Guttman, 1999):

$$f(x) = \frac{1}{\sigma \Gamma(\gamma)} \left(\frac{x}{\sigma} \right)^{\gamma-1} \exp\left(-\frac{x}{\sigma} \right) \tag{3.1}$$

Γ is the gamma function.

(ii) The Weibull distribution (WD) with the same number of parameters as the GD, is given by:

$$f(x) = \frac{\gamma}{\sigma} \left(\frac{x}{\sigma} \right)^{\gamma-1} \exp\left(-\left(\frac{x}{\sigma} \right)^{\gamma} \right) \tag{3.2}$$

The WD is widely used for the analysis of wind speed, but rarely for precipitation. An exception is Reeve (1996) applying WD for Indian rainfall.

(iii) The Burr Type III distribution (BD) extends the parameter space by an additional shape parameter (α):

$$f(x) = \alpha \gamma \left(\frac{x}{\sigma} \right)^{-\gamma+1} \left[1 + \left(\frac{x}{\sigma} \right)^{-\gamma} \right]^{-\alpha+1} \tag{3.3}$$

The BD extends the flexibility of the GD in terms of kurtosis and skewness (Rodriguez, 1977; Tadikamalla, 1980). An early precipitation application

is is the study of Mielke and Johnson (1974), using a Beta distribution associated with the BD by a parameter transformation. Note: the BD is a special case of the Kappa 4 distribution (Hosking, 1994). The Kappa distribution was applied for SPI comparison (Guttman, 1999) and heavy precipitation events (Kysely and Picek, 2007).

(iv) The exponentiated Weibull distribution (EWD) is also a three parameter distribution:

$$f(x) = \frac{\alpha\gamma}{\sigma} \left(\frac{x}{\sigma}\right)^{\gamma-1} \left[1 - \exp\left(-\left(\frac{x}{\sigma}\right)^{\gamma}\right)\right]^{\alpha-1} \exp\left(-\left(\frac{x}{\sigma}\right)^{\gamma}\right) \qquad (3.4)$$

The EWD extends the WD by a factor including a stretched exponential term and a shape parameter α. For $\alpha = 1$ the WD is obtained.

(v) The generalized Gamma distribution (GGD):

$$f(x) = \frac{\alpha}{\sigma\Gamma(\gamma)} \left(\frac{x}{\sigma}\right)^{\alpha\gamma-1} \exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right) \qquad (3.5)$$

This version includes as special cases the gamma distribution (for $\alpha = 1$) and the Weibull distribution (for $\gamma = 1$).

3.2.3   Parameter estimation

The parameters are estimated by the Maximum Likelihood Method, which is the most general approach for estimation and can be used for all selected distributions. Furthermore it gives the maximized likelihood as a result, which is the basis for Akaike's information criterion (AIC). The maximum likelihood is calculated by minimizing the negative log-likelihood function. A main criterion in the optimization algorithm is the application of equal conditions for the five distributions. To avoid erroneous detection of local minima the optimization is initialized by the minimum on a coarse grid in the parameter space. The optimization is performed by a Quasi-Newton-Method. In a few cases where convergence can not be obtained, a simplex or simulated annealing procedure is applied. Due to the combination of these methods optimized parameters can be determined at all grid points of the data sets.

3.2.4   Validation and comparison of distribution functions

There are two fundamentally different concepts for the assessment of statistical models. On the one hand there are classical tests, like goodness of fit (GOF) or Likelihood-Ratio. On the other hand models can be compared in terms of information gain and loss, for example the Akaike Information Criterion (AIC). The Likelihood-ratio test can not be applied here, because it requires nested types

of distribution functions. Sets of distribution functions are nested if functions with the smaller numbers of parameters are embedded in the parameter space of the higher dimensional distribution. Therefore, this test is only applicable for the comparison of WD and the EWD, with respect to an additional shape parameter, as well as the nesting of GD and WD into the GGD. In the following GOF, therefrom derivated quantities and the AIC are introduced. Their potential for model selection is investigated by a simulation study.

Goodness of Fit test (GOF)

GOF are applied to determine the validity of the estimates. A general approach are bootstrap procedures (Davison and Hinkley, 2003), which are applicable for all distributions. If critical values can not be calculated analytically, bootstrap procedures are a possible approach. The parametric bootstrap is able to obtain critical values by the creation of random samples from the estimates. Since there is a sensitivity to the test statistic, two different statistics, the Kolmogorov-Smirnov statistic (KSS) and Anderson-Darling statistic (ADS) are compared. The p-values, p are given by the number of bootstrapped statistics, s_i^*, which are greater than the statistic of the original sample, s_0, divided by the the number of bootstrapped replicates, n_B ($p = (\#s_i^* \geq s_0)/n_B$; $n_B = 1000$ throughout the chapter).

Derivated quantities

There are two outcomes from the GOF, which may be used for model comparison: the test statistic and the bootstrapped p-values. The test statistic describes the deviation between the estimated model and the data. The best model is chosen according to the smallest test statistic. The p-values are obtained from the statistics resulting from the bootstrap samples. In addition to the test statistic the p-values include the impact of sample variability. The largest p-value determines the preferred model.

Akaike's information criterion (AIC)

The AIC is a versatile method for model selection and is based on the Kullback-Leibler Information which compares models in terms of information gain and loss with respect to an unknown truth (Akaike, 1974; Burnham and Anderson, 2002). AIC is the relative expected Kullback-Leibler Information given by:

$$\text{AIC} = -2\log(\mathcal{L}(\hat{\theta}|y)) + 2K \tag{3.6}$$

with maximized Likelihood ($\mathcal{L}(\hat{\theta}|y)$), estimated parameters ($\hat{\theta}$) dependant on the data (y) and the number of parameters (K). The term, $2K$ corrects the maximum likelihood bias, as an estimator for the Kullback-Leibler Information and is

interpreted as penalty term for higher model dimension. For small sample sizes n in relation to the number of estimated parameters, the modification:

$$\text{AIC}_c = -2\log(\mathcal{L}(\hat{\theta}|y)) + 2K\left(\frac{n}{n-K-1}\right) \tag{3.7}$$

has found to be useful (Burnham and Anderson, 2002) and is applied in the following. AIC_c approaches AIC for large n.

The optimum model is determined by the minimum AIC value (AIC_{min}) within the set of models, i with AIC values AIC_i. The AIC differences (AICD) between the models are important for the comparison and ranking of the models:

$$\text{AICD}_i = \text{AIC}_i - \text{AIC}_{min} \tag{3.8}$$

The best model achieves $\text{AICD} = 0$ with this definition. Models with AICD near 0 are not clearly distinguishable from AIC_{min}. In cases like this it is likely, that the model with AIC_{min} will change from sample to sample. Burnham and Anderson (2002) give guidelines for the interpretation of AICD, which are reproduced in Table 3.2.

3.2.5 Data

Different kinds of data sets are used in the following. Simulated data, with known outcome is created to enable the validation of the statistical methods and to gain guidelines for the later analysis. The actual analysis ranges from a single observed time series to a global precipitation data set produced by a climate model.

- Simulated data: for different sample sizes (20 to 2500), 5000 Gamma distributed random samples have been created with given scale ($\sigma = 20$) and shape ($\gamma = 0.6$) parameters.

- England and Wales precipitation time series: one of the longest observed precipitation time series, starting in the year 1766 and reaching up to the present-day (Alexander and Jones, 2001). Here, the years up to 2007 are used.

AICD	Model support
0-2	substantial
4-7	considerably less
> 10	essentially none

Table 3.2: AICD and their interpretation in respect to the achieved strength of model support.

- Observed high-resolution precipitation data set: the Climatic Research Unit (CRU) data set covers the global land areas in 0.5 degree resolution and time period from 1901 to 2002 (Mitchell and Jones, 2005). The following analysis is restricted to grid points, where at least 1 station is present over the whole time period. This avoids problems arising from the interpolation scheme, filling observational gaps in time and space. Under this restriction two larger regions (Europe (EU; $11°W - 26°E$ and $35°N - 72°N$) and contiguous United States (US; $126°W - 60°W$ and $24°N - 50°N$)) have full data coverage and are analyzed separately.

- Precipitation from coupled atmosphere-ocean climate model: simulated precipitation in T63 spectral resolution (about $2.8°$) from the coupled climate model ECHAM5/MPI-OM (Roeckner et al., 2003; Marsland et al., 2003). The pre-industrial control experiment with constant greenhouse gas concentrations as observed in 1860 with an integration time of 500 years is used. The above given regions are investigated, together with all land and ocean grid points from $60°S - 85°N$.

A threshold of 0.035 $mm/month$ is used throughout, to separate months with and without precipitation and distributions are calculated if at least 50 values are present. Observed time series are tested for linear trends. For a positive outcome (significance level of 95%) the time series are detrended.

3.3   Simulation study

In this section the methods are assessed to determine the best distribution for precipitation data. Given that the GD is most widely applied, a set of gamma distributed random numbers is used as a reference. Since the distribution of the data is known it is possible to determine the conditions which are necessary to reject other kinds of distribution functions. Here, the alternative is limited to the WD. The impact of this restriction is discussed below.

Goodness of Fit test (GOF)

GD and WD are estimated and GOF are performed with a chosen significance level of 0.05. The expected rate of null hypothesis acceptance for the GD is 95%. In 5 % of the cases the null hypothesis will be rejected according to Error Type I. For the simulated data, good agreement with the theoretical value is achieved for all sample sizes and both statistics used, with deviations below 1% throughout (Figure 3.1 a)).

Ideally the WD should be rejected in all samples, but this is far from being reached. For small sample sizes the WD is accepted nearly as often as the GD (Figure 3.1 a)). For increasing sample sizes the acceptance rate decreases,
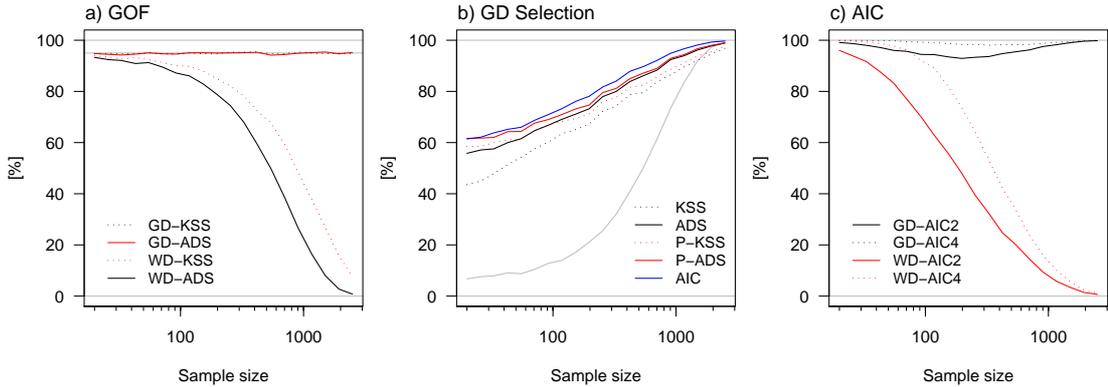
Figure 3.1: Simulation study: a) Rate of null hypothesis acceptance, with either GD or WD as 0 hypothesis. b) Number of times where the GD is selected as the preferred model. The gray line gives the rejection rate of the WD, according to GOF with ADS. c) Number of times where GD (WD) reaches AICD equal or smaller than 2 (4).

but only slowly. At sample size 100, which is close the number of observations typically available, the WD is still accepted in more than 80% of times. ADS performs considerably better than the KSS, as the faster decrease demonstrates.

Interpreting the GOF for the WD in Figure 3.1 a) in terms of power of the statistical test under the alternative hypothesis, that the data follows a GD, demonstrates that the power is low. While here, with the knowledge of the simulation setup, the alternative could be reduced to a single distribution, in the following section the power is even lower, because the set of alternative distributions is larger (and even unknown). The reason for this is due to the similarity of the chosen distribution functions and the large sample variability for small samples. To conclude, for the aim to find the best distribution for monthly precipitation GOF are a basic requisite, although they are not sufficient. The simulation study illustrates that the power of the test is weak. This is enforced for small sample sizes, where similar types of distributions are nearly indistinguishable.

Derivated quantities

Derivated quantities improve considerably the decision for the GD as the preferred model. The rejection rate of the WD, interpretable as measure for preferring the GD on the basis of GOF, is exceeded with all other quantities (Figure 3.1 b)). But sample sizes of at least 50 are needed for the KSS to favor the GD more often. The ADS and the bootstrapped p-values yield higher percentages. The GD is most often selected with p-values resulting from the ADS. It is worth noting, that ADS even outperforms bootstrapped p-values based on KSS. The reason is that the KSS is based on the maximum deviation occurring mainly at

tails. But, systematic errors will also show up with small differences near the center. They are also accounted for with the ADS, which is an integrated measure over all deviations.

Akaike's information criterion

AICD as model selection criterion outperforms the previous methods, improving the choice for the GD as the preferred model. The blue line gives the number of times where the GD yields the AIC minimum (AICD = 0; Figure 3.1 b)). In at least 70% of the cases the GD is favored for sample sizes of about 100. According to Table 3.2 models with strong support for the data are given by AICD $\leq$ 2. Most AICD of the GD are located inside this bound and the percentages exceed 90% independent of sample size (Figure 3.1 c)). This shows that, in the event of the WD reaches the AIC minimum the difference to the GD is small in the majority of cases. The GD has strong support in more than 90% (and 95% for AICD $\leq$ 4) of the cases, independent of the sample size. Since also higher WD percentages are achieved, the best distribution decision is hampered unless very large sample sizes are present.

Summarizing

The presented results are sensitive to the GD parameters of the simulated data. The same analysis redone with different combinations of scale and shape parameters yields almost no impact of the scale parameter. Higher shape parameters however, improve the decision for the GD (not shown). Due to monthly precipitation yields in most situations shape parameters exceeding the given one, the simulation study represents, at least to some kind, the lower border in decision making.

To conclude: the AIC outperforms the other methods. Therefore, the AIC is applied in the following analysis. However, as shown, the rate of preferring the GD is still low for sample sizes typically available in meteorological applications. The reasons for this are sample variability and the similarity of the chosen distributions. It can be expected, that the rate of accepting the GD as the AIC best model is further decreased if the set of alternatives is enlarged. In the later sections these problems are overcome in two ways. Sample size effects are avoided by accounting also for higher AICD. That is, if the GD is the true distribution, the achieved AICD are almost always below 4. This property has been demonstrated to be independent of the sample size (Figure 3.1 c)). Further, a stepwise comparison helps interpreting the results, if the set of alternatives is large.

3.4 Precipitation distributions and SPI

Observed and simulated precipitation data sets are investigated. The observed
data sets (Section 3.4.1 and 3.4.2) deal as case studies related to drought mon-
itoring, whereas climate model precipitation (Section 3.4.3) exemplifies drought
projection studies. Throughout AIC best distributions are determined for each
month separately analog to SPI calculation. Resulting SPI time series are com-
pared with respect to deviations from defined SPI classes.

3.4.1 England Wales precipitation time series

The England Wales precipitation data set consists of a single time series. This
eases analysis and enables the visualization of the results on a monthly basis,
contrary to the later sections where gridded precipitation fields are of interest.

Distribution comparison

The GD reaches the AIC minimum in only one month (Figure 3.2). In all
other months the AICD are greater than 2 and even exceed 7 in the majority of
cases. The WD approaches most frequently the AIC minimum (9 months) and
AICD smaller than 2 for the rest of the months, with the exception of November.
Although the higher dimensional BD reaches smaller AICD than the GD partly,
the information gain by the additional parameter is low compared to the WD.
From the above it follows that small AICD are expected for the EWD and GGD
given their relation to the WD (Section 3.2.2). This is the case and AICD around
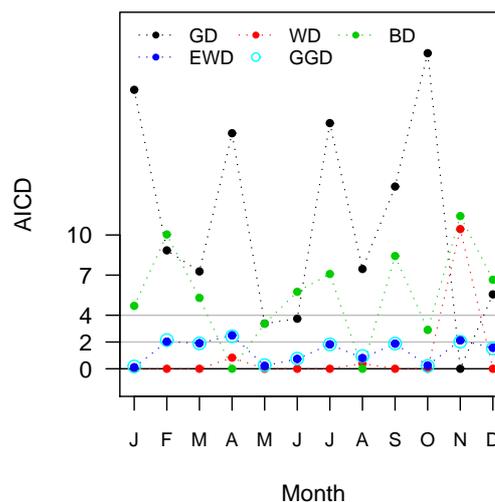
Figure 3.2: England Wales precipitation, AICD separately for each month.

2 for EWD (GGD) are a result of the additional shape parameter, penalized in the AIC calculation.

The GD fails to represent adequately England Wales monthly precipitation. With the restriction of November the WD outperforms the GD. A thereto comparable information gain is given by the EWD and GGD. The BD demonstrates that higher dimensional distributions do not necessarily improve the results. From the viewpoint of just using a single distribution the WD is preferred, due to the low number of model dimensions.

SPI

The impact of the assumed distribution on the SPI time series is analyzed. At first SPI is calculated based on either the GD and the WD and their usability for the description of precipitation and its extremes is investigated. SPI series are expected to be standard normal distributed, with SPI wet (dry) extremes exceeding 2 (falling below -2).

Even by visual inspection of SPI time series a shift to lower values is obvious for GD transformed precipitation (Figure 3.3 a)). Contrary to the expected number of extremes with probability of 2.3% (Table 3.1) extreme dry conditions occur more often (3.41%) than extreme wet conditions (0.96%). The WD transformed SPI is evenly scattered, approaching adequate equal probabilities of 2.31% (2.1%) for extreme wet (dry) conditions (Figure 3.3 b)). A summarized presentation highlights the tail deviations. Therefore the differences from the expected probabilities of the SPI classes (Table 3.1) are calculated. Consistent with the previous findings, the deviations are largest for the SPI based on the GD (Figure
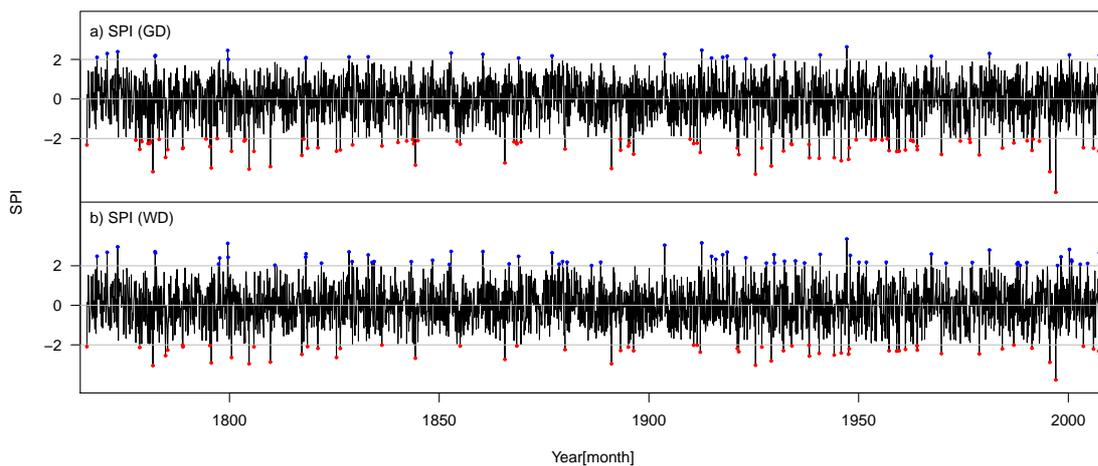
Figure 3.3: SPI time series from England Wales precipitation. SPI transformation based on a) GD and b) WD. Wet (dry) SPI extremes are highlighted in blue (red).
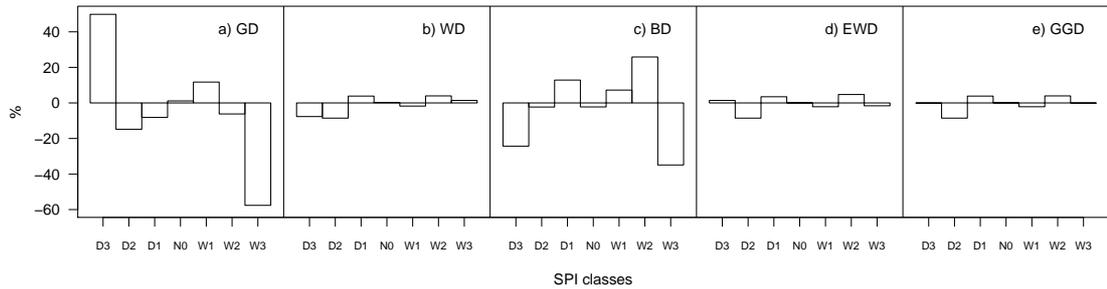
Figure 3.4: Deviations from expected SPI probabilities in percent. SPI time series are constructed with a) GD, b) WD, c) BD d) EWD and GGD e).

3.4 a)). Extreme dryness (wetness) is clearly overestimated (underestimated) and is detected around 40% (60%) more (less) often. All other distributions reduce this bias. Minor deviations are achieved with the Weibull Type distributions (WD and EWD) and the GGD.

Quantile-quantile plots in terms of SPI values are calculated to associate SPI deviations to goodness of fit. Empirical SPI values are obtained by utilizing empirical probabilities for the probability transformation. The standardization enables the presentation of all months in a single plot. The GD underestimates the SPI at the tails (Figure 3.5 a)). Most values drop below the straight line and are partly located outside the confidence bound[2]. Further, the GD shows a slight tendency to overestimate the SPI at the center. These properties together lead

---

[2]Confidence intervals are calculated with 1000 samples of standard normal distributed data.
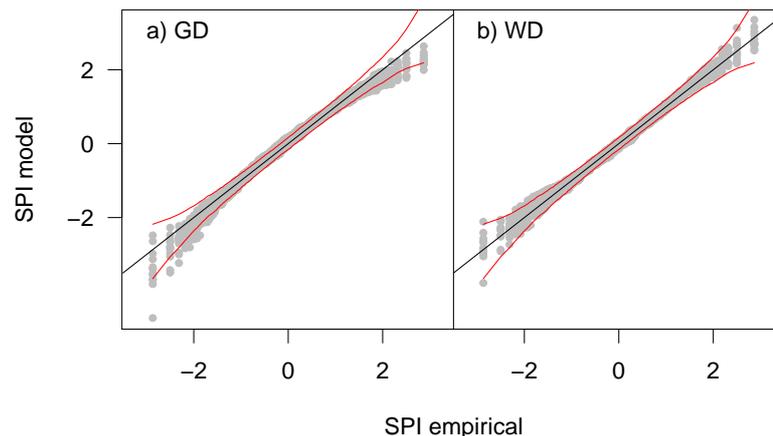
Figure 3.5: SPI quantile-quantile plots based on a) GD and b) WD transformations and empirical SPI values. 95% confidence intervals for standard normal distributed data is given by red lines.

to a curvature. In contrast, WD transformed SPI values are equally scattered around the straight line, with almost all values inside the confidence bounds (Figure 3.5 b)). The quantile-quantile plots demonstrate that the differing SPI time series (Figure 3.3) are not related to random variability. In fact, choosing the GD for SPI calculation leads to systematic deviations, most pronounced at the tails, leading to overestimation respectively underestimation of extreme dryness or wetness.

3.4.2   Observed high resolution precipitation data set

The analysis is extended to clarify, if the presented results are specific to the chosen data set. Therefore European precipitation is investigated grid point wise in the high resolution CRU data set. For the selected region this results to more than 37000 months for which distributions are estimated and compared.

Distribution comparison

A summarized presentation is used to combine calculated AICD into a single figure. Therefore the number of times a distribution reaches values equal or below to a given AICD in percent of all grid points and months are shown. In this way the percentages for AICD = 0 achieved from the single distributions sum up to 100%. Further, the cumulative way of construction leads to increasing curves for increasing AICD. Distribution functions with good data support should show a rapid increase and approach 100% quickly, preferably before AICD around 4. Higher AICD indicate considerably lower model support. For ease of interpretation the GD is at first compared to each other distribution separately, followed by an overall comparison. Additionally, a reference data set is created, consisting of values simulated with the previously estimated parameters of the GD and a sample size equivalent to the observed one. Therefrom derived results show the expected result under the gamma hypothesis.

Beginning with the reference data set, the GD gives most frequently the AIC best model in comparison with all alternatives (WD, BD, EWD and GGD; Figure 3.6, black dotted lines). The AICD rates start around 80% and more, approaching 100% quickly. The frequencies for the alternatives show different behavior reflecting their ability to reproduce the GD properties (Figure 3.6, red dotted lines). The BD shows lowest rates, exceeded by the WD, whereas the EWD and the GGD rates increase strongest and approach 100% for small AICD. This points to good estimates, which follows natural for the GGD, because the GD is a special type of the GGD, so that the differences in AIC are mainly a result of the additional parameter. This argumentation does not hold in case of the EWD. However, the similar behavior of EWD and GGD demonstrates the EWD potential to reproduce GD characteristics.
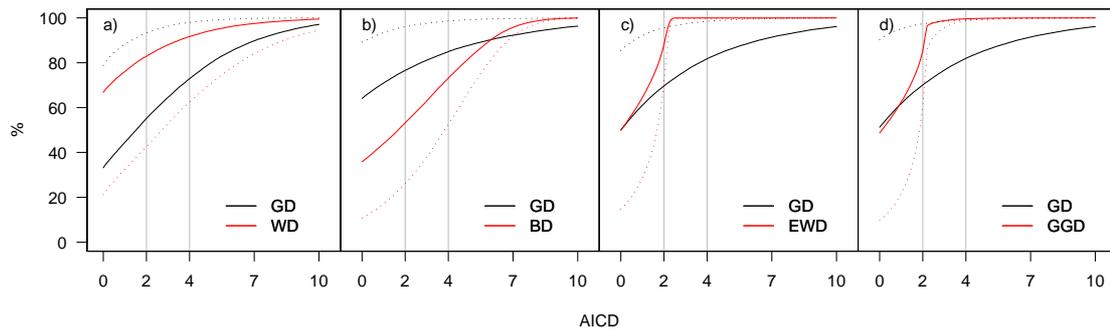
Figure 3.6: AICD frequencies: number of times a distribution yields AICD smaller or equal than a given value for the European region (CRU data set). The GD (black lines) is compared to a) WD, b) BD, c) EWD and d) GGD (red lines). Dotted lines show the respective outcome under the gamma hypothesis.

Analyzing CRU precipitation in terms of information criteria leads to pronounced different frequencies than expected, according to the GD assumption. The GD rates are lowered, whereas rates of all alternatives are increased (Figure 3.6, continuous lines). The BD achieves the smallest, but still an remarkable increase. The WD, EWD and GGD even exceed the accordant GD frequencies for Europe (GGD at least for AICD > 1). Therefore, each of the three distributions is superior to the GD in their ability to describe European monthly precipitation. It is further notable, that neither the GD nor the WD achieves frequencies of 100% for sufficient low AICD. This demonstrates that, none of the chosen two parameter distributions is able to represent European precipitation completely. Below, an overall comparison guides the decision, if one or both distributions can be excluded.

The AICD frequency comparison using all distributions negates this (Figure 3.7). Incorporating BD and EWD in the AICD calculation, resulting GD and WD
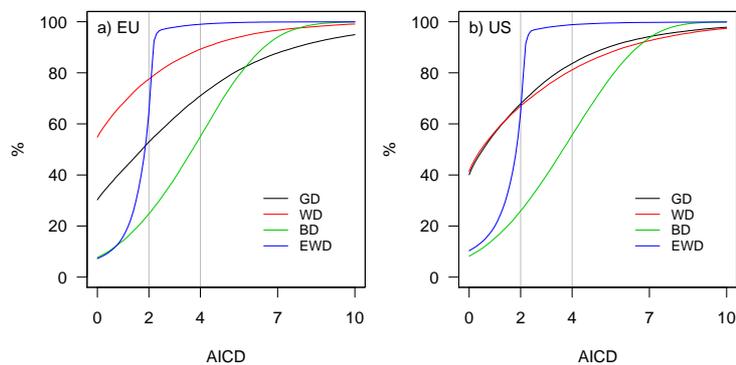
Figure 3.7: AICD frequencies for a) the European region and b) the contiguous United States. AIC comparison with GD, WD, BD and EWD.

AICD frequencies are just slightly lowered compared to Figure 3.6 a). Note that the outcome is equal in each case: replacing the EWD by the GGD or including both (not shown). Because of minor advantages of the EWD, the discussion is restricted to this distribution to simplify matters. The BD frequencies are strongly decreased in comparison to Figure 3.6 b), pointing to minor importance of this distribution for EU precipitation. The EWP frequencies yield a similar reduction for small AICD. But by contrast, the rate increases fast and exceeds the GD and WD rates towards higher AICD. This characteristic is a result of the additional parameter penalized by AIC for months where GD and WD achieve small AICD.

These results are not restricted to the European region. The outcome is similar for the contiguous United States (US; Figure 3.7 b)). The agreement is largest for the BD and EWD frequencies, with just a small offset compared to Figure 3.7 a). Notable differences are the increased (reduced) percentages for the GD (WD), leading to almost equal frequencies.

As the GD and WD are not able to describe EU (US) precipitation sufficiently for all grid points and months, a single two parameter distribution is not recommended for SPI calculation. However, due to their yielded high percentages they can not be excluded. It follows, that a mixture of the GD and WD is a possible solution, given that a complete coverage is achieved for small AICD. The higher dimensional distributions (EWD and GGD) are another suitable possibility. Here one has the advantage of using just a single distribution, although in expense to increasing variance.

SPI

Likewise for the EWP (Section 3.4.1, Figure 3.4) deviations from the SPI classes are shown, including all grid points and months. Noting that, due to the large number of distributions SPI biases may cancel each other out and further, inclusion of months with small AICD will reduce the overall SPI differences. Nonetheless, the GD transformed SPI leads to largest deviations at the tails (Figure 3.8 a)). As before extreme dryness (wetness) is overestimated (underestimated). More than 30% to much (less) extreme drought (wet) events are observed in the considered time period and region. The BD underestimates both extremes (Figure 3.8 c)), whereas the WD and even more the EWD and GGD reduce SPI differences (Figure 3.8 b), d) and e)). Additional SPI deviations for month with AICD ≤ 2 are given, verifying bias reduction for distributions with strong support for the data. Contrary to the GD, WD and BD, the EWD and GGD yield no further reduction, consistent with the fast approach of 100% coverage in Figure 3.7 a).
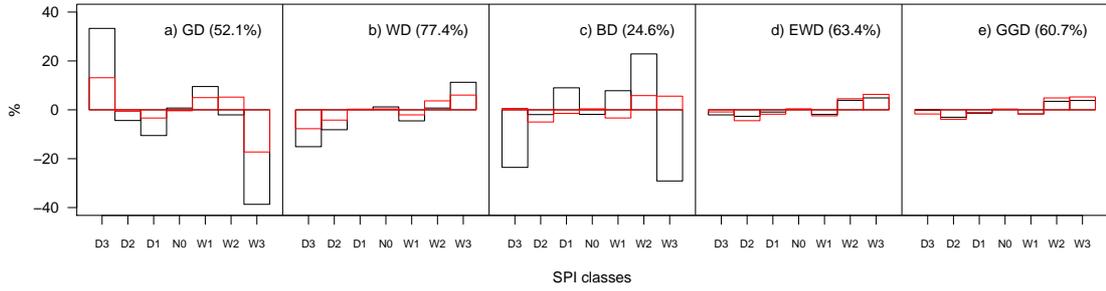
Figure 3.8: Deviations from expected SPI probabilities in percent for all grid points and months (black) and for distributions achieving AICD ≤ 2 (red). SPI time series are constructed with a) GD, b) WD, c) BD, d) EWD and e) GGD. The number of times each distribution approaches AICD ≤ 2 is given in parentheses.

3.4.3 Precipitation from a coupled atmosphere-ocean climate model

For climate change projections of future drought occurrences in terms of the SPI, an essential requirement is the SPI calculation of a reference climate state. This might be either the present or a climate undisturbed by anthropogenic greenhouse gas emissions. Therefore precipitation distributions and therefrom derived SPI are evaluated in a pre-industrial control simulation (ECHAM5/MPI-OM), with an integration time of 500 years and constant greenhouse gas levels fixed at their year 1860 values.

Distributions

At first, European precipitation is analyzed, comparing the distributions separately. Under the Gamma hypothesis (Figure 3.9, dotted lines), the GD yields higher frequencies in relation to the CRU data set (Figure 3.6), whereas the the frequencies of the alternatives (WD and BD) are reduced. This difference is due to the larger sample size, helping to distinguish between the distributions. The sample size is of minor importance for distributions including the GD as subset (EWD and GGD), leading to similar frequencies in Figure 3.6 and Figure 3.9 c) and d).

The GD frequencies are strongly reduced for ECHAM5/MPI-OM precipitation (Figure 3.9, continuous lines) and each of the alternatives outperforms the GD in terms of AIC. Depending on the alternative the GD is not supported according to AIC in around 60% (BD) or even more (WD, EWD and GGD) of all grid points and months. The marginal increase of the frequencies in Figure 3.9 a) and b), remaining below 100% for high AICD demonstrates that beside the GD, neither the WD nor the BD alone is able to cover European precipitation
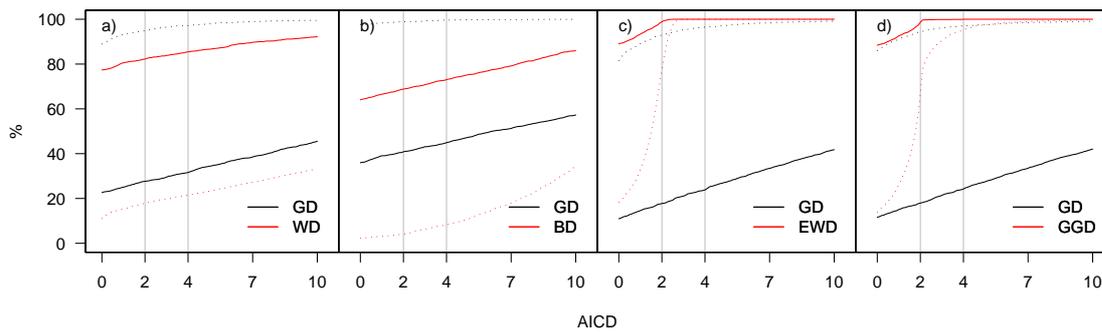
Figure 3.9: Like Figure 3.6, AICD frequencies for the European region (ECHAM5 data set).

completely. This is in contrast to the EWD and GGD, yielding AICD below 3 in all cases.

A comparison using all distributions guides the decision for the preferred one. Analog to the CRU data set, the EWD and GGD are found to be exchangeable, resulting to similar results, but slightly higher frequencies for the EWD through-out. Therefore, the GGD is omitted for AICD calculation below. The GD yields the lowest frequencies, which are even exceeded by the BD (Figure 3.10 a) and b)). This is in contrast to the CRU data set, together with lower WD frequencies (Figure 3.7 a) and b)). The WD however, achieves still high rates in the European region, but not for the contiguous United States. The EWD outperforms all other distributions, even for AICD below 2 and yields the AIC best model in around 40% of times. This is mainly related to the minor importance of GD (GD and WD) in Figure 3.10 a) (Figure 3.10 b)) and therefrom reduced impact of nesting. Concluding that, for a complete coverage the higher dimensional distributions (EWD or GGD) are essential. However, the EWD frequencies do not achieve 100% sufficiently fast, most noticeable in Figure 3.10 b) and c). Given that the EWD is an extention of the WD and is able to reproduce GD characteristics,
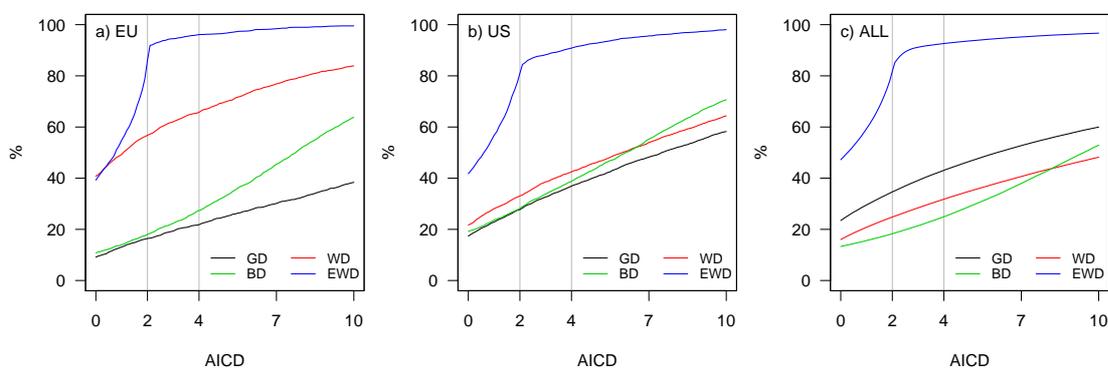
Figure 3.10: Like Figure 3.7, based on the ECHAM5 data set. Additional a comparison is given, using all land and ocean grid points (c).
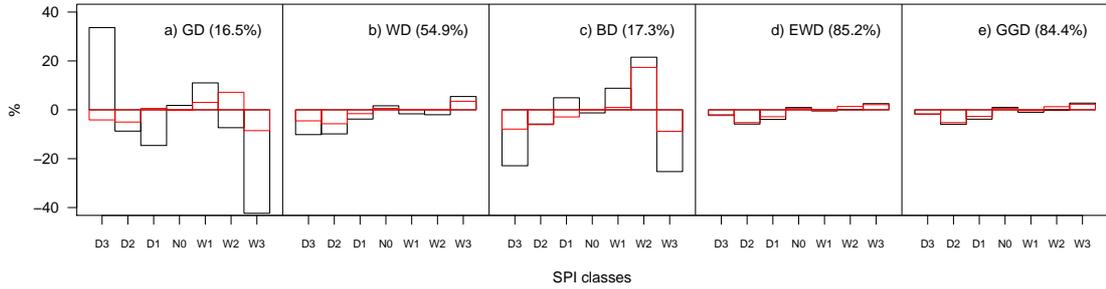
Figure 3.11: Like Figure 3.8, for the European region (ECHAM5)

this limit has to be related to the BD and implies that the BD is needed as well. It is further noteworthy, that the global analysis, including land and ocean grid points yields a distinct difference to the aforementioned results. There, the GD frequencies exceed the ones from the WD (Figure 3.10 c)). This reversed order results from the involved ocean grid points, as the global land area comparison is rather similar to US region (not shown).

SPI

Based on ECHAM5 precipitation the deviations from the expected SPI probabilities are given in Figure 3.11 restricted to the EU region. The largest differences occur again with the GD. The error is scaled down with the WD and minimized with the EWD and GGD. Further, selecting those distributions given by strong support for the data (AICD $\leq$ 2) reduces the deviations.

3.5   Summary and conclusions

Different data sets are analyzed with the aim to verify the Gamma distribution for monthly precipitation sums. The motivation is given thereby, that the GD is widely applied for SPI calculation, model evaluation and climate change assessment. The validation uses the AIC, accounting for information gain or loss, weighted by the numbers of parameters involved. This approach is founded by a simulation study. The main results are summarized:

- The GD fails to represent precipitation in considerable areas of global observed and simulated data. This failure leads to an overestimating (underestimating) of extreme dryness (wetness).

- Improved results are attained with the WD, mainly for the European region, but also for the contiguous United States and global land areas in ECHAM5.

- It is noteworthy, that higher dimensional distributions do not necessarily improve the results, as shown by the BD.

- The EWD and the GGD are the most flexible distributions. They are able to yield accurate estimates in almost all areas, resulting in smallest SPI deviations.

Furthermore, notable differences are detected between the two data sets CRU and ECHAM5, for example the pronounced lower GD frequencies in the model simulation. This points to climate model bias in the precipitation routines. However, the precipitation time series in the two data sets represent grid boxes of different size and it is so far unknown, how interpolation algorithms change distributional properties or the distribution itself.

Concluding that the chosen distribution has a large impact on the SPI, there are two different ways to overcome this problem. Firstly, from the viewpoint of bias-variance adjustment, the preferred SPI calculation should be done stepwise with multiple distributions. That is, using lower dimensional distributions as long as they are appropriate (AICD $\leq$ 2) and changing to the higher dimensional ones for the remaining grid points. Secondly, if comparability and reproducibility are important a feasible solution is the usage of the most general distribution (EWD or GGD). However, this involves an increase of variance.

4 Extreme value statistics for North Atlantic cyclones

Extremes of the cyclone intensity measures geopotential height (z_{1000}), mean horizontal gradient (∇z), cyclone depth (D), and relative vorticity (ζ_{850}), are analyzed in re-analysis data (ERA40) and model simulations (ECHAM5/MPI-OM) in the North Atlantic region for extended winter seasons. Generalized Pareto distributions (GPD) are estimated for model validation and climate change assessment. Covariates, linear trend and North Atlantic Oscillation (NAO) are included to analyze the dependancies of the extremes.

In ERA40 no significant linear trend can be detected, while evidence for a NAO impact on z_{1000}, ∇z, and ζ_{850} extremes is found. Model validation yields good agreement with consistent scale and shape, but a shift to lower values is notable. Like in ERA40 no trend is found in the simulation. The evidence for a NAO impact on cyclone extremes is less corroborated in the simulation, pointing to sample size effects.

In the warmer climate scenario (A1BS) extreme value statistics shows an intensification for all variables. Significant differences in GPD are obtained through testing for lower (higher) parameters. In contrast, considering all cyclones an increase is only present for z_{1000}, while a decrease is found for ∇z and ζ_{850} and no change for D.

4.1 Introduction

Extratropical cyclones are the major source of intra-annual climate variability in midlatitudes. Huge damage is caused by intense storms and heavy precipitation associated with extraordinary intense baroclinic vortices. The growth and intensity of these vortices are determined by sea surface temperatures, baroclinicity and large scale teleconnections (for example the North Atlantic Oscillation, NAO) which might be altered in an anthropogenic climate change (Pinto et al., 2009). Thus, the identification of possible changes of extratropical cyclone extremes is an important issue in the assessment of anthropogenic climate change.

The dominant mode of variability in the North Atlantic region is the NAO. During the strong positive phase of the NAO cyclone tracks tend to have a more

northeastward orientation (Gulev et al., 2001). On the other hand, the variability of the NAO itself is substantially influenced by extratropical cyclones (Löptien and Ruprecht, 2005). However, the cyclone influence on the NAO typically occurs at different time scales (> 10 days) than the influence of the NAO to cyclone activity itself (Benedict et al., 2004). Furthermore, suitable growth conditions for extreme cyclones occur in wider areas during the positive phase of the NAO (Pinto et al., 2009).

A common approach to investigate the characteristics of extratropical cyclones in large data sets is to determine individual cyclones, their tracks and their life cycles by using numerical detection and tracking algorithms. Various methods have been proposed and successfully applied to observations and model data for present day, paleo-climate and anthropogenic climate change scenarios (for example Murray and Simmonds (1991), Hodges (1994), Blender et al. (1997)). Evidence for anthropogenic changes in extratropical cyclone activity and characteristics is documented in many recent model studies (see Ulbrich et al. (2009) and references therein).

For the northern hemispheric winter the majority of scenario simulations show a slight decrease of the total number of cyclones (Bengtsson et al., 2009), while, on the other hand, there are hints that the number of intense cyclones increases (Lambert and Fyfe (2006); defining intense cyclones as systems with core pressure less than 970 hPa). However, models do not agree with respect to these conclusions, in particular if individual regions are considered.

The definition of an extreme event and how its been analyzed influences the results. In most studies, extreme cyclones are defined by a measure of their strength (e.g. the central pressure or the pressure gradient). A cyclone is considered extreme if its strength exceeds a subjectively chosen threshold or if it belongs to upper percentiles of the distribution associated with intense cyclones.

Recently, the assessment of exceptional and rare events by extreme value statistics has become a widely-used technique in the analysis of observational and model data (Katz et al., 2002; Kharin and Zwiers, 2005). Methods and concepts like, for example, the generalized extreme value (GEV) distribution, return values and statistical modeling based on maximum likelihood estimation in presence of covariates contribute substantially to the understanding of extreme events and to quantify potential changes of climate extremes due to anthropogenic forcing. An early application of extreme value statistics for extratropical cyclones is the work of Della-Marta and Pinto (2009). Here, their work is extended through analyzing influences of linear trend and NAO. Further, detection of changes in cyclone parameters due to greenhouse gas warming are not restricted to return value analysis: additionally, the changes in the parameters of the GPD distributions are analyzed.

The aim of this chapter is to analyze the life cycles of intense extratropical cyclones using extreme value statistics. Using a detection and tracking algorithm individual cyclone tracks are extracted from re-analysis data and coupled model

simulations for present-day (20th century) and a moderate greenhouse warming scenario (IPCC A1B). Extreme value statistics is applied to compare cyclone extremes simulated by the model with observations and to assess possible changes due to anthropogenic forcing. Furthermore, the relation between the extremes in cyclone parameters and the NAO are investigated. The analyses concentrate on the North Atlantic region during the extended Northern Hemispheric winter. In Section 4.2 the data used and the methodologies are introduced. The results are presented in Section 4.3. Conclusions, discussion and an outlook follows in Section 4.4.

4.2   Data and Methods

North Atlantic/European cyclones are analyzed in the re-analysis data of the European Center for Medium-Range Weather Forecasts (ERA40) and in simulations with the coupled atmosphere ocean model ECHAM5/MPI-OM. Extreme value statistics is applied to these data sets for extended winters (October-March). The relationships between extreme cyclone properties and the North Atlantic Oscillation (NAO) are considered.

Data

The re-analysis encompasses the last half of the twentieth century from September 1, 1957 to August 31, 2002 (Uppala et al., 2005). The data are computed at T159 spectral resolution on 60 vertical levels and stored with 6 hourly time steps. Here they are mapped to a spatial grid corresponding to T63 spectral resolution to enable an unbiased comparison with the model simulation.

The coupled atmosphere ocean model ECHAM5/MPI-OM uses an atmosphere with T63 spectral resolution on 31 vertical levels (Roeckner et al., 2003). The ocean MPI-OM model has a 1.5° resolution on 40 vertical levels (Marsland et al., 2003). The coupling between atmosphere and ocean is implemented without flux corrections (Jungclaus et al., 2006). The data are available every 6 hours. Three model experiments are analyzed which are part of the IPCC contribution (for further details on the IPCC scenarios see Nakićenović et al. (2000)):

(i) 20C, 20th century: The twentieth century simulation (20C) for the period from 1957 to 2000 simulated with observed greenhouse gas concentrations, aerosols, and solar and volcanic forcing is compared with ERA40 re-analysis data.

(ii) 20CS stabilization: After the year 2000, greenhouse gas concentrations are fixed at their 2000 level. This commitment experiment 20CS is analyzed during the 2005-2100 period.

(iii) A1BS stabilization: To assess the potential impact of anthropogenic greenhouse forcing, the stabilization run of a A1B scenario is analyzed. The A1B scenario describes a climate under gradual CO_2 increase from 2000 to 2100. A stabilization run extends this scenario with fixed greenhouse gas concentrations at the level of 2100. For one ensemble member of the A1B scenario, the stabilization run is extended about another one hundred years (2200-2300). To ensure the best stationarity, this extended stabilization time period is chosen for the analysis, beginning at 2205. However, the stabilization run of the twentieth century (20C) ended in 2100. Therefore the analysis is based on the stabilization run for the time period from 2005-2100.

The cyclone identification is based on the detection of minima in the 1000 hPa geopotential height field (Blender et al., 1997). To avoid erroneous detections, the cyclones must exist at least two days with a minimum mean horizontal gradient of 30gpm/1000km. The cyclones are identified in the Northern Hemisphere (20°N−80°N) but the extreme value statistics is restricted to cyclones which attain an extremum in the North Atlantic sector (30°N - 80°N, 80°W - 40°E).

This study focuses on the following cyclone parameters: geopotential height in the center of a cyclone (z_{1000}), mean horizontal gradient of the geopotential height in the neighborhood (∇z), cyclone depth (D), all measured at 1000 hPa, and relative vorticity in the cyclone center at 850hPa (ζ_{850}). z_{1000} and ζ_{850} are directly extracted from the data sets. ∇z is measured by the horizontal average of the calculated grid point gradients in a region corresponding to roughly 1000 km distance (Sickmöller et al., 2000). By fitting a Gaussian function to the cyclone center and the surrounding height field, D is defined as the difference between a fitted environmental value and the geopotential height in the cyclone center (i.e. D is positive). A description of this method is given in Schneidereit et al. (2010). These four quantities describe different physical meaning and spatial characteristics. Although z_{1000} characterizes the local geopotential height field at the cyclone center, the variable D imply information of the wider environment (roughly 1000 km). ∇z and ζ_{850} characterize the extratropical cyclone on the local scale in the cyclone center. The dynamical characteristics ∇z and ζ_{850} are in close relationship. The extremum for each cyclone property (minima for z_{1000}, maxima for ∇z, D and ζ_{850}) during their individual cyclone life cycles is chosen for further analyses. For convenience, the signs for the z_{1000} minima are reversed to attain positive values for all variables considered.

EOF analysis is applied to monthly mean sea level pressure anomalies and used to determine the NAO by the leading principal component, in the region 20°N−80°N and 80°W−30°E. For ERA40 data the pressure field is interpolated to T63 resolution.

Extreme value statistics

There are two widely applied approaches to achieve extreme value distributions for identical and independent distributed random samples, $X_1, X_2, ..., X_n$. One is based on division into blocks of the sample for which the block maxima are used to estimate the generalized extreme value distribution. The other method uses all data exceeding some well defined threshold, u. The resulting distribution function, for $y = X_i - u$ given $X_i > u$, is the generalized Pareto distribution (GPD):

$$H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \text{, for } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{, for } \xi = 0 \end{cases} \qquad (4.1)$$

with the parameters: scale, σ and shape, ξ. ξ determines the overall behavior of the distribution. The three possible extremal types are: Fréchet $\xi > 0$, Gumbel $\xi = 0$ (achieved by taking the limit $\xi \to 0$), and Weibull $\xi < 0$. In this chapter, the threshold method is applied because more values are incorporated into distribution estimation than with the block maxima approach. However, finding an optimal threshold, u is critical. Two graphical methods exist which are used in the following: the mean residual life plot and fitting over a range of thresholds (for details see Coles (2003)).

The estimated distributions are interpreted in terms of their return levels (quantiles), z_N, which are exceeded once every N years:

$$z_N = \begin{cases} u + \frac{\sigma}{\xi}\left[(Nn_y k_u/n)^\xi - 1\right] & \text{, for } \xi \neq 0 \\ u + \sigma \log\left(Nn_y k_u/n\right) & \text{, for } \xi = 0 \end{cases} \qquad (4.2)$$

n_y are the number of cyclones per year and the number of values exceeding u are given by k_u. Note that k_u/n is the estimate for the exceedance probability. Plotting z_N on a logarithmic scale gives return level plots guiding the decision which extremal type is present. That is, return level curves following a straight line result from Gumbel type distributions; the Fréchet (Weibull) type shows up with concave (convex) curves.

If data points and confidence intervals are included into the return level plots goodness of fit can be derived, which are performing well in the subsequent estimations. The confidence intervals are calculated with profile-likelihood method (Coles, 2003) and enable also statements about the significance of return level differences. According to Kharin and Zwiers (2000) differences between two return levels are significant at the 1% level if their 90% confidence intervals do not overlap.

Cyclone life cycles are serial dependant, with increasing (decreasing) values before (after) an approached maximum. This will harm the independence assumption and affects the uncertainty analysis, so that confidence bounds are

expected to be too narrow. Furthermore, GPD estimates on short time spans may be dominated by few strong and long-lasting cyclones. This problem is omitted by using one value, that is, the maximum during the cyclone life cycle. This is similar to standard declustering schemes but, for the problem at hand, the clusters are physically defined. Note that, beside serial dependant life-cycles, cyclone occurrences cluster in time (Mailier et al. 2006). This clustering may also hold for extreme cyclones, but is not considered in the following.

For observations and transient climate model simulations the extremes may potentially change in time. Such nonstationarities are accounted for by including covariates in the scale parameter and assuming a linear change with time:

$$\sigma(t) = \alpha_0 + \alpha_1 t \ , \tag{4.3}$$

For the cyclone intensity measures, the time steps have to account for discontinuous occurrence of the life cycle maxima. The method is analog to generalized linear modeling. Comparing stationary and trend models enables conclusions about trend significance with the test and criteria described below. This method for trend detection in extremes outperforms other methods, like linear least squares estimation or even Kendall's trend test (Zhang et al., 2004). The main advantage is, that the residuals are not restricted to normal distributed values, which is a wrong assumption in the extreme value context. In the same way, the scale parameter dependance of the NAO time series can be modeled as:

$$\sigma(t) = \beta_0 + \beta_1 \mathrm{NAO}(t) \tag{4.4}$$

Combining Equation 4.3 and 4.4 allows to distinguish between the two different impacts. Equivalent, the shape parameter can be modeled with time dependancies as well.

Each of the assumptions increases the number of parameters to be estimated. Instead of σ, one has to estimate α_0 and α_1 to include the linear trend. To avoid overfitting and to test for significant improvement of higher dimensional statistical models log-likelihood ratio tests (LLR-test) are performed (details are given in Coles (2003)). The significance level is set to $\alpha = 0.05$ throughout (LLR-tests and other applied tests). This kind of test is only applicable in the case of nested models, i.e. models with less parameters have to be in subset of models with more parameters. This restriction is overcome by using Akaike's information criterion (AIC, Akaike (1974)):

$$\mathrm{AIC} = -2\log(\mathcal{L}(\hat{\theta}|y)) + 2K\left(\frac{n}{n-K-1}\right) \tag{4.5}$$

with maximized Likelihood, $\mathcal{L}(\hat{\theta}|y)$, number of estimated parameters, K, and multiplier, $n/(n-K-1)$. This is a small sample size extension of Akaike's original definition (Burnham and Anderson, 2002). From a set of models with

AIC_i (i is the model number), the best model is the one with the minimum AIC value AIC_{min}. Akaike differences (AICD) are calculated to rank and compare the models:

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{min} \tag{4.6}$$

A guideline for AICD is: $\Delta_i < 2$ gives models with strong, $4 < \Delta_i < 7$ considerably less and $\Delta_i > 10$ no support (Burnham and Anderson, 2002). Further interpretation is achieved with Akaike weights (AICW)

$$w_i = \frac{\exp(-0.5\Delta_i)}{\sum_{j=1}^{J} \exp(-0.5\Delta_j)} \tag{4.7}$$

which give the probability that model i is the best one for the given data.

Differences in extremes are commonly analyzed in terms of return levels and respective confidence intervals. Additional, extreme value distributions can be compared in terms of their parameters by performing tests for increased or decreased scale or shape parameters. In the following, this is done with the method of time dependant parameters. The data sets of interest (for example ERA40 and 20C) are merged and analyzed together. Instead of time t (Equation 4.3), a step function is included with values of -1 (1) over the time period of the first (second) data set. The scale parameter depends on the selected threshold. To simplify interpretation the higher threshold of the two data sets is applied. With this concept different statistical models are built, ranging from the "stationary" model with no difference between the parameters of the two data sets to the model, where both the scale and shape parameter are allowed to change. The different models can then easily be compared and tested employing LLR-test and AIC. With the best model obtained in this way it is possible to reconstruct return level plots for each data set separately. Scale and shape parameters are achieved from the combined model. According to Equation 4.2 the number of cyclones per year and the exceedance probability are also required which, however, are easily calculated for the corresponding data sets (note: these quantities do not affect model estimation). Confidence intervals can be calculated by resampling.

4.3   Extreme value statistics and cyclone life cycles

In the following, life cycle maxima of the cyclone intensity measures geopotential height (z_{1000}) mean horizontal gradient (∇z), cyclone depth (D) and relative vorticity (ζ_{850}) are analyzed. First, overall density estimates are compared, followed by an investigation of present day extremes and their potential dependance on covariates. The impact of greenhouse gas warming on the extreme statistics concludes this section.

4.3.1   Density estimates

Before analyzing extreme values, estimates for the overall distributions of the variables of interest are evaluated for all data sets (Figure 4.1). All estimates show uniform distribution functions. While the central geopotential height, z_{1000} is nearest to a Gaussian distribution, the other quantities are positively skewed (Figure 4.1 a)). The skewness, an important quantity affecting the extremes, is reproduced by ECHAM5/MPI-OM. The deviations between observed (ERA40) and model simulated data (20C) are small. The densities agree well, with the exception of the relative vorticity, ζ_{850} where the ECHAM5/MPI-OM density is slightly shifted to lower values (Figure 4.1 d)). Comparing the differences in the mean underlines this. A t-test shows no significant differences between ERA40 and 20C in the z_{1000}, ∇z, and D. But the mean ζ_{850} in 20C is significantly lower than in ERA40.

The transient (20C) and stabilization (20CS) model runs give nearly identical results. Despite an assumed higher variability in 20C, probably associated with increasing greenhouse gas forcing, the distributions of the cyclone parameters remain largely unaffected compared to 20CS. This may be an indication of the small importance of the forcing during this (relatively short) time period compared to the natural variability of the system. Larger, but still small differences are visible, if the two stabilization runs 20CS and A1BS are compared. The tail
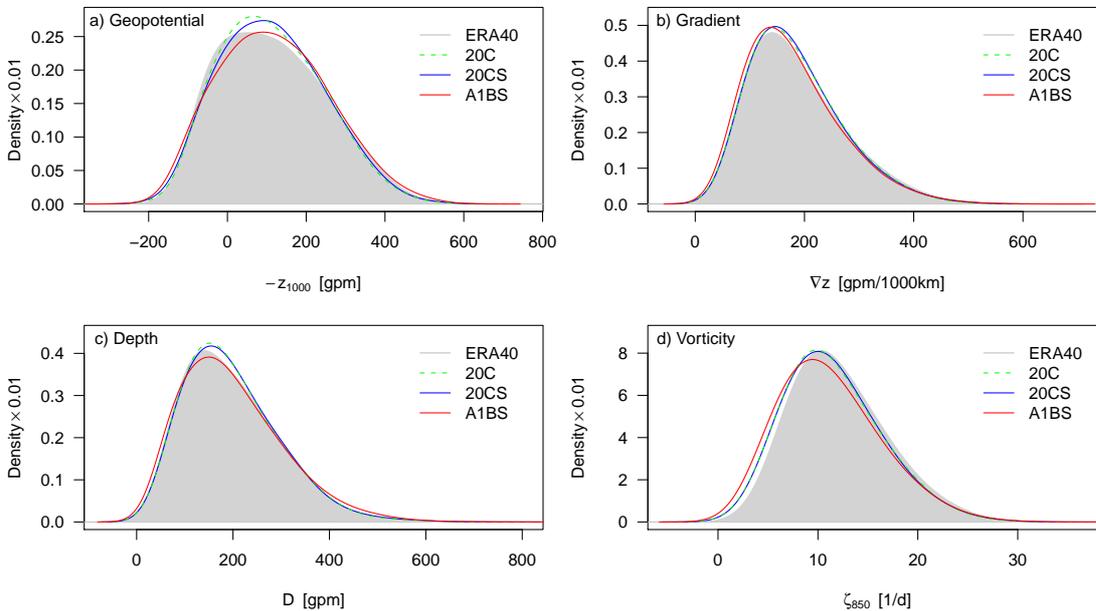
Figure 4.1: Kernel density estimates for maximum cyclone life cycle properties of a) geopotential height, z_{1000}, b) mean horizontal gradient, ∇z, c) cyclone depth, D and d) relative vorticity, ζ_{850}. Used are all data sets in this study, re-analysis (ERA40), transient (20C) and the stabilization model runs (20CS and A1BS).

behavior shows slight differences for z_{1000} and D (Figure 4.1 a) and c)). The distributions of ∇z and ζ_{850} are shifted to lower values (Figure 4.1 b) and d)). Testing for differences in the mean gives a significant higher z_{1000} in A1BS, no change for D, and significantly reduced values for ∇z and ζ_{850}.

4.3.2 Extremes of the present day climate

Extreme value statistics is applied to compare the models (20C) present day climate with observations (ERA40) and to access the dependance on linear trend and NAO.

Return level estimates

Comparing ERA40 and 20C, Figure 4.2 displays return level plots. For return periods below 1 year the return values are significantly lower in 20C for z_{1000}, ∇z and ζ_{850}. This shift is not present for D. For the local measures z_{1000}, ∇z and ζ_{850}, the results indicate Weibull-type distributions for both data sets with comparable shapes. The estimated shape parameters, $\hat{\xi}$, together with scales, $\hat{\sigma}$ and thresholds, u are given in Table 4.1 and 4.2. For these two parameters the model exhibits longer return times (lower return levels) compared to the observations. For the integral (large scale) measure D virtually no difference occur for return periods up to about 10 years. However, while the shape hints to a Weibull-type distribution for ERA40, a Gumbel-type distribution is plausible for the model ($\hat{\xi} = -0.003$; see Table 4.2), though this might be caused by sampling errors for long return periods.

To clarify the observed differences in return levels, the GPD are investigated. Return levels are controlled by scale and shape. For z_{1000} the estimated parameters, $\hat{\sigma}$ and $\hat{\xi}$ are similar for 20C and ERA40 (Table 4.1 and 4.2). Larger differences occur for the other quantities, where the scale, the shape or both are affected. One example is the shape parameter for the depth, D, which is consistent with $\xi = 0$ in 20C, but not in ERA40. To decide whether the extreme value distributions are significantly different between the two data sets or, if one distribution could be found that fits the quantities from both data sets equally well combined estimates are calculated (see Section 4.2). The statistical model set consists of four models: (i) "stationary", no time dependance, i.e. the same parameters in ERA40 and 20C, (ii) $\sigma(t)$, (iii) $\xi(t)$ and (vi) $\sigma(t)$ and $\xi(t)$. In (ii) - (iv) the parameters are allowed to be different in ERA40 and 20C.

Applying LLR-test to the statistical models reveals no significant differences in the parameters between ERA40 and 20C for all four cyclone properties (all p-values are higher than 0.15). Furthermore, the AIC minimum is reached with the "stationary" model, with just one exception (not shown). The vorticity reaches the AIC minimum with the $\sigma(t)$ model. However, the "stationary" model is very near to this minimum (with AICD $= 0.003$) and is preferred due to lower number
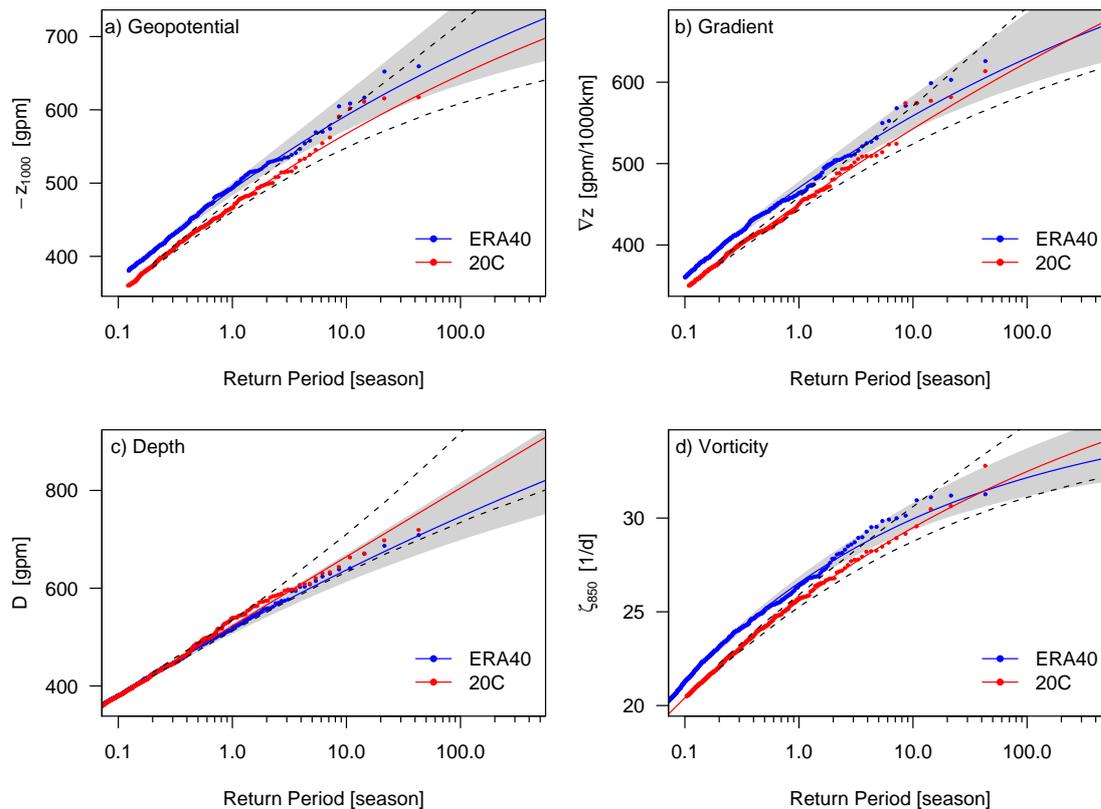
Figure 4.2: Return level plots for a) geopotential height, z_{1000}, b) mean horizontal gradient, ∇z, c) cyclone depth, D and d) relative vorticity, ζ_{850}. The empirical (points) and estimated return levels (lines) have the same colors for ERA40 (blue) and 20C (red). 95% confidence intervals are shown for ERA40 (gray shaded) and 20C (dashed lines).

of parameters. Note that from the combined models reconstructed return levels (not shown) are similar to the curves in Figure 4.2. However, the crossing between 20C and ERA40 curves in Figure 4.2 b) (∇z) and Figure 4.2 d) (ζ_{850}) vanishes as an effect of the similar extreme value distributions.

Since, the estimated parameters agree in both data sets, the differences in the return levels result solely from a shift in the location of the distributions. If the thresholds agree, this shift is expressed through differing exceedance probabilities (Equation 4.2). For ∇z the exceedance probabilities are 0.044 (0.033) in ERA40 (20C). The same argument holds for z_{1000} (ζ_{850}) with probabilities 0.036 (0.067) in ERA40 and 0.026 (0.052) in 20C respectively. The exceedance probabilities for D reach the same rounded value of 0.073 in ERA40 and 20C. The effect of consistent parameters and exceedance probabilities leads to identical return levels from combined model approach (not shown).

To summarize, GPD with the same scale and shape parameters are found in ERA40 and 20C for each cyclone quantity. Reduced return levels in 20C (z_{1000},

ERA40	u	$\hat{\sigma}$	$\hat{\xi}$	Δ_{stat}	Δ_{trend}	Δ_{NAO}	$\hat{\beta_1}$
z_{1000}	380	58.39 (4.33)	-0.089 (0.052)	6.5 (0.037)	7.7 (0.02)	0.0 (0.943)	7.42 (2.34)
∇z	360	53.31 (3.42)	-0.095 (0.043)	1.7 (0.273)	3.7 (0.099)	0.0 (0.628)	3.9 (1.97)
D	320	64.38 (2.66)	-0.046 (0.029)	0.0 (0.554)	2.0 (0.203)	1.7 (0.243)	1.08 (1.8)
ζ_{850}	20	3.08 (0.16)	-0.191 (0.033)	1.9 (0.255)	3.9 (0.094)	0.0 (0.651)	0.171 (0.085)

Table 4.1: Estimated parameters scale, $\hat{\sigma}$ and shape, $\hat{\xi}$ as well as determined thresholds, u for the selected cyclone properties: geopotential height, z_{1000}, mean horizontal gradient, ∇z, cyclone depth, D and relative vorticity, ζ_{850} for observations (ERA40). AICD for the "stationary" (Δ_{stat}), linear trend (Δ_{trend}) and NAO model (Δ_{NAO}) are given with AICW in parentheses, as well as estimated slope, $\hat{\beta_1}$ for the NAO model. The standard deviation for estimated parameters is given in parentheses.

20C	u	$\hat{\sigma}$	$\hat{\xi}$	Δ_{stat}	Δ_{trend}	Δ_{NAO}	$\hat{\beta_1}$
z_{1000}	360	56.8 (4.17)	-0.088 (0.051)	0.0 (0.415)	1.1 (0.243)	0.4 (0.342)	3.435 (2.678)
∇z	350	48 (3.33)	-0.054 (0.048)	0.9 (0.335)	2.6 (0.143)	0.0 (0.522)	3.906 (2.235)
D	350	62.23 (3.51)	-0.003 (0.042)	1.8 (0.251)	2.9 (0.141)	0.0 (0.608)	4.599 (2.308)
ζ_{850}	20.5	2.55 (0.17)	-0.118 (0.045)	0.0 (0.528)	2.0 (0.199)	1.3 (0.273)	0.088 (0.103)

Table 4.2: Like Table 4.1, but for the transient climate model simulation (20C).

∇z and ζ_{850}) result from lower locations of the GPD, which are expressed through smaller exceedance probabilities for agreeing thresholds. A smaller number of threshold crossings may be a direct effect of the lower resolution in 20C, since higher resolution also results in strengthening of cyclones on larger scales as show by Jung et al. (2006) for the ECMWF model.

Trend and NAO dependance of extremes

There is evidence, that cyclone properties have changed in the recent past (Ulbrich et al. (2009) and references therein). However, trend analysis using a generalized linear models approach with extreme value distributions has not been undertaken so far. Impacts on the extremes by the covariates, linear trend and North Atlantic Oscillation (NAO), are investigated in ERA40.

LLR-tests applied for the trend model yield p-values higher than 0.3 for each cyclone property. This excludes an influence on the extremes by a linear trend. This is further confirmed by the AICD, where the ranking gives the weakest support for the model with linear trend (Table 4.1). So that, the trend component (Equation 4.3) does not improve model fits and the stationary model would be preferred, if this two models are considered.

Using the NAO as covariate (Equation 4.4) leads to differing findings. For the geopotential height, z_{1000}, the applied LLR-test gives high significance for the NAO outperforming the stationary model (p-value lower than 0.005). The LLR-

test is passed too, if the vorticity, ζ_{850}, is considered. The depth, D, fails the test. The p-value for the gradient, ∇z, is just slightly higher than the chosen α (p-value = 0.054). The AIC best model is achieved with the NAO as a covariate with ∇z, as for z_{1000} and ζ_{850}. Furthermore the estimated $\hat{\beta}_1$ is high enough to exclude $\beta_1 = 0$ compared to the given range by the standard deviation. Giving weight to the conclusion that the NAO improves model fits also for ∇z. However, the strength of support is lower as for z_{1000} and reaches nearly the same probability as ζ_{850}.

The observed NAO exhibits decadal-like trends, being upward directed from the sixties until the end of the eighties. This together with an assumed increase in the cyclone quantities through greenhouse gas warming might be interpreted as the NAO just reflecting this upward trend. This possibility however, could be excluded through the given trend analysis, stating no significant trend.

In summary, strong support is found that the NAO has an impact on the geopotential height, z_{1000}, extremes. The same holds also for ∇z and ζ_{850}, but with restrictions on the strength of evidence. This is consistent with results found by Pinto et al. (2009), that is, higher cyclone intensities ($\nabla^2 p$) and deeper cores are related to the daily projected positive NAO phases. Here we extend this finding to the gradient, ∇z. Furthermore covariate modeling with the monthly NAO leads to the probabilistic interpretation that, in months with positive (negative) NAO, the probability of extreme intense cyclones is increased (decreased).

In agreement with ERA40 no significant trend could be found in 20C in none of the cyclones quantities (p-values higher than 0.3). With AIC decision making is hampered because all AICD are smaller than 3, demonstrating plausibility for each model. But the trend model gives highest AICD and the estimated slope, $\hat{\alpha}_1$, is smaller than the corresponding standard deviation (not shown). Therefore, the trend exclusion with AIC is not as strong as in ERA40.

With the NAO-model an improvement is only present for depth, D, with an achieved p-value near the significance level (p-value: 0.052). Most striking is the strongly reduced evidence of the NAO impact on z_{1000} extremes in 20C, which is clearly present in ERA40 (see AICW values in Table 4.2). The small AICD however, indicates an improvement of the model fit, confirmed by $\hat{\beta}_1$ with smaller standard deviation than parameter estimate. But the improvement is to weak to separate it from the stationary model on the basis of AICD alone.

Further differences are found for the NAO influence on D and ζ_{850} extremes: AICD gives weak support for D in 20C (probability around 60%), which is not present in ERA40. The weak support for ζ_{850} in ERA40 is vanished in 20C showed by the parameter estimate, $\hat{\beta}_1$, including 0 if referred to the standard deviation. The dependance of ∇z on the NAO is the same in ERA40 and 20C, with slightly reduced probability.

To summarize, the differences between observed (ERA40) and modeled (20C) cyclone dependancies consist mainly in the strength of support for the statistical models chosen. Differing findings may not necessarily result from a dynamic

response missing in the model. Another possibility is the time evolution of the observed NAO in the ERA40 period, which exhibits decadal trends. A phase shift of these trends in the model could be an explanation. Note that the overall agreement with ERA40 is higher in 20CS (see below) reflecting a strong sensitivity to the sampling size (which is, of course, also true for ERA40).

Excursion: NAO and z_{1000} extremes in ERA40

The impact of the NAO on the z_{1000} return levels is illustrated in Figure 4.3 a). The lines correspond to "NAO-worlds", i. e. the NAO is hold fixed on the selected levels from -2 to 2 and are calculated with the determined parameters from the covariate model. Using Equation 4.4 with estimated $\hat{\beta}_1$ from Table 4.1 results in a scale parameter difference of almost 30 gpm between NAO = -2 and NAO = 2. This difference has a large impact on the calculated return levels. The return level which is exceeded once every 100 years with NAO = -2 is roughly 570 gpm (lower horizontal gray line). The return period is largely reduced with NAO = 2, the same level is now exceeded every 3-4 years or, on the other hand, the 100 year return level is increased up to more than 700 gpm (upper horizontal gray line).

Note that the stationary model (gray line in Figure 4.3 a)) gives similar results as NAO = 1 and not, as one might expect the neutral NAO = 0 case. The reason for this departure is the disproportional occurrence of threshold exceedances, as there are 73.9% (26.1%) threshold crossings in the positive (negative) NAO phase. This difference in number affects the maximization procedure, which is dominated by the majority of values.

Figure 4.3 b) confirms the results with another method. Here, compositing extremes is applied to achieve GPD for the case where the NAO is greater (less)
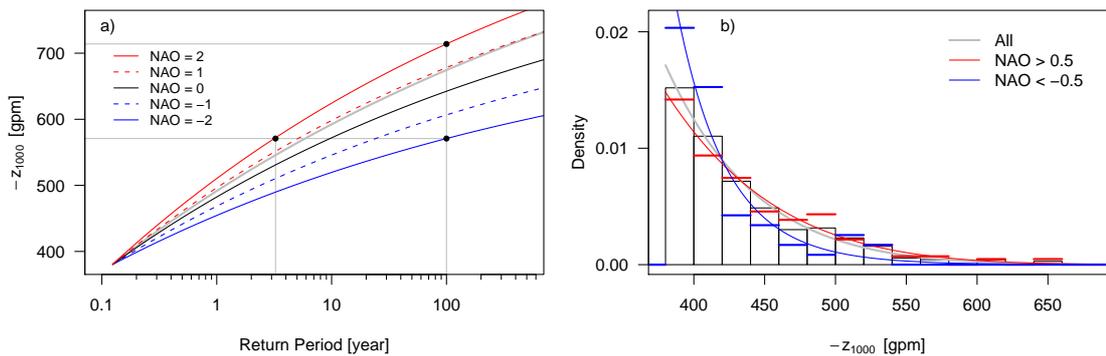

Figure 4.3: NAO impact on geopotential height, z_{1000} extremes. a) Return levels for given NAO value from -2 to 2, influencing the GPD scale parameter. b) Histograms, as well as estimated density for GPD, for composites of z_{1000} extremes, where NAO exceeding (falling below) 0.5 (-0.5) and all values together.

than 0.5 (-0.5). The gray line gives the resulting distribution of the stationary fit (using all values above the threshold). The estimated distributions agree well with the corresponding histograms. The impact of the NAO on the extremes gives higher (lower) densities for NAO < -0.5 (NAO > 0.5) for small values. This relation is reversed for values exceeding ≈ 425 gpm. The deviations of the NAO > 0.5 (NAO < -0.5) distributions, compared to the fit with all values, are small (large).

Life cycles and spatial occurrence of extremes

The mean life cycles of the four cyclone parameters are shown in Figure 4.4 for all cyclones in ERA40 and 20C, as well as the 95% and 5% quantiles. Additionally, the individual life cycles of the five most extreme cyclones are included, whereas extreme is related to z_{1000}. For all parameters, the mean life cycles are in good agreement between ERA40 and 20C. If the life cycles of the five most extreme cyclones in the respective parameter are considered, the good agreement is confirmed (not shown). An extreme cyclone in z_{1000} may not naturally belong to
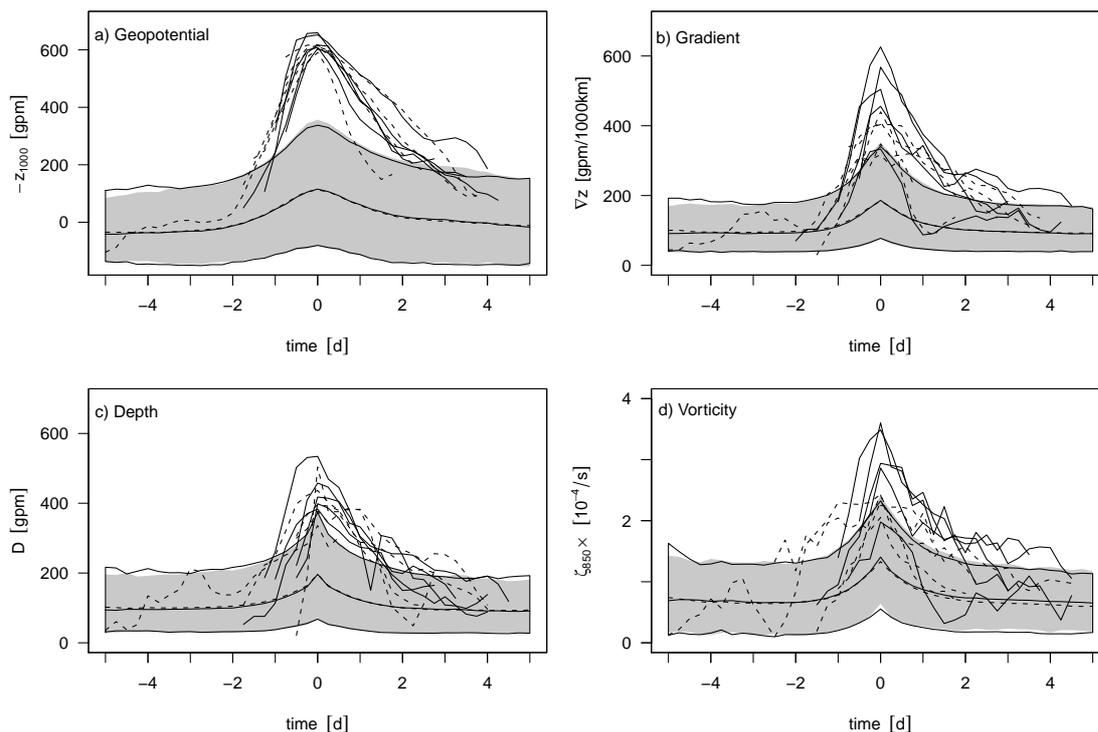
**Figure 4.4:** The five most extreme life cycles in respect of z_{1000} centered about the extremum for a) geopotential height, z_{1000}, b) mean horizontal gradient, ∇z, c) cyclone depth, D and d) relative vorticity, ζ_{850} for ERA40 (solid) and 20C (dashed). Ninety percent of the data (confined by the 5 and the 95 quantile) are presented by shaded (framed) region for ERA40 (20C).

an extreme cyclone in the other variables. However, nearly all extreme cyclones with respect to z_{1000} belong to the most extreme upper 5% in the other variables for ERA40. The same holds if another reference parameter is chosen. Similar results but with reduced absolute values in terms of z_{1000} occur in 20C.

The spatial density estimates of the locations where the life cycle maxima occur are shown in Figure 4.5 for the selected quantities geopotential height, z_{1000} and the relative vorticity, ζ_{850}. Only cyclones are included whose maxima pass the according threshold (see Table 4.1 and 4.2). Comparing the density maxima of z_{1000} and ζ_{850} for ERA40 (Figure 4.5 a) and c), respectively) yield two different positions of the centers. The spatial density estimates with respect to the gradient (not shown) indicate a maximum in between the maxima of ζ_{850} and z_{1000}. Note that, the spatial density estimates consider the maxima of extreme cyclones in the accordant parameter only. Therefore, the number of the considered cyclones varies.

For example, a spatial large cyclone with very deep core pressure does not necessarily exhibit a strong wind field and high vorticity values. This circumstance could be responsible for the different cyclone extreme behavior. Considering only cyclones belonging to the intersection of extreme events in both cyclone variables,
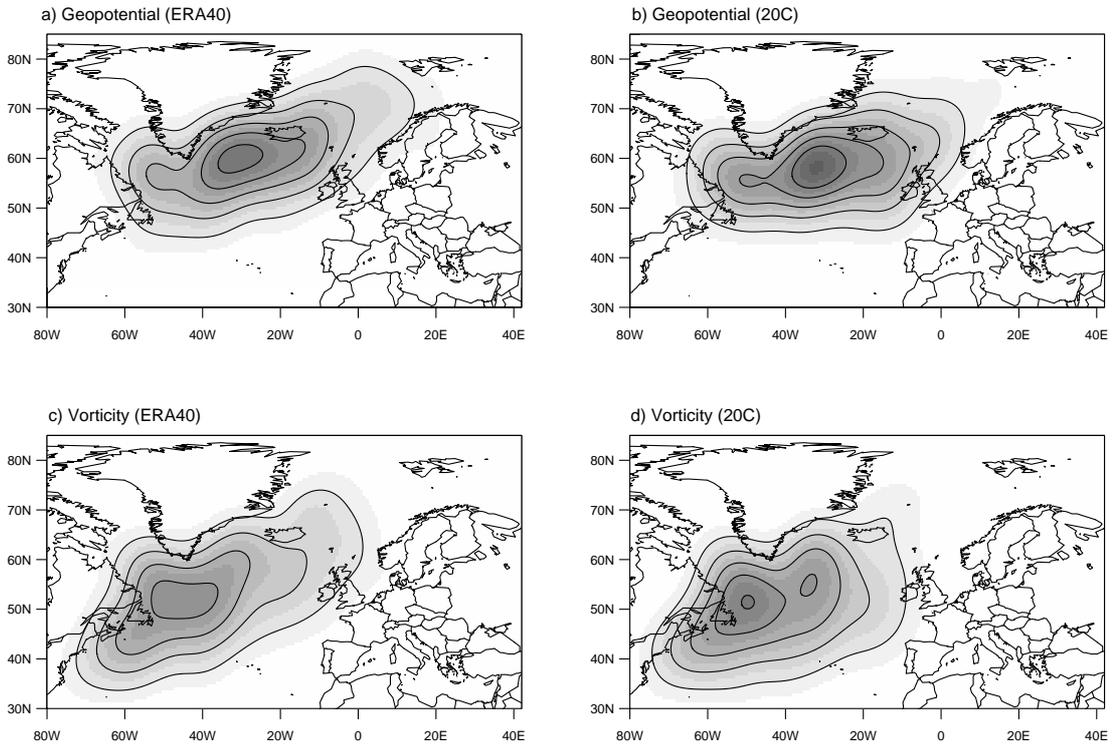
Figure 4.5: Spatial density estimates of the most extreme cyclones (values above thresholds) in the centered geopotential height, z_{1000} (upper panels) and the relative vorticity, ζ_{850} (lower panels) for ERA40 (left) and 20C (right).

z_{1000} and ζ_{850}, the spatial density maxima of z_{1000} is shifted to lower latitudes (not shown) overlapping with the maxima in ζ_{850}. The determined cross correlation function (not shown) of the extreme cyclone life cycles suggest to a time shift (roughly 6 hours) between the vorticity and central geopotential height. This is consistent with Bengtsson et al. (2009) showing that vorticity reaches the maximum before surface pressure. From the geostrophic adjustment it would follow that the mass field leads the wind field if the considered scale is greater than the Rossby radius of deformation. Note that our results indicate the opposite case.

In 20C, the spatial density estimates of the four parameters correspond to the densities of ERA40. Due to a coarser resolution the spatial variability is smaller in the simulation.

### 4.3.3   Extremes changing with greenhouse warming

The stabilization model runs 20CS (present day) and A1BS (greenhouse warming scenario) are investigated for changing cyclone extremes.

Return level estimates

Changes of the return periods (levels) are displayed in Figure 4.6 comparing 20CS and A1BS. For a better comparison, the stabilization run 20CS is used instead of the transient run 20C. It should be noted that a comparison between 20C and 20CS yields differences in the estimated parameters (Table 4.2 and 4.3). For all four cyclone properties a higher scale and lower shape parameter is found in 20CS. The 20CS return level curves, however, are located inside the 20C confidence bounds over the full range of return periods (not shown). An explanation for the differences is the smaller sample size in 20C, resulting in large variability from sample to sample, together with sensitive shape parameter estimation and compensating effects between the two parameters.

The return levels show qualitatively consistent results for all variables which, however, vary in detail. In the A1BS climate, a significant shift towards higher return levels (shorter periods) is only found for z_{1000} up to return periods of about 10 years. The tendency to stronger cyclones is still present but less notable for ∇z and ζ_{850}. For D a consistent shift can be seen for longer (> 6 years) and shorter (< 1 year) periods while medium periods show only a very week signal. Remarkable is the change from Weibull to Gumble type distribution, indicated by the approximate straight return level curve for D in A1BS. This is confirmed through the estimated shape parameter, where $\xi = 0$ is included in the interval given by the standard deviation. However, changes in terms of return levels are not significant for ∇z, ζ_{850} and D.

Deeper insight into the observed return level changes can be achieved by analyzing the GPD parameters. In 20CS and A1BS similar shape parameters are found for z_{1000}, ∇z and ζ_{850} (Table 4.3 and 4.4), while the scale parameters
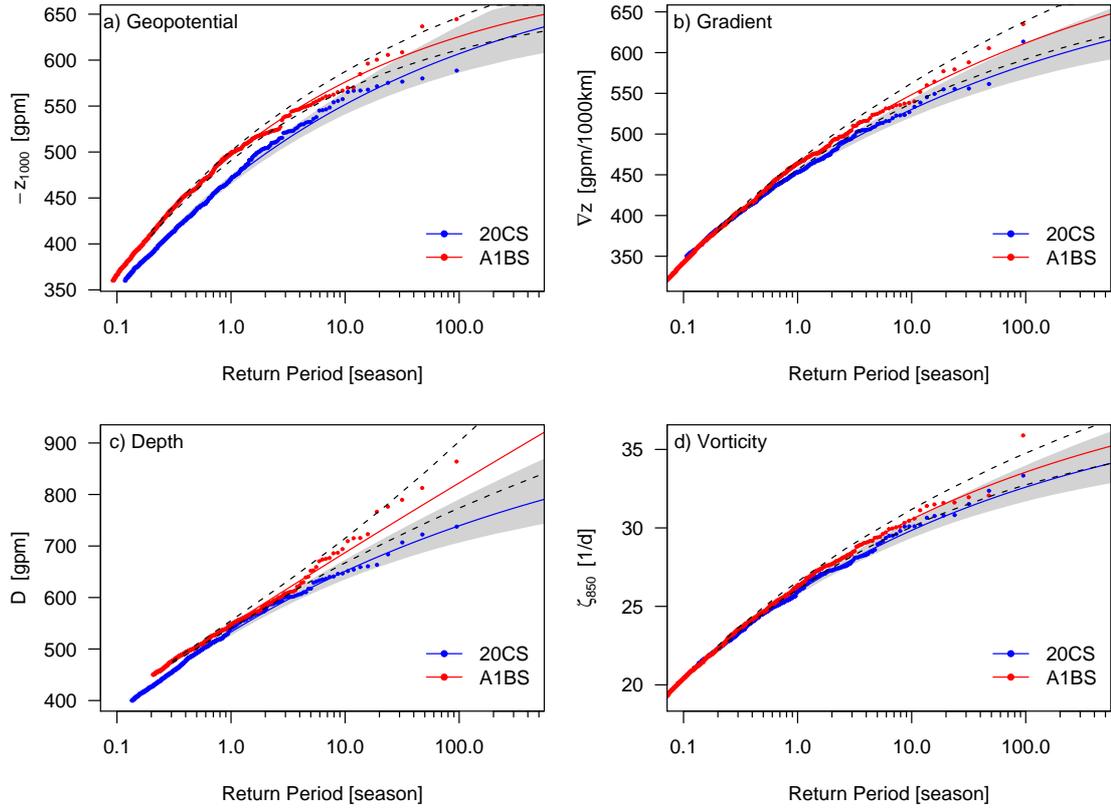
Figure 4.6: Like Figure 4.2, for the 21th (22th) stabilization run 20CS (A1BS).

are increased in A1BS. Different behavior is found for depth, D with lower scale and higher shape parameter in A1BS. Combined estimates are calculated and LLR-tests are applied in two steps: first, $\sigma(t)$- and $\xi(t)$-models are compared to the "stationary" model (LRT-test 1) and second, the model allowing for different σ and ξ parameters in the data sets is compared to all lower dimensional models (LRT-test 2). For z_{1000}, ∇z and ζ_{850} the "stationary" model with the same parameters in 20CS and A1BS is implausible (Table 4.5): LLR-test1 favors the $\sigma(t)$-model against the stationary model since the p-values are all lower than 0.01. Also, the $\xi(t)$-model shows an improvement, but the reason for this seems to be a compensating effect. This is demonstrated by LLR-test2, because the combined model (σ and ξ are allowed to change; Table 4.5) outperforms the $\xi(t)$-, but not the $\sigma(t)$-model, demonstrating that a change in the scale is sufficient to describe the changes that occur in A1BS. Accordingly the AIC minimum is achieved by the $\sigma(t)$-model with AICW above 0.5. Interestingly, for ζ_{850} extremes, where the smallest differences in the return levels are obtained, the highest confidence is found for different scale parameters (AICW; $w_\zeta = 0.726$). For D, a single time dependant model does not give an improvement compared to the "stationary"

20CS	u	$\hat{\sigma}$	$\hat{\xi}$	Δ_{stat}	Δ_{NAO}	$\hat{\beta}_1$
z_{1000}	360	61.37 (2.98)	-0.169 (0.034)	17.1 (0)	0 (1)	7.32 (1.543)
∇z	350	54.33 (2.37)	-0.146 (0.028)	5.2 (0.068)	0 (0.932)	4.19 (1.509)
D	400	75.31 (4)	-0.124 (0.038)	0 (0.687)	1.6 (0.313)	1.771 (2.674)
ζ_{850}	21	2.75 (0.13)	-0.151 (0.031)	0 (0.712)	1.8 (0.288)	0.034 (0.076)

Table 4.3: Like Table 4.1, for the 21th. century stabilization run (20CS).

A1BS	u	$\hat{\sigma}$	$\hat{\xi}$	Δ_{stat}	Δ_{NAO}	$\hat{\beta}_1$
z_{1000}	360	72.43 (2.8)	-0.21 (0.024)	3.9 (0.122)	0 (0.878)	4.497 (1.737)
∇z	300	66.27 (2.04)	-0.138 (0.02)	0 (0.664)	1.4 (0.336)	1.091 (1.356)
D	450	62.75 (4.04)	-0.013 (0.044)	0 (0.712)	1.8 (0.288)	1.26 (2.71)
ζ_{850}	19	3.26 (0.11)	-0.149 (0.02)	0.5 (0.438)	0 (0.562)	-0.111 (0.071)

Table 4.4: Like Table 4.1, for the 22th. century stabilization run (A1BS).

model, but allowing for changes in both parameters outperforms all other models.

NAO dependance of extremes

For both runs (20CS and A1BS) strong support is found for a link to the NAO for z_{1000}, but with reduced strength in A1BS ($\hat{\beta}_1$ reaches 7.3 gpm in 20CS, but only 4.5 gpm in A1BS). That is, other parameters controlling z_{1000} are more important in A1BS than in 20CS. The results may suggest that the increased return levels in A1BS do not result from changed NAO regimes.

For D both data sets do not support a link. Differences are found for ∇z and ζ_{850}: While for ∇z a link is supported in 20CS only, the opposite is true for ζ_{850}. Note that the results for 20CS and 20C differ with respect to the NAO link, which is most probably due to the different sample size.

	LLR-test 1		LLR-test 2			AICD (AICW)			
	$\sigma(t)$	$\xi(t)$	Stat.	$\sigma(t)$	$\xi(t)$	Stat.	$\sigma(t)$	$\xi(t)$	$\sigma(t)$; $\xi(t)$
z_{1000}	0.002	0.057	0.001	0.329	0.008	7.7 (0.013)	0 (0.603)	6.1 (0.028)	1.1 (0.356)
∇z	0.006	0.025	0.006	0.78	0.114	5.4 (0.038)	0 (0.573)	2.4 (0.171)	1.9 (0.218)
D	0.259	0.635	0.025	0.053	0.029	1 (0.268)	1.7 (0.185)	2.8 (0.11)	0 (0.437)
ζ_{850}	0	0.001	0	0.966	0.003	18.6 (0)	0 (0.726)	9 (0.008)	2 (0.266)

Table 4.5: Statistical model building to test for differences between 20CS and A1BS. Estimated p-values according to log-likelihood ratio tests (LLR-test). LLR-test1 tests for significant improvement of the time dependant models ($\sigma(t)$ or $\xi(t)$) against the "stationary" model, while LLR-test2 tests the combined model ($\sigma(t)$ and $\xi(t)$ are time dependant) against the lower dimensional models ("stationary", $\sigma(t)$ or $\xi(t)$). Further, AIC differences (AICD) are given with corresponding AIC weights (AICW) in parentheses for all models in the set.

In contrast to the trend analysis (Section 4.1), differences are found between 20CS and A1BS cyclone extremes, indicating that a linear trend should be present in 20C. The trend on the chosen ERA40 time period, however, is too small to be detectable. Trend detection is further hampered through the presence of NAO, which increases the variability, but this problem is minimized through the method of statistical model building, which is able to separate different impacts on the extremes.

Life cycles and spatial occurrence of extremes

Similar to Figure 4.4, the five most extreme life cycles with respect to z_{1000} in the four parameters and the mean life cycles are shown in Figure 4.7 for 20CS and A1BS. If the mean life cycles in z_{1000} are considered, a deepening of the cyclones is found in A1BS. Additionally, the variability in z_{1000} increases. The five most extreme life cycles corroborate this finding by showing a decrease in the central geopotential height (higher z_{1000} values). Considering the mean life cycles in the gradient, ∇z, the relative vorticity, ζ_{850}, and the depth, D, yields opposite conclusions. While the maxima in the gradient and the vorticity decrease, the maxima in the depth are unchanged. Note that the life cycles with respect to z_{1000} in ∇z, ζ_{850} and D belong to the upper 5% of the cyclones.
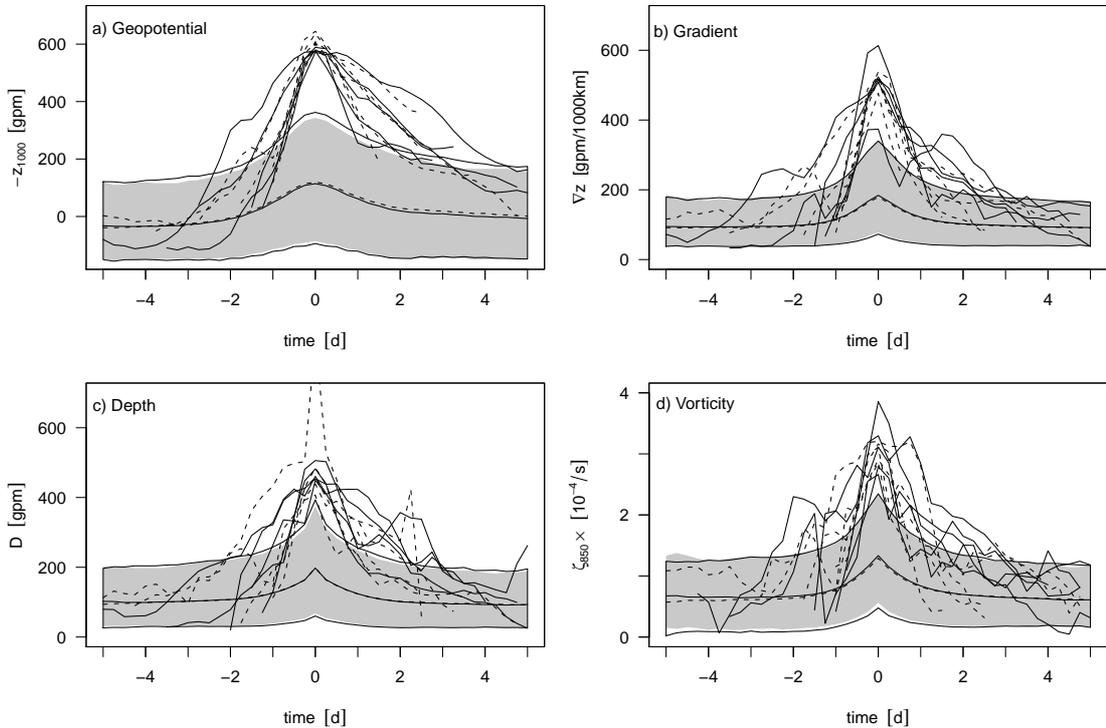

Figure 4.7: Like Figure 4.4, for the stabilization runs 20CS (solid) and A1BS (dashed).

Indicated by Figure 4.6 the extreme cyclones intensify in the warmer climate. Ensuring that this strengthening is not a consequence of a northward displacement of the cyclone tracks, spatial density estimates of the extreme cyclones are compared in Figure 4.8. The comparison between the density maxima in the geopotential height reveal no appreciable northward shift of the extreme cyclones between 20CS and A1BS. This is confirmed by spatial density differences yielding no systematic, but rather small and noisy departures (not shown). The density estimates of the other three parameters (ζ_{850}, ∇z and D) indicate a slight increase of the affected area with no northward displacement in connection with a smaller density maximum. Therefore, the intensification of the extreme cyclones found in a warmer climate scenario is supported.

4.4   Summary, conclusion and discussion

Extreme value statistics is applied to extremes of four parameters characterizing cyclone life cycles: central geopotential height (z_{1000}), mean geopotential height gradient (∇z), cyclone depth (D) and relative vorticity (ζ_{850}). The present-day climate of the coupled atmosphere ocean model ECHAM5/MPI-OM is compared with ERA40 re-analysis and with a global warming scenario (A1BS). The study focuses on extreme midlatitude cyclones in the North Atlantic during extended boreal winters (October-March). Additionally, the present-day data (ERA40, 20C) is analyzed with respect to a possible trend and a potential link to the North Atlantic Oscillation (NAO) is investigated.

In general, the model results are similar in terms of distributions, trends, life cycles and density estimates compared to ERA40. The extreme value statistics (return level plots) for z_{1000}, ∇z and ζ_{850} indicate reversed Weibull-type distributions, implying an upper bound. When analyzing the relatively short ERA40 time period, no significant linear trend is detected for ERA40 and 20C simula-
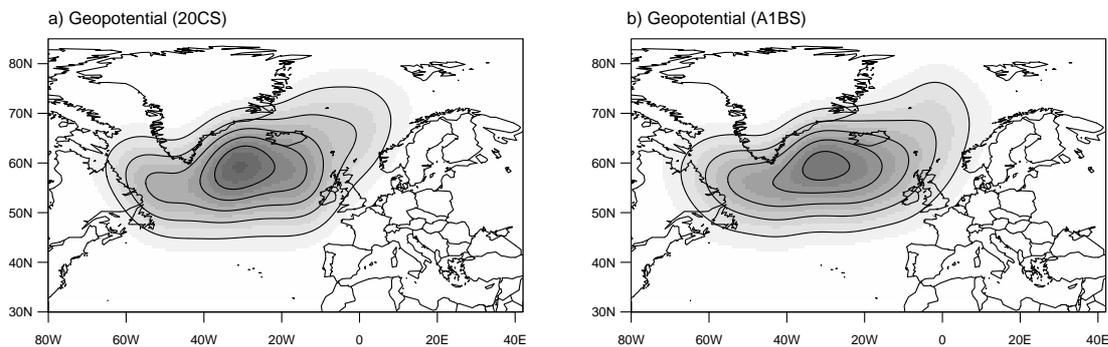
Figure 4.8: Spatial density estimates of the most extreme cyclones (values above thresholds) in the centered geopotential height, z_{1000} for 20CS (left) and A1BS (right).

tion. The use of a linear trend model is questionable under the assumption of a nonlinear forcing, but higher order polynomial or exponential trends can be approximated through linear models on short time spans. Life cycles of extreme cyclones respective to one cyclone parameter show that these cyclones also belong to the upper tails of the distributions in the other parameters for both, ERA40 and 20C. Density estimates of extreme cyclones in the four cyclone variables of 20C show similar spatial distributions as ERA40. For both data sets, the maximum density in vorticity is located more south than the maxima of the geopotential height.

Apart from these similarities, some differences between the data sets can be noted: The depth, D, shows a Gumbel-type distribution for 20C, while a Weibull-type is obtained for ERA40. In addition, 20C shows lower return levels. The threshold is rarely exceeded, which might be a direct consequence of the lower resolution. In 20C the link between NAO and z_{1000} is less evident than in ERA40. However, much stronger evidence is found in 20CS which may indicate that 40 years of data are not enough to give robust results.

The comparison between the two stabilization runs (20CS, A1BS) using extreme value statistics yields an intensification of the extreme cyclones in the North Atlantic region. The return levels show an increase for all parameters under consideration, but different in strength: smaller changes for gradient and vorticity and stronger for geopotential and depth. This intensification results in increased scale parameters for z_{1000}, ∇z and ζ_{850}. Additionally, the exceedance probability increases for z_{1000}. For D a decreased scale and increased shape parameter is found. Further analysis indicates that the deepening in z_{1000} is not a consequence of a northward shift, but could be attributed to changes in the mean sea level pressure (Fink et al., 2009).

Comparing this results with the findings of Della-Marta and Pinto (2009), some differences can be identified. In contrast to their work z_{1000} return values increase for return periods up to 10 years. However, the results of the vorticity agree. The difference could be attributed to the distinct analyzed time period (transient instead of stabilization model run) or the applied tracking algorithm. An extension is the analysis of GPD parameters. With the exception of D, each cyclone quantity (z_{1000}, ∇z and ζ_{850}) shows a higher scale parameter in A1BS. Demonstrating that significant different parameters do not necessarily yield a significant difference in return values.

Extreme cyclones and their trends have been analyzed in both model simulations and observations (re-analyses). However, a comparison with the results presented here is hampered by (i) different cyclone identification and tracking, (ii) different definitions of cyclone extremes and intensity measures, and (iii) different methodology to assess there statistics. Although for the entire Hemisphere, Ulbrich et al. (2009) already noted that dependant on the definition of extreme cyclones both an increase and a decrease in extreme cyclones can be found for the Northern Hemisphere: analyzing extreme cyclones in the upper percentage

of extremely low sea level pressure they detect an increase of extreme cyclones in a A1BS greenhouse warming scenario, while the definition of extremes by high values in the pressures Laplacian yields a decrease.

Despite the limited comparability the following similarities and differences with regard to recent studies might be noted: A link to NAO is confirmed by Pinto et al. (2009) showing that the number of extreme cyclones is enhanced in positive NAO phase in the NCEP/NCAR re-analyses. The absence of a significant trend in the cyclone parameters for the whole North Atlantic is consistent with the findings of Raible et al. (2008) (using NCEP/NCAR and ERA40 re-analyses). But, dividing the North Atlantic region in a high latitude and a midlatitude part, similar to the defined regions of Wang et al. (2008), Raible et al. (2008) find a dipole-like trend pattern in cyclone activity (in agreement with Wang et al. (2008)) and intensity.

Based on a multi-model perspective, there is a reduction in the total number of cyclones on the hemisphere in warmer climate simulations, whereas an increase is found in intense systems (Lambert and Fyfe, 2006). The decrease of the total number of cyclones in the Northern Hemisphere is supported by several other studies (Bengtsson et al., 2006; Finnis et al., 2007; Löptien et al., 2008; Pinto et al., 2009). Comparing twentieth century to warmer climate simulations, a northward shift of all detected cyclones is found. Cyclone density increases near the British Isles during the transient A1B scenario, while this increase is absent in the stabilization runs (not shown).

Consistent with the results presented here, Löptien et al. (2008) observe stronger deepening rates. However, they do not find significant changes in cyclone intensity (minimum central pressure during the cyclone life cycle). Deeper central core pressures of the 100 most intense cyclones were found in scenario A1BS, but interpreted as a result of a northward shift of cyclone tracks (Bengtsson et al., 2009).

An increased track density and intensity of extratropical cyclones is found near the British Isles in warmer climate scenarios (Ulbrich et al. (2009) and references therein). The same holds for extreme cyclones (Pinto et al., 2009) and is detected in multi-model analysis (Leckebusch et al., 2006).

Evaluating the present study concerning a possible change it is interesting to note that if all detected cyclones are used a strengthening can only be found for z_{1000}. The other parameters would indicate no change (D) or a weakening (∇z, ζ_{850}). This demonstrates potential advantages of analyzing the extremes directly. As presented here, extreme value statistics appears to be an appropriate and powerful method. In addition, accounting every cyclone only once to avoid intense and long living cyclones to be disproportionately weighted and to ensure independent data is preferable.

To supplement this study it may be useful to analyze and compare cyclones detected from the vorticity field and to ascertain the causes of the different positions of the density maxima found for the different cyclone parameters. Regarding the

cyclone variable D, one can only speculate why D differs in general for the other results. Here, further analyses are needed, for example, regarding the sensitivity to the particular definition of D.

Furthermore, possible links between North Atlantic cyclone extremes and variability modes other than NAO may be detectable (for example, El Niño / Southern Oscillation (ENSO) or the stratospheric circulation). Since it is possible to identify a significant change of extreme cyclones analyzing the A1BS stabilization run, but not within the ERA40 period, the question arises how sufficiently large the sample size should be for trend identification.

Considering the North Atlantic basin the determining factors for a cyclone to belong to an extreme event changes in A1B. Due to an enhanced low-level temperature gradient in the central North Atlantic, the low-level baroclinicity increase (Bengtsson et al., 2006). This region is also characterize by an eastward shift of the polar jet stream into Europe and increased upper air baroclinity (Pinto et al., 2007), which are related to extreme cyclones.

The present work can be integrated in the context of statistical analysis of cyclone extremes in re-analysis, present-day and future scenarios. Understanding the physical mechanisms behind the changes on cyclone extremes and the different behavior of cyclone variables on NAO are part of further analysis.

5 Extreme event return times in long-term memory processes near 1/f

The distribution of extreme event return times and their correlations are analyzed in observed and simulated long-term memory (LTM) time series with $1/f$ power spectra. The analysis is based on tropical temperature and mixing ratio (specific humidity) time series from TOGA COARE with 1 min resolution and an approximate $1/f$ power spectrum. Extreme events are determined by Peak-Over-Threshold (POT) crossing. The Weibull distribution represents a reasonable fit to the return time distributions while the power-law predicted by the stretched exponential for $1/f$ deviates considerably.

For a comparison and an analysis of the return time predictability, a very long simulated time series with an approximate $1/f$ spectrum is produced by a fractionally differenced (FD) process. This simulated data confirms the Weibull distribution (a power law can be excluded). The return time sequences show distinctly weaker long-term correlations than the original time series (correlation exponent $\bar{\gamma} \approx 0.56$).

5.1 Introduction

Long-term memory (LTM) is a ubiquitous phenomenon in natural time series and mainly identified by power-laws characterized by a single correlation exponent γ in the correlation function, $C(t) \sim t^{-\gamma}$ (Fraedrich and Blender, 2003). In many observed time series, predominantly sea surface temperatures, $1/f$ power spectra are found related to small γ (Weissman, 1988; Monetti et al., 2003). In the current discussion on anthropogenic climate change, the simulation of LTM becomes relevant since anthropogenic trends may be masked by low frequency internal variability (Blender and Fraedrich, 2003).

Even weak LTM (with γ slightly below 1) has considerable impacts on return times of extreme events (Altmann and Kantz, 2005; Eichner et al., 2007). An obvious reason for this effect is the clustering of threshold crossings during periods with high averages (Bunde et al., 2005). The distribution of return times t_r in the presence of LTM is approximately given by a stretched exponential, $p \sim \exp(-t_r^\gamma)$, where the exponent is assumed to be identical to the correlation exponent γ.

The stretched exponential is motivated by the study of Newell and Rosenblatt (1962) who derived an upper bound for the probability of no zero crossings in power-law correlated Gaussian processes. Olla (2007) applied an ϵ-expansion for $\gamma = 1 - \epsilon$ and obtained a stretched exponential distribution with exponent γ. Stretched exponential distributions are found for linear systems with LTM (Altmann and Kantz, 2005). There are classes of nonlinear dynamical systems which show algebraic (power-law) distributions (Zaslavsky, 2002). For inter-event distributions of earth quakes Corral (2004) suggests a gamma distribution. The long-term memory does not only alter the distribution of return times but also their temporal correlations which are the basis for the return time predictability (Bunde et al., 2004; Altmann and Kantz, 2005).

The present study is motivated by the abundance of observed nonstationary $1/f$ time series which are at the border of stationarity. The aim of this chapter is to analyze extreme event return time statistics in high resolution observations of tropical boundary layer temperature and humidity which both reveal a $1/f-$spectrum. The results are compared with a simulated time series obtained by a long simulation of a stationary fractionally differenced process (FD) with a power spectrum in the vicinity of the $1/f$. To evaluate potential predictability, long-term correlations of return times in this time series are estimated.

The chapter is organized as follows: In Section 5.2 LTM is defined and available results on return time distributions are summarized. The long term memory properties and the return time distributions of the observational data are determined in Section 5.3. In Section 5.4 simulated time series are compared and the correlation properties of the extreme event intervals are analyzed. The Section 5.5 concludes with a summary and discussion.

5.2   Estimating long-term memory and extreme event return time statistics

For the estimation of long-term memory (LTM, Beran, 1994) several methods are available. We compare results of the Detrended Fluctuation Analysis (DFA, Peng et al., 1994) with fits of FARIMA(p,d,0) processes (Hosking, 1981). The FARIMA processes are able to assess the contributions of short- and long-term components. The distribution of the extreme event return times is altered in the presence of LTM since long periods with anomalous low or high persistent deviations occur. The correlations between successive extreme event return times are useful for the prediction of extreme event return times.

5.2.1 Long-term memory analysis

A time series has long-term memory (LTM, also denoted as long-term persistence) if the correlation function $C(t)$ is not integrable (Beran, 1994). For a long-term power-law decay, $C(t) \sim t^{-\gamma}$, LTM is equivalent to $\gamma > 0$. Empirical time series have LTM if the autocorrelation follows a power-law with exponent $0 < \gamma < 1$ for the largest time scales present. LTM is ubiquitous in nature and shows up mainly in temperature records (Fraedrich and Blender, 2003; Huybers and Curry, 2006). The exponent β of the power spectrum, $S(f) \sim f^{-\beta}$, and the correlation exponent are related by $\beta = 1 - \gamma$, hence the power spectrum increases with decreasing frequency for $\gamma < 1$.

To determine LTM properties two methods are applied: (i) Detrended fluctuation analysis (DFA, Peng et al., 1994), and (ii) the estimation of the parameters in FARIMA(p,d,0) processes (Hosking, 1981). The two methods are independent complements for the analysis of our data and inhibit an erroneous detection of LTM: While there is a known LTM detection problem in the DFA in short term memory time series (Maraun et al., 2004), this method does not require any model assumption (for example normality of the data). The FARIMA process is ideal for the detection of short- and long-term memory, in addition, it allows a significance test for the number of parameters, however, normality of the data is required.

(i) The DFA determines fluctuations $F(\tau)$ on time scales τ in stationary anomaly sequences with LTM. Trends in the time series can be eliminated by extensions of the DFA (Fraedrich and Blender, 2003).

(ii) To assess the contributions of short- and long term memory components, fits of autoregressive processes (AR) and fractionally integrated autoregressive process are considered. In the following, FAR is used as a short notation for FARIMA(p,d,0) (Hosking, 1981) which includes an autoregressive (AR) process of order p and a fractionally differenced (FD) process with dimension d.

The AR process is defined by

$$\phi(B)x_t = \epsilon_t \tag{5.1}$$

where B is the backshift operator defined by $Bx_t = x_{t-1}$, and ϵ_t is white noise. Using the coefficients a_n, the polynomial $\phi(B)$ is

$$\phi(B)x_t = x_t - \sum_{n=1}^{p} a_n x_{t-n} \tag{5.2}$$

The FD process (Hosking, 1981) is derived from

$$(1 - B)^d x_t = \epsilon_t \tag{5.3}$$

and leads to an AR process of infinite order

$$x_t = \sum_{n=1}^{\infty} a_n x_{t-n} + \epsilon_t, \quad a_n = -\frac{\Gamma(n - d)}{\Gamma(-d)\Gamma(n + 1)} \tag{5.4}$$

For low frequencies the FD process shows a scaling power spectrum $S(f) \sim f^{-\beta}$ with spectral exponent $\beta = 2d$ and correlation exponent $\gamma = 1 - 2d$.

The FAR process is given by the combination

$$\phi(B)(1 - B)^d x_t = \epsilon_t \tag{5.5}$$

and is determined by p coefficients in the AR and the dimension d.

5.2.2　Extreme event return distributions

An extreme event in a time series x_i, $i = 1, \ldots, N$, crosses a given threshold q with $x_i > q$. The return time t_r between two extreme events is the time interval between two events with $x_i > q$ and $x_{i+t_r} > q$ and lower values $x_j < q$ in between, $i < j < i + t_r$. The mean return time R_q depends on the threshold q and is approximated by the probability distribution function (pdf) $D(x)$ of the time series

$$R_q^{-1} = \int_q^\infty D(x)dx \tag{5.6}$$

In the present chapter, the threshold q is determined to obtain a specific value of R_q. For uncorrelated data, the return times are exponentially distributed following a Poisson process

$$p_q(t_r) = \frac{1}{R_q} \exp(-t_r/R_q) \tag{5.7}$$

LTM leads to periods with anomalous persistent low or high deviations. During such periods extreme high values are either rare (for low anomalies) or frequent (during high anomalies). Thus return time statistics shows clustering which is not observed in time series without memory.

For LTM time series stretched exponential return time distributions are suggested (Bunde et al., 2004; Altmann and Kantz, 2005; Eichner et al., 2007)

$$p_q(t_r) \approx \frac{a_\gamma}{R_q} \exp[-(b_\gamma t_r/R_q)^\gamma] \tag{5.8}$$

Note that the scaling exponent γ is conjectured to be equal to the correlation exponent which characterizes LTM. The coefficients $a_\gamma = \gamma\Gamma(2/\gamma)/\Gamma^2(1/\gamma)$ and $b_\gamma = \Gamma(2/\gamma)/\Gamma(1/\gamma)$ are determined by normalization of p_q and the condition for the mean, $R_q = <t_r>$; Γ is the gamma-function.

In the limit $\gamma \to 0$, the stretched exponential approaches a power law

$$\log p_q(t_r) \sim -s \log t_r + const \tag{5.9}$$

with the exponent

$$s = \lim_{\gamma \to 0} \gamma b_\gamma^\gamma = 1.5 \tag{5.10}$$

Altmann and Kantz (2005) and Eichner et al. (2007) consider the correlation exponents $0.05 < \gamma < 1$ that is, between almost $1/f$ and white noise. Eichner et al. (2007) show that the stretched exponential is valid for several types of distributions D.

For small return times, $t_r \ll R_q$, the observed distribution deviates from the stretched exponential (5.8) and scales as (see Eqs. (10) in Eichner et al., 2007).

$$R_q p_q(t_r) \sim \left(\frac{t_r}{R_q}\right)^{s'} \tag{5.11}$$

with the proposed value $s' = \gamma' - 1$, $\gamma' \approx \gamma$ for Gaussian density. For large return times ($t_r \gg R_q$) the limit of the distribution (5.8) is

$$R_q p_q(t_r) \sim \exp[-(bt_r/R_q)^{\gamma}] \tag{5.12}$$

The stretched exponential is accepted as an approximate representation for linear LTM processes (Altmann and Kantz, 2005; Eichner et al., 2007).

An alternative to the stretched exponential distribution (5.8) is the Weibull distribution (Sornette, 2006; Abaimov et al., 2007) with the scale parameter τ and the shape parameter γ

$$p_W(t_r) = \frac{\gamma}{\tau} \left(\frac{t_r}{\tau}\right)^{\gamma-1} \exp[-(t_r/\tau)^{\gamma}] \tag{5.13}$$

Note that the Weibull distribution for $\gamma < 1$ is frequently denoted stretched exponential distribution; this, however, differs from (5.8) by the prefactor $\sim t^{\gamma-1}$. Without reference to Weibull, the power-law (5.11) is suggested by Eichner et al. (2007) in their Eq. (10) to correct the pure stretched exponential, $\sim \exp(-t^{\gamma})$, for small return times.

The advantages of the Weibull distribution for the characterization of extreme event return times are:

(i) The Weibull distribution (5.13) combines (5.8) and the short time limit (5.11) and describes the observed distribution in a wide range of return times.

(ii) The cumulative distribution function is known, $F_W(t_r) = 1 - \exp[-(t_r/\tau)^{\gamma}]$, and the mean recurrence time is determined by $R = \tau\Gamma(1 + 1/\gamma)$. This cumulative distribution function is useful for statistical analyses.

(iii) According to Sornette (2006) power-laws can be approximated by the Weibull distribution in arbitrary intervals to any prescribed accuracy.

The fit of the discrete power-law and Weibull distribution to the return time series is performed following Clausset et al. (2007). The approach fits the parameters of the distributions (exponents for power-laws; shape and scale parameters for Weibull) using Maximum Likelihood estimation and determines an

optimal range (restricted by a minimum return time cutoff) by minimizing the Kolmogorov-Smirnov distance.

The correlations between successive extreme event return times are one of the most useful aspects in practical applications of extreme value theory. Given time series with weak LTM (correlation exponents $\gamma = 0.4$ and 0.7), Bunde et al. (2004) analyze the respective return times arranged in a sequence, and find their long-term correlation exponents γ to be similar to the exponents of the original time series. It is expected that this relationship changes distinctly for very strong LTM due to its close vicinity to the nonstationarity threshold $1/f$.

In this chapter, extreme events are determined by the Peak-Over-Threshold (POT) method with different thresholds q, which are adjusted for mean return times R_q. The detrended fluctuation analysis (DFA) is employed to determine the LTM of the observational data and the recurrence times in the simulated data. Fits of FARIMA process support the LTM analysis of observational data. The time series are simulated by fractionally differenced processes (FD, Hosking (1981)). The fits of the power-law and the Weibull distributions are performed by the code available from Clausset et al. (2007). For all other calculations we use the statistics software R (R Development Core Team, 2005).

5.3   High resolution observational data

The observed time series analyzed in this study are obtained during the TOGA-COARE experiment (November 1992 – February 1993, Data Processing Center/Data Archive and Distribution Center for COARE Surface Meteorological Data, Florida State University, COARE-MET; Webster and Lukas (1992)). The aim of the international field experiment TOGA COARE during 1992-1993 was to study the atmospheric and oceanic processes over the western Pacific. The data measured at the Research Vessel (R/V) Kexue (3.9°S, 155.9°E) encompasses boundary layer near surface air temperature and the mixing ratio with one minute resolution (Figure 5.1); this data set has been corrected by Lucas and Zipser (2000). In the air temperature time series the diurnal cycle (daily mean with 1 min resolution) is removed for the analysis. The weak diurnal cycle in the mixing ratio is not removed since this does not change the result.

The mixing ratio (Figure 5.1 c)) reveals the presence of a large scale event during the first part of the time series (due to a passing 40-day wave). The fluctuations of the temperature and the mixing ratio are characterized by a $1/f$ power spectrum in a wide range of time scales (Yano et al., 2001; 2004). Only a part of this time series ($8.8 \cdot 10^4$ time steps, roughly 61 days) is analyzed to keep the number of missing values below $< 5\%$. The missing values are replaced by the mean and no attempt has been made to determine the effect of these replacements.
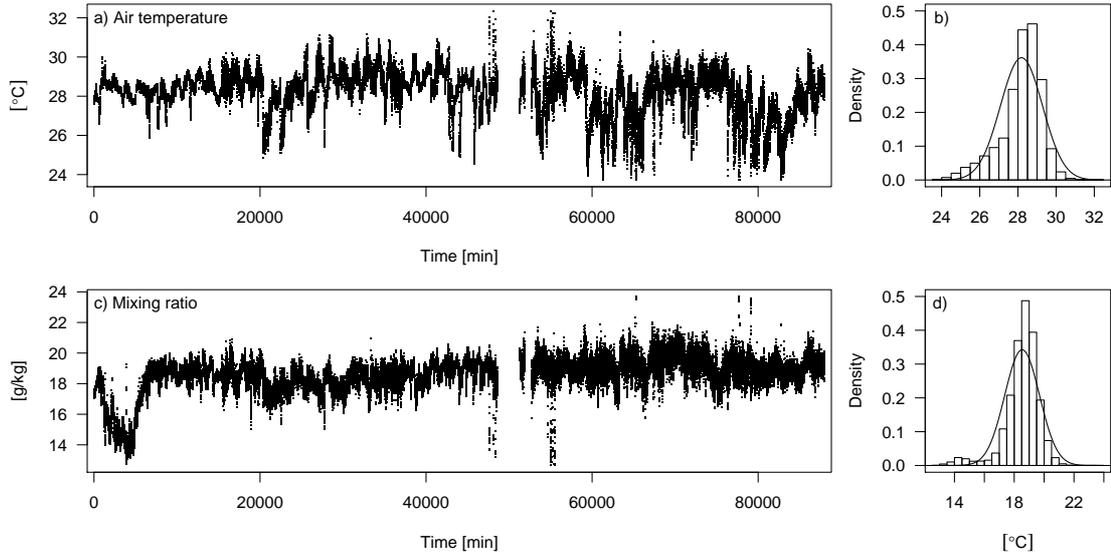
Figure 5.1: Observations of a) atmospheric near surface temperature and and c) mixing ratio; corresponding frequency distributions in b) and d).

The overall behavior of the data indicates nonstationarities in both time series. The frequency distributions for both time series (Figure 5.1 b) and d)) show deviations from Gaussian, which are, however, not substantial and presumably related to the nonstationarity.

5.3.1   Long term memory analysis

To determine LTM properties of the two observed time series two methods are applied (see Section 5.2.1): (i) Detrended fluctuation analysis (DFA, Peng et al., 1994), and (ii) an estimation of the parameters in FARIMA(p,d,0) process (Hosking, 1981).

(i) The DFA spectra in Figure 5.2 a) and b) show scaling fluctuation spectra, $F(\tau) \sim \tau^{\alpha}$, with exponents $\alpha \approx 1 \ldots 1.1$ close to a $1/f-$spectrum ($\alpha = 1$) for the temperature and the mixing ratio. The power spectrum is closely related to $F(\tau)$ and scales as $S(f) \sim f^{-\beta}$ with exponents $\beta = 2\alpha - 1 \approx 1 \ldots 1.2$; the correlation exponents are $\gamma = 1 - \beta \approx 0 \cdots - 0.2$. Note that the temperature fluctuation spectrum in Figure 5.2 a) approaches $\alpha \approx 1$ ($\gamma = 0$) for long time periods ($t_r > 10^3$ min) during two decades. An analysis of a trend-eliminating version of the DFA yields the same exponents.

(ii) Short- and long term memory contributions are assessed by fits of autoregressive processes (AR) and fractionally integrated autoregressive process (FAR); see Section 5.2.1. To determine the optimal number of parameters in the FAR
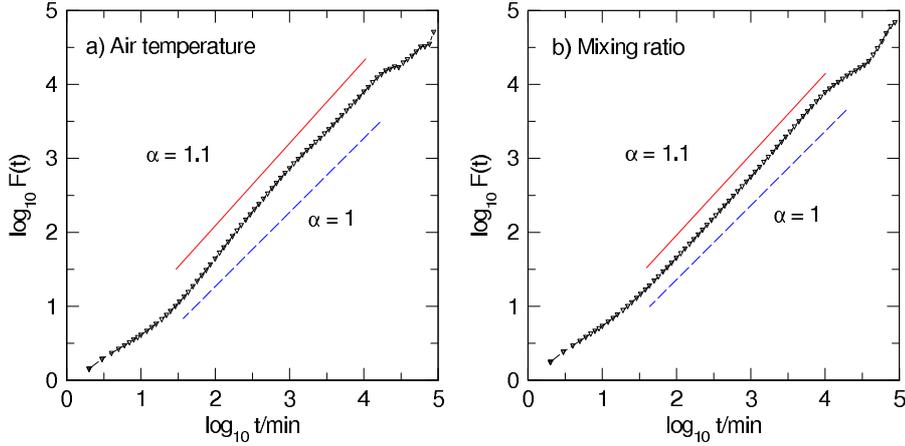
Figure 5.2: DFA fluctuation function of a) atmospheric near surface temperature and b) mixing ratio at R/V Kexue. The solid (red) lines indicates $\alpha = 1.1$, the dashed (blue) lines represents a $1/f-$spectrum ($\alpha = 1$).
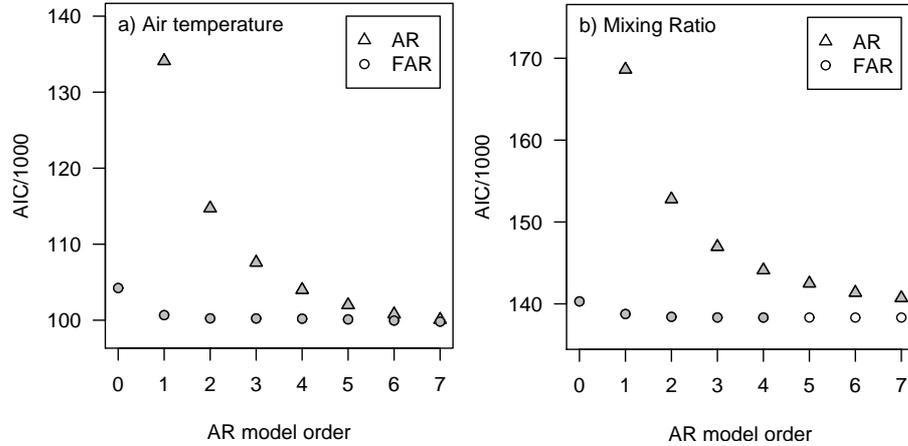
fit, the Akaike information criterion (AIC) is used based on the minimum of

$$\text{AIC} = -2\log(\mathcal{L}) + 2k \tag{5.14}$$

where $\mathcal{L}$ is the maximized likelihood function and k the number of estimated parameters. For temperature and mixing ratio it appears that the FAR process is superior to AR processes for small numbers of coefficients (Figure 5.3 a) and b)). The mixing ratio shows a higher preference for the FAR than temperature, which can be explained by the higher degree of scaling (Figure 5.2 b)). Furthermore, a maximum likelihood ratio test (99% significance) supports a lower degree of the autoregressive component in the FAR for the mixing ratio.

Likelihood ratio tests are performed to test whether higher order models give significant improvement compared to lower order models. FAR-models are tested against all (AR and FAR) lower order models, while the test for the AR-models is only performed for lower order. Filled symbols (Figure 5.3) show significance against lower order models on the 99% significance level.

For both observed data sets FAR-models outperform the ARs for all model orders below $p = 7$ according to the Akaike information criterion and the likelihood ratio test. Even for temperature (Figure 5.3 a)), where the information criterion looks quite similar for higher model orders, the likelihood ratio test prefers the FAR-models. However, the likelihood ratio test for temperature does not indicate an optimal model order. The mixing ratio (Figure 5.3 b)) can best be characterized by an FAR-model including four additional AR coefficients with the correlation exponent $\gamma \approx 5 \cdot 10^{-4}$. Since the model of choice is less clear for the temperature we consider the correlation exponents and their standard deviations. The correlation exponents decay from $\gamma = 0.016$ for $p = 0$ to $\gamma \approx 10^{-4}$ for

Figure 5.3: Akaike Information Criterion (AIC) for AR ($\triangle$) and FAR ($\circ$) models: a) air temperature and b) mixing ratio. Filled symbols show significance against lower order models according to likelihood-ratio tests.

$p \geq 2$; the standard deviations are below $\approx 2 \cdot 10^{-3}$. In summarizing we conclude that the spectra of both observed time series can be considered as $1/f$.

5.3.2 Return time distributions

The return time distributions $p_q(t_r)$ for temperature and mixing ratio are determined for the mean return time $R_q = 100$ (5.6). For the data the complementary cumulative distribution function (CCDF) is determined, $C_F(t_r) = 1 - F(t_r)$, where $F(t_r)$ is the cumulative distribution function. Scaling of the distribution function is preserved in the CCDF with an exponent reduced by 1. Unfortunately, a fit of the stretched exponential distribution is inhibited by insurmountable numerical difficulties.

The distributions in Figure 5.4 a) and b) are compared with the fits of Weibull distributions (5.13) and power-laws. The Weibull parameters, the power-law exponents, and the cutoffs are given in Table 5.1 together with confidence intervals, which are determined by resampling with replacement (1000 samples). For temperature and mixing ratio the power law exponent $s = 1.5$ lies outside the confidence intervals. The cutoffs are determined by minimizing the Kolmogorov-Smirnov test statistic. The power-law fits are compared with the power-law $s = 1.5$ predicted by the limit of the stretched exponential for $1/f$ noise (5.9, 5.10). The return time distributions for the two observed time series are reasonably well approximated by Weibull distributions in a wide range of return times. Note that the power-law fits are restricted to narrow ranges (in particular for the mixing ratio) and are obviously worse approximations for the observed distribu-
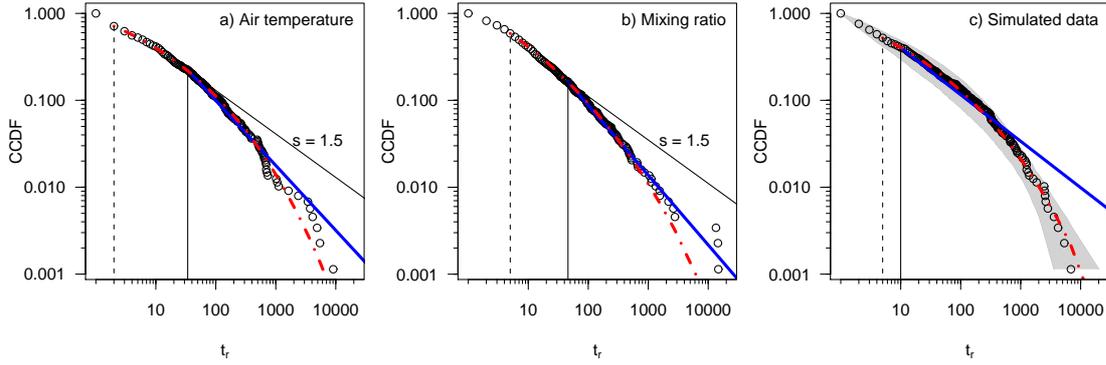
Figure 5.4: Complementary cumulative distribution functions for the return times in a) air temperature, b) mixing ratio, and c) $1/f$ simulated data. Dashed (red) curves denote Weibull distributions (5.13) and solid (blue) power-laws distributions. Vertical lines denote cutoffs k_{min}. The solid black lines denote the exponent 1.5 for the same cutoffs as the fits (this overlaps with the blue line in c)). In c) the 95% confidence interval is gray shaded.

tions. The power-law exponents for air temperature ($s = 1.74$) and mixing ratio ($s = 1.8$) differ substantially from 1.5, since this value is beyond the confidence intervals (Table 5.1).

5.4   Simulated data

Simulated time series with self-similar LTM are generated by a linear autoregressive process. As the AR part (5.2) is responsible for short memory, the simulated data is simulated by an FD process (5.4), see Figure 5.5 for the time series and the Gaussian frequency distribution. The power spectrum exponent is chosen as $\beta = 0.99$ ($d = \beta/2 = 0.495, \gamma = 0.01$) to obtain an approximation for a stationary time series with a $1/f$ power spectrum. To inhibit the impact of finite size effects in the comparison with observational data, the total length of the simulated time

	temperature	mixing ratio	simulated data
shape γ	0.21 (0.17, 0.27)	0.15 (0.11, 1.88)	0.19 (0.15, 0.25)
scale τ	0.36 (0.04, 1.94)	0.003 (10^{-5}, 0.04)	0.29 (0.01, 2.68)
exponent s	1.74 (1.66, 1.85)	1.80 (1.69, 1.94)	1.53 (1.49, 1.57)
cutoff k_{min}	4 (8)	5 (46)	5 (10)

Table 5.1: Values of: estimated parameters for Weibull and power-law distributions (with 95% confidence intervals), and cutoffs for the Weibull (power-law) distribution.
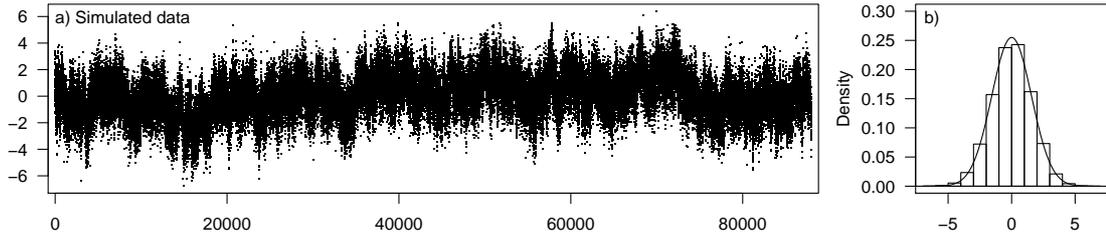
**Figure 5.5:** a) Simulated data (FD with $\gamma = 0.01$) and b) frequency distribution.

series is identical with that of the observed data ($N = 8.8 \cdot 10^4$). For the following analysis of the LTM in the return time sequence, however, a very long time series of $N = 10^8$ is simulated.

5.4.1 Return time distributions

The distribution $p_q(t_r)$ for the return times t_r is determined for a threshold q related to a mean return time $R_q = 100$. The analysis is analogous to Section 5.3.2. The distribution (Figure 5.4 c)) is well approximated by a Weibull distribution. The 95% confidence interval (gray shaded) is determined by creating 1000 time series with the same parameters. The shape and scale parameters are $\gamma = 0.2$ and $\tau = 0.29$, respectively. The validity of the power law fit, $p_q \sim t_r^{-s}$ with the exponent $s = 1.53$, is restricted to $t_r \approx 10 \ldots 500$. The deviation from $s = 1.5$, which is the $1/f$ limit of the stretched exponential, might originate in either: (i) The conjecture that the stretched exponential exponent (see Equation 5.8) is identical to the correlation exponent γ is not valid. (ii) The stretched exponential is not valid for very small $\gamma = 0.01$, i. e. near $1/f$.

5.4.2 Potential predictability of extreme event return times

For time series with weak LTM (correlation exponents $\gamma = 0.4, 0.7$, Bunde et al., 2004) the sequences $t_r(n)$ composed of extreme event return times show long-term correlations with similar LTM as the time series itself. To analyze this behavior in the vicinity of $1/f$ noise, the correlation exponent is $\gamma = 0.01$ (as in Section 5.4.1) and the extreme events are based on different thresholds providing the mean return times $R_q = 10, 100, 1000, 10000$. The total length of the time series is $N = 10^8$. Note that this is two decades longer than $N = 2^{21} \approx 2.1 \cdot 10^6$ in Eichner et al. (2007). The sequence of the return times is analyzed by detrended fluctuation analysis (DFA). Figure 5.6 shows that the long-term correlation of the return times is described by a power-law fluctuation function $F(n) \sim n^\alpha$ with $\alpha \approx 0.72$ independent of the threshold and the mean return time; the index n enumerates the return times. This corresponds to the power spectrum exponent $\beta \approx 0.44$ (using $\beta = 2\alpha - 1$) and the correlation exponent $\gamma \approx 0.56$. Thus the
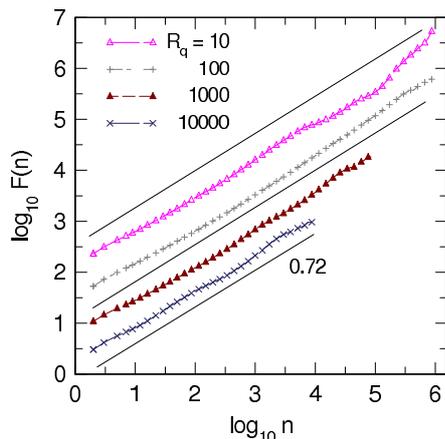
Figure 5.6: DFA fluctuation functions for the sequence of return times obtained with a FD ($\gamma = 0.01$) and different thresholds corresponding to the mean return times R_q as indicated. n enumerates the return time sequence. The solid lines show the LTM exponent $\alpha = 0.72$.

sequence of extreme events in time series near the $1/f-$limit shows distinctly weaker long-term correlation properties than the original time series.

5.5   Conclusions

This chapter presents an analysis of the extreme event return time statistics for observed and simulated data with $1/f$ power spectra. The observed data is given by measurements of temperature and mixing ratio during TOGA-COARE (November 1992 – February 1993) at the research vessel Kexue. In the time series of one minute resolution, 61 days with low numbers of missing values are extracted. Both time series show a scaling power spectrum, $S(f) \sim f^{-\beta}$, with $\beta = 1 \ldots 1.2$; the correlation exponent in $C(t) \sim t^{-\gamma}$ is related by $\gamma = 1 - \beta$. This result is determined by detrended fluctuation analysis and substantiated by a fit of a FARIMA(p,d,0) fractionally differenced autoregressive process which yields $d \approx 0.5$ for the long-term behavior ($\beta = 2d$). Hence, both time series are considered as $1/f$ noise.

Extreme events are determined by Peak-Over-Threshold (POT) crossing. The observed return time distributions $p_q(t_r)$ are compared to a stretched exponential, $\sim \exp(-t^\gamma)$, and a Weibull distribution, $\sim t^{\gamma-1} \exp(-t^\gamma)$. According to the approach by Altmann and Kantz (2005) and Eichner et al. (2007), the stretched exponential distribution converges to a power-law $p_q(t_r) \sim t_r^{-s}$ with $s = 1.5$ for $\gamma \to 0$.

The return time distributions for the two observed time series are better approximated by a Weibull distribution than by a power-law. If a power-law is

fitted in the intermediate range of return times, the temperature yields a power-law exponent $s = 1.74$, while the mixing ratio yields $s = 1.8$; both are distinctly different from the stretched exponential limit.

Simulated data is generated by a fractionally differenced autoregressive process with a power spectrum in the vicinity of the stationarity threshold, $\beta = 0.99$ ($\gamma = 0.01$). As for the observational data, the Weibull distribution yields a convincing representation of the return time distributions, while a power-law can be excluded.

The simulated data is used to evaluate the potential predictability of the extreme event return times. The LTM in the sequence of return times is analyzed by the detrended fluctuation analysis and reveals a power law fluctuation function $F(n) \sim n^\alpha$ with $\alpha \approx 0.72$ (correlation exponent $\bar{\gamma} \approx 0.56$). Thus, the return time sequences show weaker long-term correlations than the original time series. This values is independent of the threshold and the mean return time. A possible reason is that short term random effects lead to level crossings which, thereby, perturb the overall LTM of the original time series.

The analysis leads to the following main conclusions for the behavior of extreme event return times in the $1/f$ limit:

(i) The return time distributions for time series in the vicinity of a $1/f$ power spectrum are well approximated by the standard Weibull distribution. This is suggested by the observed time series and substantiated by the simulated data. The stretched exponential (which differs from Weibull by the absence of a pre-factor) is likely to be convenient for weak LTM.

(ii) The sequence of return times shows LTM with $F(n) \sim n^\alpha$ with $\alpha \approx 0.72$ which is weaker than in the original time series ($\alpha = 0.995$). However, the correlation $C(n) \sim n^{-0.56}$ is still promising for the prediction of return times.

Future analyses should consider the Weibull distribution as an alternative to the stretched exponential return time distribution for a wide range of LTM correlation coefficients γ.

6 Summary and conclusions

Different aspects of extreme events are considered. This involes their description, as well as their relation to climate variability and their potential intensificaion under changing climate conditions. The influence of long-term dependence on the extreme event return times is investigated. The main findings are restated.

Standardised Precipitation Index (SPI)

The SPI is introduced conceptual in Chapter 2. An extension is presented to apply the SPI in climate change scenarios. This enables the detection of extreme dryness and wetness changes due to greenhouse gas forcing. Problems going along with the distribution function of precipitation are noted briefly in Chapter 2 and discussed in detail in Chapter 3. The Gamma distribution (GD) is found the less flexible one, leading to systematic overestimation (underestimation) of extreme dryness (wetness). Solutions are presented to minimize this bias. That are: multiple distributions or a more general one, such as the exponentiated Weibull (EWD) or generalized Gamma (GGD) distribution can be used for the SPI calculation. The consideration of comparability, reproducibility and bias-variance adjustement guides the decision for the preferred method, depending on the situation at hand.

The SPI is used to analyse extreme dryness and wetness in Iceland (Chapter 2). These extremes depend on a distinct dipol pattern in the geopotential height field. The Europe-Greenland Index is derived from the geopotential height differences between the two poles and therewith the extremes are related to the sign of the index. Severe and extreme wet conditions occur significant more often in the warmer climate scenario of the 21th century (A1B). The increase is seasonally different, which results from intensified (reduced) flow anomalies.

Extreme value statistics

Extreme value statistics is applied to North Atlantic cyclones in Chapter 4. Generalized Pareto distributions are analysed in terms of return levels and the parameters of the distribution to detect changes in the cyclone extremes. A method based on covariates is proposed for the identification of differenences between climate change scenarios. This method enables the application of likelihood-ratio

test or information criteria (AIC) for the distribution parameters between two climate states. It is demonstrated that, in situtaions where return level changes are small and not significant, it is still possible to detect significant differences in the parameters, if this approach is applied.

Different measures are analysed, which are related to cyclone intensity: geopotential height, mean horizontal gradient, cyclone depth and relative vorticity. The extremes of the chosen cyclone properties depend on the NAO, with exception of the cyclone depth. In re-analysis data (ERA40) no significant linear trend is detected. However, all variables show an intensification in the warmer climate scenario (A1B).

Return time analysis

The presence of long-term memory in time series alters the occurrence of extremes. The extremes tend to cluster in time, depending on the the strength of memory. One characteristic of long-term memory processes is a power-spectrum, that follows a power-law. If the strength of memory increases, the spectrum approaches the limit $1/f$ (f, the frequency). Processes with this property are called $1/f$ noise. Observed air temperature and mixing ratio time series in high temporal resolution exhibit spectra sufficiently close to the given limit and can be considered as $1/f$ noise (Chapter 5). These time series, together with simulated ones, reproducing their properties are investigated in respect to their extreme event return times.

It is shown, that the return time distributions follow a discrete Weibull distribution. This is in contrast to preceding studies of other authors, who proposed a power-law distribution for the $1/f$ limit. However, a direct comparison yields improved estimates for the Weibull, whereas a power-law distribution deviates substantially. Further, the return time series shows distinctly weaker long-term correlation properties than the original time series.

Bibliography

Abaimov, S. G., D. L. Turcotte, R. Shcherbakov, and J. B. Rundle, 2007: Recurrence and interoccurrence behavior of self-organized complex phenomena. *Nonlin. Processes Geophys.*, **14**, 455–464.

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19 (6)**, 716– 723.

Alexander, L. V. and P. D. Jones, 2001: Updated precipitation series for the U.K. and discussion of recent extremes. *Atmos. Sci. Lett.*, **1**, 142–150.

Altmann, E. G. and H. Kantz, 2005: Recurrence time analysis, long-term correlations and extreme events. *Phys. Rev. E*, **71**, 056 106.

Beck, C., J. Grieser, and B. Rudolf, 2005: A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. Climate status report 2004, German Weather Service, Offenbach, Germany. 181 - 190.

Benedict, J. J., S. Lee, and S. B.Feldstein, 2004: Synoptic view of the North Atlantic Oscillation. *J. Atmos. Sci.*, **61**, 121–144.

Bengtsson, L., K. I. Hodges, and N. Keenlyside, 2009: Will extratropical storms intensify in a warmer climate? *J. Clim.*, **22**, 2276–2301.

Bengtsson, L., K. I. Hodges, and E. Roeckner, 2006: Storm tracks and climate change. *J. Clim.*, **19**, 3518–3543.

Beran, J., 1994: *Statistics for long-memory processes*. Chapman & Hall/CRC Press, London, 315 pp.

Blender, R. and K. Fraedrich, 2003: Long time memory in global warming simulations. *Geophys. Res. Lett.*, **30 (14)**, 1769–1772.

Blender, R., K. Fraedrich, and F. Lunkeit, 1997: Identification of cyclone-track regimes in the North Atlantic. *Q. J. R. Meteorol. Soc.*, **123**, 727 – 741.

Blender, R., K. Fraedrich, and F. Sienz, 2008: Extreme event return times in long-term memory processes near $1/f$. *Nonlin. Processes Geophys.*, **15**, 557–565.

Bordi, I., K. Fraedrich, M. Petitta, and A. Sutera, 2006: Extreme value analysis of wet and dry periods in sicily. *Theor. Appl. Climatol.*, **87**, 61–71.

Bordi, I. and A. Sutera, 2001: Fifty years of precipitation: Some spatially remote teleconnections. *Water Resour. Manag.*, **15**, 247–280.

Bunde, A., J. F. Eichner, S. Havlin, and J. W. Kantelhardt, 2004: Return intervals of rare events in records with long-term persistence. *Physica A*, **342**, 308–314.

Bunde, A., J. F. Eichner, J. W. Kantelhardt, and S. Havlin, 2005: Long-term memory: A natural mechanism for the clustering of extrem events and anomalous residual times in climate records. *Phys. Rev. Lett.*, **94**, 048 701.

Burke, E. J. and S. J. Brown, 2008: Evaluating uncertainties in the projection of future drought. *J. Hydrometeorol.*, **9**, 292–299.

Burnham, K. P. and D. R. Anderson, 2002: *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach*. 2d ed., New York: Springer-Verlag.

Cassou, C., L. Terray, J. W. Hurrell, and C. Deser, 2004: North atlantic winter climate regimes: spatial asymmetry, stationarity with time, and oceanic forcing. *J. Clim.*, **17 (5)**, 1055–1068.

Clauset, A., C. R. Shalizi, and M. E. J. Newman, 2007: Power-law distributions in empirical data. URL `http://arxiv.org/abs/0706.1062`.

Coles, S., 2003: *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, New York, 224 pp.

Coles, S. and L. Pericchi, 2003: Anticipating catastrophes through extreme value modeling. *Appl. Statist.*, **52 (4)**, 405–416.

Coles, S., L. R. Pericchi, and S. Sisson, 2003: A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, **273 (1-4)**, 35–50.

Corral, A., 2004: Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Phys. Rev. Lett.*, **92**, 108 501.

Davison, A. C. and D. V. Hinkley, 2003: *Bootstrap methods and their application*. Cambridge Univ. Press.

Della-Marta, P. M. and J. G. Pinto, 2009: Statistical uncertainty of changes in winter storms over the North Atlantic in an ensemble of transient climate simulations. *Geophys. Res. Lett.*, **36**, L14 703.

Dress, H., 2008: Some aspects of extreme value statistics under serial depenence. *Extremes*, **11**, 35–53.

Eichner, J., J. W. Kantelhardt, A. Bunde, and S. Havlin, 2007: Statistics of return intervals in long-term correlated records. *Phys. Rev. E*, **75**, 011 128.

Fink, A. H., T. Brücher, V. Ermert, A. Krüger, and J. G. Pinto, 2009: The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Nat. Hazards Earth Syst. Sci.*, **9**, 405–423.

Finnis, J., M. M. Holland, M. C. Serreze, and J. J. Cassano, 2007: Response of Northern Hemisphere extratropical cyclone activity and associated precipitation to climate change, as represented by the Community Climate System Model. *J. Geophys. Res.*, **112**, G04S42.

Fisher, R. A., 1922: On the mathematical foundations of theoretical statistics. *Philos. T. Roy. Soc. A*, **222**, 309–368.

Fisher, R. A. and L. H. C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180–190.

Fraedrich, K., 1994: ENSO impact on Europe? - A review. *Tellus*, **46A**, 541–552.

Fraedrich, K. and R. Blender, 2003: Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.*, **90**, 108 501.

García-Herrera, R., D. Paredes, R. M. Trigo, I. F. Trigo, E. Hernández, D. Barriopedro, and M. A. Mendes, 2006: The outstanding 2004/05 drought in the Iberian Peninsula: Associated atmospheric circulation. *J. Hydrometeorol.*, **8**, 483–498.

Groisman, P. Y., et al., 1999: Changes in the probability of heavy precipitation: Important indicators of climate change. *Climate Change*, **42 (1)**, 243–283.

Gulev, S. K., O. Zolina, and S. Grigoriev, 2001: Extratropical cyclone variability in the Northern Hemisphere winter from the NCEP/NCAR reanalysis data. *Clim. Dyn.*, **17**, 795 – 809.

Guttman, N. B., 1999: Accepting the standardized precipitation index. *J. Am. Water Resour. As.*, **35 (2)**, 311–322.

Hanna, E., T. Jónsson, and J. E. Box, 2004: An analysis of icelandic climate since the nineteenth century. *Int. J. Climatol.*, **24 (10)**, 1193–1210.

Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.*, **122**, 2573 –2586.

Hosking, J. R. M., 1981: Fractional differencing. *Biometrica*, **68 (1)**, 165–176.

Hosking, J. R. M., 1994: The four-parameter kappa distribution. *IBM J. Res. Dev.*, **38 (3)**, 251 – 258.

Huybers, P. and W. Curry, 2006: Links between annual, milankovitch and continuum temperature variability. *Nature*, **441**, 329–332.

Jung, T., S. K. Gulev, I. Rudeva, and V. Soloviov, 2006: Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model. *Q. J. R. Meteorol. Soc.*, **132**, 1839 – 1857.

Jungclaus, J. H., et al., 2006: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *J. Clim.*, **19**, 3952 – 3972.

Katz, R. W. and B. G. Brown, 1992: Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, **21**, 289–302.

Katz, R. W., M. B. Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Adv. Water Resour.*, **25**, 1287–1304.

Kharin, V. V. and F. W. Zwiers, 2000: Changes in the extremes in an ensamble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Clim.*, **13**, 3760–3788.

Kharin, V. V. and F. W. Zwiers, 2005: Estimating extremes in transient climate simulations. *J. Clim.*, **18**, 1156–1173.

Klein Tank, A. M., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Climate Data and Monitoring WCDMP-No. 72 (WMO-TD No. 1500), World Meteorological Institution.

Klein Tank, A. M. G. and G. P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946-99. *J. Clim.*, **16 (22)**, 3665–3680.

Klein Tank, A. M. G. et al., 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment. *Int. J. Climatol.*, **22**, 1441–1453.

Kysely, J. and J. Picek, 2007: Probability estimates of heavy precipitation events in flood-prone central-european region with enhanced influence of mediterranean cyclones. *Adv. Geosci.*, **12**, 43–50.

Lambert, S. J. and J. C. Fyfe, 2006: Changes in winter cyclone frequencies and strengths simulated in enhanced greenhouse warming experiments: results from the models participating in the IPCC diagnostic exercise. *Clim. Dyn.*, **26**, 713–728.

Lawless, J. F., 1982: *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons, Inc., New York, 580 pp.

Leckebusch, G. C., B. Koffi, U. Ulbrich, J. G. Pinto, T. Spangehl, and S. Zacharias, 2006: Analysis of frequency and intensity of European winter storm events from a multi-model perspective, at synoptic and regional scales. *Clim. Res.*, **31**, 59–74.

Lloyd-Hughes, B. and M. A. Saunders, 2002: A drought climatology for europe. *Int. J. Climatol.*, **22**, 1571–1592.

Löptien, U. and E. Ruprecht, 2005: Effect of synoptic systems on the variability of the North Atlantic Oscillation. *Mon. Wea. Rev.*, **133**, 2894 – 2904.

Löptien, U., O. Zolina, S. Gulev, M. Latif, and V. Soloviov, 2008: Cyclone life cycle characteristics over the Northern Hemisphere in coupled GCMs. *Clim. Dyn.*, **31**, 507–532.

Lucas, C. and E. J. Zipser, 2000: Environmental variability during TOGA COARE. *J. Atmos. Sci.*, **57**, 2333–2350.

Maraun, D., H. W. Rust, and J. Timmer, 2004: Tempting long-memory - on the interpretation of DFA results. *Nonlin. Processes Geophys.*, **11**, 495–503.

Marsland, S. J., H. Haak, J. H. Jungclaus, M. Latif, and F. Röske, 2003: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Model.*, **5 (2)**, 91–127.

McKee, T. B., N. J. Doeskin, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. *8th Conf. on applied climatology*, American Meteorological Society, Anaheim, Canada, 179–184.

Mielke, P. W. and E. S. Johnson, 1974: Some geralized Beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resour. Res.*, **10 (2)**, 223–226.

Mitchell, T. D. and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25 (6)**, 693–712.

Monetti, R. A., S. Havlin, and A. Bunde, 2003: Long-term persistence in the sea surface temperature fluctuations. *Physica A*, **320**, 581–589.

Murray, R. and I. Simmonds, 1991: A numerical scheme for tracking cyclone centers from digital data. part i: Development and operation of the scheme. *Aust. Meteor. Mag.*, **39**, 155–166.

Nakićenović, N. J. et al., 2000: *IPCC Special Report on Emissions Scenarios.* Cambridge University Press, Cambridge, United Kingdom and New York, USA, 599 pp.

Newell, G. F. and M. Rosenblatt, 1962: Zero crossing probabilities for gaussian stationary processes. *Ann. Math. Stat.*, **33**, 1306.

Olla, P., 2007: Return times for stochastic processes with power-law scaling. *Phys. Rev. E*, **76**, 011 122.

Palmer, W. C., 1965: Meteorological drought. Research paper No. 45, US Weather Bureau, Washington, DC.

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, 1994: On the mosaic organization of DNA sequences. *Phys. Rev. E*, **49**, 1685–1689.

Pinto, J. G., U. Ulbrich, G. C. Leckebusch, T. Spangehl, M. Reyers, and S. Zacharias, 2007: Changes in storm track and cyclone activity in three SRES ensemble experiments with the ECHAM5/MPI-OM1 GCM. *Clim. Dyn.*, **29**, 195–210.

Pinto, J. G., S. Zacharias, A. H. Fink, G. C. Leckebusch, and U. Ulbrich, 2009: Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Clim. Dyn.*, **32**, 711–737.

R Development Core Team, 2009: *R: A language and environment for statistical computing.* Vienna, Austria, R Foundation for Statistical Computing, URL `http://www.R-project.org`.

Raible, C. C., P. M. Della-Marta, C. Schwierz, H. Wernli, and R. Blender, 2008: Northern Hemisphere extratropical cyclones: A comparison of detection and tracking methods and different reanalyses. *Mon. Wea. Rev.*, **136**, 880–897.

Reeve, D. E., 1996: Estimation of extreme indian monsoon rainfall. *Int. J. Climatol.*, **16**, 105–112.

Rodriguez, R. N., 1977: A guide to the Burr type XII distribution. *Biometrika*, **64 (1)**, 129–134.

Roeckner, E., et al., 2003: The atmopheric general circulation model ECHAM5, part 1: Model description. Report 349, Max-Planck-Institut for Meteorology, Hamburg, Germany. 127 pp.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336.

Schneidereit, A., R. Blender, and K. Fraedrich, 2010: Radius-depth model for midlatitude cyclones in re-analysis data and simulations. *Q. J. R. Meteor. Soc.*, **136**, 50–6.

Schneidereit, A., R. Blender, K. Fraedrich, and F. Lunkeit, 2007: Iceland climate and North Atlantic cyclones in ERA40 reanalysis. *Meteorol. Z.*, **16**, 17–23.

Serreze, M. C., F. Carse, R. G. Barry, and J. C. Rogers, 1997: Icelandic low cyclone activity: climatological features, linkages with the NAO, and relationships with recent changes in the northern hemisphere circulation. *J. Clim.*, **10 (3)**, 453–464.

SFB-512, 2005: Cyclones and the North Atlantic climate system. *Meteorol. Z.*, **14 (6)**, 693–821.

Sickmöller, M., R. Blender, and K. Fraedrich, 2000: Observed winter cyclone tracks in the Northern Hemisphere in re-analysed ECMWF data. *Q. J. R. Meteor. Soc.*, **126**, 591 – 620.

Sienz, F., I. Bordi, K. Fraedrich, and A. Schneidereit, 2007: Extreme dry and wet events in Iceland: Observations, simulations and scenarios. *Meteorol. Z.*, **16 (1)**, 9–16.

Sienz, F., A. Schneidereit, R. Blender, K. Fraedrich, and F. Lunkeit, 2010: Extreme value statistics for North Atlantic cyclones. *Tellus A*, accepted.

Simmons, A. J. and J. K. Gibson, 2000: The ERA40 project plan. ERA40 Report Series 1, European Center for Medium-Range Weather Forcasts, Reading, United Kingdom. 66 pp.

Sornette, D., 2006: *Critical phenomena in natural sciences*. Springer Series in Synergetics, Springer-Verlag, 528 pp.

Tadikamalla, P. R., 1980: A look at the Burr and related distributions. *Int. Stat. Rev.*, **48 (3)**, 337–344.

Ulbrich, U., G. C. Leckebusch, and J. G. Pinto, 2009: Extra-tropical cyclones in the present and future climate: a review. *Theor. Appl. Climatol.*, **96**, 117–131.

Uppala, S. M., et al., 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131 (612)**, 2961 – 3012.

Wang, X. L., V. R. Swail, and F. W. Zwiers, 2008: Climatology and changes of the extratropical cyclone activity: comparison of ERA-40 with NCEP/NCAR reanalysis for 1958-2001. *J. Clim.*, **19**, 3145 – 3166.

Webster, P. and R. Lukas, 1992: TOGA COARE: the coupled ocean-atmosphere response experiment. *Bull. Amer. Meteor. Soc.*, **73**, 1377–1416.

Weissman, M. B., 1988: 1/f noise and other slow, nonexponential kinetics in condensed matter. *Rev. Mod. Phys.*, **60**, 537–571.

Wu, H., M. J. Hayes, D. A. Wilhite, and M. D. Svoboda, 2005: The effect of the length of record on the standardized precipitation index calculation. *Int. J. Climatol.*, **25**, 505–520.

Wu, H., M. D. Svoboda, et al., 2007: Appropriate application of the standardized precipitation index in arid locations and dry seasons. *Int. J. Climatol.*, **27**, 65–79.

Yano, J.-I., R. Blender, C. Zhang, and K. Fraedrich, 2004: 1/f-noise and pulse-like events in the tropical atmospheric surface variabilities: Convective self-criticality and impacts on ENSO. *Quart. J. Roy. Meteorol. Soc.*, **130**, 1697–1721.

Yano, J.-I., K. Fraedrich, and R. Blender, 2001: Tropical convective variability as 1/f-noise. *J. Climate*, **14**, 3608–3616.

Zaslavsky, G. M., 2002: Chaos, fractional kinetics, and anomalous transport. *Physics Reports*, **371**, 461–580.

Zhang, X., F. Zwiers, and G. Li, 2004: Monte carlo experiments on the detection of trends in extreme values. *J. Clim.*, **17**, 1945–1952.

Considerable parts of this dissertation are already published in scientific journals. This concerns the Chapters 2, 3 and 5, representing contents of the following articles:

- Sienz F., I. Bordi, K. Fraedrich and A. Schneidereit, 2007: Extreme dry and wet events in Iceland: Observations, simulations and scenarios. *Meteorologische Zeitschrift*, **16**, 9-16.

- Sienz F., A. Schneidereit, R. Blender, K. Fraedrich und F. Lunkeit, 2010: Extreme value statistics for North Atlantic cyclones. *Tellus A*, accepted.

- Blender R., K. Fraedrich and F. Sienz, 2008: Extreme event return times in long-term memory processes near $1/f$. *Nonlinear Processes in Geophysics*, **15**, 557-565.

Chapter 3 is intended for publication shortly:

- Sienz F., R. Blender, O. Bothe und K. Fraedrich, 2010: Implications for drought monitoring and projection - SPI bias, in preparation.

Acknowledgments