Trends and Extreme Values of River Discharge Time Series

Dipl.-Syst. Wiss. Malaak Kallache

Potsdam, April 2007

Dissertation zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
Eingereicht im Fachbereich Geowissenschaften
der Fakultät II – Biologie, Chemie und Geowissenschaften
der Universität Bayreuth, Deutschland

Dissertation submitted for obtaining the degree of
Doctor rerum naturalium
in the department of Geosciences
at the faculty of Biology, Chemistry and Geosciences
of the University of Bayreuth, Germany

Vollständiger Abdruck der von der Fakultät Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades Doktor der Naturwissenschaften (Dr. rer. nat.).

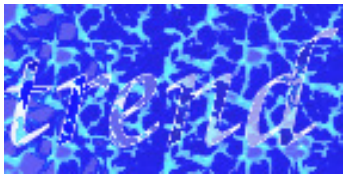
Die vorliegende Arbeit wurde in der Zeit vom November 2002 bis November 2007 am Lehrstuhl für Ökologische Modellbildung der Universität Bayreuth angefertigt.

Die Arbeit wurde eingereicht am 16. April 2007

Das Rigorosum fand statt am 02. November 2007

Der Prüfungsausschuss bestand aus:

Prof. Dr.	C. Beierkuhnlein	(Vorsitzender)
Prof. Dr.	H. Lange	(Erstgutachter)
Prof. Dr.	H.J. Schellnhuber	(Zweitgutachter)
Prof. Dr.	M. Hauhs	
PD Dr.	G. Lischeid	



.....you cannot step twice into the same river
Heraclitus

Contents

List of Figures

ABSTRACT

The central goal of this thesis is the development and application of sound data analysis frameworks to assess trends in mean and extreme river run-off. This includes the consideration of the auto-correlation structure (memory) of the data. Such an approach is very useful to, e.g., assess the anticipated intensification of the hydrological cycle due to anthropogenic climate change. The costs related to more frequent or more severe floods and droughts are enormous. Therefore an adequate estimation of these hazards and the related uncertainties is of major concern. We analyse discharge of basins of the Danube River and Neckar River in Southern Germany. Thereby we compare our frameworks with methods common in hydrology with respect to the calculation of assessment measures used by water management authorities. Furthermore, we look for spatial patterns of trends.

In the first part of this thesis, we evaluate trends of average discharge data. We use a trend test, where the auto-correlation structure of the data is explicitly modelled using stochastic FARIMA processes. This is a crucial task, because auto-correlations are capable of producing spurious trends. We assume that the data consists of an additive combination of natural variability, which is represented by the stochastic process, and potentially a deterministic trend component. The trend is estimated using wavelets and represents the fluctuation of the data on large time scales. In case the trend adds more variability to the data than the natural variability is likely to generate we consider the trend as significant. Hydrological data possesses short- and long-term memory, due to, e.g., weather regimes or consequences of large scale atmospheric patterns. This is confirmed by our results when analysing about 90 discharge records in the Neckar and Danube River basin: We find a spatially heterogeneous auto-correlation structure incorporating short- and long-term correlations. Thus, the complexity of our novel approach is necessary to adequately represent the data. Our results therefore differ from those of the Mann-Kendall trend test which is a common tool in hydrology. This test too often detects significant trends in case memory is present in the data. Trends and memory are interrelated phenomena. Thus, when applying our trend test we find that the chosen stochastic model influences the uncertainty of the trend estimate as well as the trend test result. We detect increasing and decreasing trends in both basins considered and the resulting spatial pattern is not ad hoc interpretable.

In the second part of this thesis, we use a point process to model trends in the extremes of river discharge data. Thereby we use exceedances over a threshold as extremes which we assume to be distributed according to a generalized Pareto distribution. In order to eliminate auto-correlation, the data are thinned out. Contrary to ordinary extreme value statistics, potential non-stationarity is included by allowing the model parameters to vary with time. By this, changes in frequency and magnitude of the extremes can be tracked. The model which best suits the data is selected out of a set of models which comprises the stationary model and models with a variety of polynomial and exponential trend assumptions. Common assessment measures, such as 100-year return levels,

can be calculated from this model. Analysing winter discharge data of about 50 gauges within the Danube River basin, we find trends in the extremes in about one third of the gauges examined. The spatial pattern of the trends in extremes is not immediately interpretable. Importantly, assuming stationary models for non-stationary extremes results in biased assessment measures. This underlines the necessity of the proposed methodology. The magnitude of the bias depends on the trend strength. Furthermore, the proposed approach allows to quantify the uncertainty of assessment measures such as return levels.

Summing up, this thesis provides enhanced trend and extreme value assessment frameworks, which account for the idea of memory in the data and environmental change. Trends in the mean values and extremes are found to be common in discharge when analysing river basins in Southern Germany. However, as yet no ubiquitous signal of increasing or decreasing discharge trends primarily related to climate change can be detected in these river basins. We observe neighbouring gauges often to display distinct behaviour, possibly due to non-climatic factors such as changes in land use or soil conditions. In any case, a sound trend assessment is the necessary basis for any process oriented, physical interpretation. Moreover, common practice of water management authorities can be improved by applying the proposed methods, and costs for flood protection buildings can be calculated with higher accuracy. That is, construction costs are reduced in case return level estimates are evidenced to be too high and possible future damage costs are dampened in case it is shown that those estimates are too low.

Keywords: trend assessment, auto-correlation, extreme value analysis, threshold excesses, declustering, wavelets, stochastic modelling, discharge

ZUSAMMENFASSUNG

In dieser Dissertation wird ein profunder Datenanalyse-Methodenkanon entwickelt und angewendet, um Trends im Mittelwert und den Extremen von Abflusszeitreihen zu untersuchen. Insbesondere findet dabei die Autokorrelationsstruktur der Daten (das Gedächtnis) Berücksichtigung. Ein solcher Ansatz ist zum Beispiel sehr nützlich, um eine mögliche Intensivierung des globalen Wasserkreislaufes aufgrund des anthropogenen Klimawandels zu untersuchen. Die Kosten häufigerer oder schwererer Hochwasser und Dürren sind sehr hoch, deshalb ist eine akkurate Abschätzung dieser Gefahren samt ihrer Unsicherheiten von großer Bedeutung. Wir analysieren Abflüsse von Einzugsgebieten der Donau und des Neckar in Süddeutschland. Hierbei vergleichen wir unseren Methodenkanon mit in der Hydrologie gängigen Methoden, um Unterschiede und Gemeinsamkeiten bezüglich der Berechnung von Bemessungsgrößen, die von Wasserwirtschaftsämtern genutzt werden, feststellen zu können. Außerdem suchen wir nach räumlichen Mustern von Trends.

Im ersten Teil dieser Arbeit untersuchen wir Trends im Mittelwert von Abflussdaten. Beim Trendtest wird hierbei die Autokorrelationsstruktur der Daten mittels stochastischer FARIMA-Prozesse modelliert. Dies ist ein entscheidendes Merkmal des Trendtests, da durch Autokorrelationen unechte Trends vorgetäuscht werden können. Wir nehmen an, dass die Daten additiv zusammengesetzt sind aus natürlicher Variabilität, die durch den stochastischen Prozeß repräsentiert wird, und möglicherweise einer deterministischen Trendkomponente. Dieser Trend wird mit Hilfe von Wavelets bestimmt und repräsentiert die Fluktuationen der Zeitreihe auf großen Skalen. Falls der Trend mehr Variabilität verursacht als die natürliche Variablität zu generieren vermag, betrachten wir den Trend als signifikant. Hydrologische Zeitreihen zeichnen sich durch Kurzzeit- und Langzeit-Gedächtnis aus, das zum Beispiel durch Wetter oder Folgen großskaliger atmosphärischer Muster verursacht werden kann. Dies wird durch die Ergebnisse unserer Analyse von ca. 90 Abflusspegeln der Einzugsgebiete von Neckar und Donau bestätigt: Wir finden räumlich sehr heterogene Autokorrelationsstrukturen der Daten, die sowohl Kurzzeit- als auch Langzeit-Korrelationen beinhalten. Dies zeigt, dass die Komplexität des hier präsentierten Ansatzes zur Bewertung von Trends notwendig ist, um die Zeitreihen adäquat zu repräsentieren. Unsere Ergebnisse unterscheiden sich deshalb auch von denen des in der Hydrologie gebräuchlichen Mann-Kendall-Trendtests, der zu oft signifikante Trends diagnostiziert, falls die analysierten Daten Autokorrelationen besitzen. Trends und Gedächtnis sind zusammenhängende Phänomene. Beim Anwenden unseres Tests stellen wir also fest, dass das gewählte stochastische Modell sowohl die Unsicherheit des Trendschätzers beeinflusst als auch das Ergebnis des Trendtests. Wir finden sowohl steigende als auch fallende Trends in beiden Einzugsgebieten und es ist kein ad hoc interpretierbares räumliches Muster auszumachen.

Im zweiten Teil dieser Arbeit modellieren wir Trends in den Extremwerten von Abflussdaten mittels Punktprozessen. Hierbei nehmen wir an, dass Überschreitungen eines

Schwellenwertes gemäß der verallgemeinerten Pareto-Verteilung verteilt sind. Autokorrelationen in den Extremwerten werden entfernt, indem man die Daten ausdünnt. Im Gegensatz zu konventioneller Extremwertstatistik wird eine mögliche Instationarität über zeitabhängige Modellparameter zugelassen. Auf diese Weise können Änderungen sowohl in der Häufigkeit als auch im Betrag der Extremwerte berücksichtigt werden. Das beste Modell wird dann aus einer Menge gewählt, zu der sowohl das stationäre Modell als auch Modelle mit einer Auswahl an polynomialen und exponentiellen Trendannahmen gehören. Gängige Bemessungsgrößen, wie das Jahrhunderthochwasser, können nun mit Hilfe dieses Modells berechnet werden. Wir analysieren Winter-Abflussdaten von ca. 50 Pegeln innerhalb des Einzugsgebietes der Donau und finden in ungefähr einem Drittel der Daten Trends in den Extremwerten. Das räumliche Muster der Trends in den Extremwerten ist nicht direkt interpretierbar. Wesentlich ist, dass die Anpassung stationärer Modelle an instationäre Extrema zu einer Verzerrung der Bemessungsgrößen führt. Dies unterstreicht die Notwendigkeit der vorgeschlagenen Methodik. Die Größe der Verzerrung hängt von der Stärke des Trends ab. Die hier vorgestellte Methode bietet außerdem die Möglichkeit, die Unsicherheit der ermittelten Bemessungsgrößen, wie zum Beispiel Wiederkehrschwellen, abzuschätzen.

Die vorliegende Dissertation liefert einen Methodenkanon zur verbesserten Trend- und Extremwertanalyse, die die Vorstellung von Gedächtnis in den Daten und einer sich mit der Zeit verändernden Umwelt aufgreift. Bei unseren Untersuchungen von Abflüssen in Flusseinzugsgebieten von Süddeutschland ermitteln wir oft Trends im Mittelwert und den Extremen. Dennoch ist in diesen Flusseinzugsgebieten kein einheitliches, in Beziehung zum Klimawandel stehendes, Signal fallender oder steigender Abfluss-Trends zu erkennen. Oft stellen wir vielmehr fest, dass sich benachbarte Pegel unterschiedlich verhalten. Dies wird möglicherweise durch Faktoren wie Änderung der Landnutzung oder der Bodenbeschaffenheit bestimmt, die primär nichts mit dem Klimawandel zu tun haben. Auf jeden Fall ist eine fundierte Trendbewertung die notwendige Grundlage für jegliche prozeßorientierte, physikalische Interpretation. Weiterhin kann die gängige Praxis von Wasserwirtschaftsämtern verbessert werden indem die vorgestellte Methodik angewandt wird und Kosten für den Hochwasserschutz können genauer kalkuliert werden. So werden zum Beispiel Baukosten reduziert, wenn die Schätzung zu hoher Wiederkehrschwellen aufgezeigt werden kann und Kosten möglicher zukünftiger Schäden werden vermindert, falls die Unterschätzung solcher Wiederkehrschwellen belegt werden kann.

Stichworte: Trendbewertung, Autokorrelation, Extremwertanalyse, Schwellenwert-Überschreitungen, Entclustern, Wavelets, stochastische Modellierung, Abfluss

CHAPTER 1

MOTIVATION

River discharge data may exhibit trends because of a variety of reasons, thereby climate change is an anticipated factor. This thesis analyses non-stationarities of river discharge during the 20th century in Southern Germany and evaluates the significance of such trends for mean values as well as for extremes, i.e. floods. Regional patterns are looked for to come to a better understanding of the potential for change in river discharge data.

1.1 CLIMATE CHANGE

Since several decades mankind has become aware of its ability to change its environmental conditions, even on a global scale. The emergence and the containment of the ozone hole may serve as an impressive example for the interaction of ecosphere and anthroposphere on a global scale. Several approaches have been made to formalise this interaction at least in a qualitative way (cf. Schellnhuber and Kropp 1998 and Schellnhuber 1998).

Already in the 1860's scientists like Tyndall (1861) recognised the earth's natural greenhouse effect. During the 1980s and 1990s then a scientific awareness began to emerge that the emission of "greenhouse gases" (basically carbon dioxide and methane) into the atmosphere was leading to an increase in global temperatures. This effect nowadays is labelled as *global warming* and has been demonstrated in many ways by science (see, e.g., IPCC 2001, Rahmstorf and Schellnhuber 2006 and IPCC 2007). In this context, the field of time series analysis, in which also this thesis is located, contributed notably. Reconstructions of Northern Hemisphere temperatures for the last 1 000 years have been studied, for example, to assess the question of global warming (cf. Mann et al. 1998 and Mann et al. 2004). Luterbacher et al. (2004), for example, affirm an increase in temperature in the late 20th and early 21th century as highest in the record regarding European climate. The detection of potential anthropogenic impacts is as well addressed using data analysis methods (see Bloomfield 1992, Smith 1993 and Schär et al. 2004).

To examine the evidence for anthropogenic effects on climate change the Intergovernmental Panel on Climate Change (IPCC) was established by the United Nations and the World Meteorological Organisation. This institution also worked on the formulation of plausible greenhouse gas emission scenarios in order to predict future climatic changes. The scientific assessment reports of the IPCC (IPCC 2007) state that most of the observed increase in globally averaged temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations. The associated impact assessments (cf. Stern 2006) suggest that the social, economic and environmental consequences of climate change are likely to be severe.

Regarding the identification of changes in extreme events, a widespread assumption is that climate change leads to an increase in the frequency and magnitude of severe weather events. This is confirmed by a large number of meteorological records (such as highest temperature events, most intense precipitation or longest dry spell), which have been surpassed within the past two decades (FOE 1997). The IPCC states (IPCC 2001) that "small changes in the mean climate or climate variability can produce relatively large changes in the frequency of extreme events", acknowledging that substantial changes in the extreme behaviour are not implausible from a scientific perspective.

Thereby the concern is not so much the mean temperature increase of about 0.74°C that has occurred since 1906, but the best estimates for projections of globally average surface air warming, which range from 1.8°C for the low emission scenario (B1) up to 4.0°C for the high emission scenario (A1FI) (IPCC 2007). Climate researchers now have to interact with economists, sociologists and political scientists in order to assess the full implications of climate change for sustainable development. Nowadays the existence of a global warming phenomenon is widely accepted by politicians, which is documented by various self-commitments of a lot of states to take the necessary steps to protect the climate.

1.2 INFLUENCE OF CLIMATE CHANGE ON THE HYDROLOGICAL CYCLE

Changes in the atmospheric composition may add to the risk of floods (Frei et al. 2000 and IPCC 2001). This is because the capacity of the atmosphere to hold water grows with increasing temperature. In this way the potential for intense precipitation also increases. Higher and more intense rainfall has already been observed (cf. Osborn et al. 2000 or Kundzewicz and Schellnhuber 2004). This trend is expected to continue, which gives reason to anticipate a rise of flood hazards because of global climate change as well.

In Europe floods have been a known natural hazard for centuries (cf. Pfister et al. 1999, Brázdil et al. 1999, Brázdil et al. 2005 and DFO 2004). Nevertheless, the occurrence of extreme flood events seems to have grown considerably over recent decades. IPCC (2001) and Kundzewicz and Simonovic (2002), for example, find a higher occurrence of flood events worldwide in the time span of 1990-1998 than in the nearly four times longer period of 1950-1985. Also recently occurred extreme river floods have had severe effects in central Europe. The Elbe flood in August 2002, for example, caused 36 deaths and over 15 billion USD damages and the Oder flood in July 1997 caused 114 deaths and around 5 billion USD damages (Mudelsee et al. 2004). This might support the assumption that the anticipated influence of global warming on the hydrological cycle also has an effect on hydrological extremes. Therefore the question of detection of changes in various flood and drought-related indices attracts more and more interest. This thesis also aims at investigating trends in extreme values of hydrological data.

Furthermore, average river run-off is an important hydrological statistics for assessing river catchments. Therefore systemic change in average river run-off is as well investigated in this dissertation. The importance of the topic is shown by various studies (cf. Burlando and Rosso 2002 and KLIWA 2003).

Yet, there is no common trend tendency detectable in drainage basins of rivers (catchments) or even larger regions. Therefore the picture composed by various studies, which assess trend tendencies in hydrological data, is very diverse. Black (1995) analyses annual mean run-off time series in Scotland and partly detects an increasing trend, predominantly after 1989. Caspary and Bárdossy (1995) find an increasing trend at some stations in Germany for annual maxima data. Robson et al. (1998) and Robson (2002) analyse run-off extremes of UK rivers and cannot find evidence of significant trends. For U.S. rivers Douglas et al. (2000) do not detect trends in flood flows, but they find indication for an upward trend in low flows. Analysing stream flow data in Switzerland, the results of Bîrsan et al. (2003) indicate an increasing run-off in winter and a decrease in summer. Zhang et al. (2001) detect no increasing trend for Canadian rivers, but in some cases decreasing changes. Kundzewicz et al. (2004), who analyse maximum annual flow all over the world, find significant increasing and decreasing trends – but no regional patterns.

Systemic change is difficult to detect in river run-off data in any case. The observed system is complex and anyway possesses a large variability, which in addition is delayed by large storage capacities such as groundwater reservoirs. Furthermore, regarding the detection of climate change, the interactions between precipitation, temperature and river run-off are not universal or linear. Last but not least, the separation of systemic change and natural variability, that is the detection of trends, does not yet solve the attribution problem. The systemic change may be caused by a lot of different phenomena. Socio-economic factors such as population increase (and therefore different settlement patterns, cf. Haase and Nuissl 2007) or accumulation of wealth in vulnerable areas may be a reason for an increase in impacts. Changes of local or site-related factors such as soil type, land use, urbanisation patterns or river engineering (e.g. drainage or flood alleviation schemes) are important sources for trends in the data and related effects may overlay a possible trend due to global warming.

Hence up to now, there is no general and conclusive proof of anthropogenically caused climate change having an impact on floods (cf. Kundzewicz and Schellnhuber 2004 and Katz et al. 2002), but also no disproof. This task remains an important field of scientific investigation. For practical water management reasons, a reliable estimation and prediction of peak flows under changing climatic conditions, preferably with an assessment of the uncertainty of these estimates, is desirable. These problem fields and requirements motivated our thesis, which presents a framework for assessing non-stationary peak flow data and incorporating confidence intervals.

1.3 APPROACHES TO ASSESS TRENDS

One way to distinguish scientific techniques to assess changes may be the division into a process oriented, explanatory modelling approach on one side and a descriptive data analysis approach on the other.

By functional, process oriented modelling, parts of a real system (i.e. a range of natural processes) are mapped to a mathematical model. Such models can be used for simulation studies. The output then is evaluated and the conclusions drawn are used to explain and interpret possible reactions of the real system. This procedure is especially useful for studying behaviour of the system for yet not observed circumstances, e.g. scenarios or future settings. In the field of hydro-meteorology there exists a wide range of process

oriented models for a variety of purposes. Regional climate models (cf. Thodsen 2007) provide rainfall and temperature scenarios as input for, e.g., hydrological models, which generate the run-off. They include processes of interception, evapotranspiration and water storage in soils and aquifers (see, e.g., Bremicker 2000 and Jones et al. 2006). Models on a global scale, such as general circulation models (GCM), contain process oriented parts. They are designed to conduct global climate studies and they are widely used for modelling the development of the global climate under various emission scenarios (see, for example, Claussen 1996, Stott et al. 2004, Schneider von Deimling et al. 2006 and Baettig et al. 2007). These approaches are established by now, their results have advanced scientific research and form a basis for political decision making. However, some drawbacks also appeared. Kirchner (2006), for example, questions the upscaling premise used for a lot of physically based models of hydrologic systems. Besides, the GCM models are tuned to reproduce the 30 year *average* behaviour of the system adequately. Therefore the reproduction of extreme events is difficult. Thus the anticipation of extreme scenarios by the help of GCM model runs, for example, does not seem to be satisfactory (Lucio 2004). Furthermore, no probabilistic forecasts with numerical models exist, which systematically investigate the range of possible responses of the hydrological cycle to anthropogenic climate change (Allen and Ingram 2002). This is due to the complexity and the computational expenses of the issue.

Data analysis possesses a long tradition in the field of hydro-meteorology as well. Among others Bloomfield (1992) and Smith (1993) have assessed trends in global temperature by analysing observed data. Hurst (1951) investigated the Nile River and formulated a universal law concerning water storage. Extreme value analysis theory, which has especially been developed to deal with extreme events (Embrechts et al. 1997), is used by water management authorities and engineers to estimate return levels of floods and to set up protecting constructions. A variety of statistical models and analysis techniques are utilised in time series analysis and modelling (cf. Lange 2006). The variability in a system is partly represented using probability distributions and random variables to model variations in empirical data rather than using non-linear deterministic functions. Time series analysis also has to deal with several drawbacks: Analysing data, one has to cope with the relative shortness of the empirical record in comparison to the time span of processes which might influence the variability of the system. Predictions are made by extrapolating statistical models into the future, but usually the conditions under which the model has been constructed are assumed to stay constant, which may be misleading. Furthermore, all time series methods require a certain length of the data to obtain credible results. Measurement errors (which may be as high as 10% of the total variability of the data in the case of river run-off) might also affect the findings (this also holds for process oriented models, which naturally rely on empirical data).

Nevertheless, the modelling and the data analysis approach both contribute to the assessment of the current and important issue of the detection and attribution of climate change and resulting changes in the global hydrological cycle. Furthermore, they intertwine fruitfully for a variety of purposes. Time series analysis and statistical methods are used for model validation by comparing model output and empirical data according to some statistical criterion (see, e.g., Fuentes et al. 2003 and Vyushin et al. 2004). Stochastic models may be used to build simple and fast emulators of complex process oriented models. In this way Monte Carlo studies of these models can be mimicked, which helps to assess the uncertainty of deterministic flood risk models (Apel et al. 2004) or GCM output (cf. Oakley and O'Hagan 2002, Lefebvre 2002, and Goldstein and Rougier 2006).

Time series regression analysis is capable to use model output as deterministic predictor variable, which e.g. is utilised by Smith et al. (2003) to detect trends in hemispheric mean temperatures. These few examples of the combination of modelling and data analysis methods underline the necessity of both approaches.

1.4 AIMS AND OUTLINE OF THE THESIS

The appearance of floods is affected by short-term phenomena in the first place. Heavy precipitation and other weather conditions have a large impact on the evolution of river discharge. However, the examination of processes, which act on large time scales, also have to be taken into account to adequately assess the evolution of river discharge and to make predictions for the nearer future. In this context, the separation of systemic changes and natural variability is a key question, especially in the context of the anticipated impact of global warming on the hydrological cycle. This work aims at contributing to these topics by means of time series analysis, that is by examining observed records. Thereby a special focus is set on the modelling of non-stationary data, i.e. of observations where a trend in the mean or other moments of the distribution is present. It is assumed that this trend is traceable as a slow change, which takes effect on large time scales. Thereby the trend is explicitly modelled. By this means, all empirical information available can be used at once. This is contrary to, e.g., comparing statistical characteristics of two time windows. The non-stationarity may be caused by a variety of factors, such as climate change due to anthropogenic influence, local river regulation schemes or by natural factors, for example, large scale atmospheric patterns influenced by El Niño and general weather circulation patterns. Therefore we do not presume a certain trend shape, e.g. "linearly increasing", but we test for a deterministic signal, which may also exhibit periodic oscillations.

We expect this signal to be hidden by the "natural variability" of the system, which we model via the auto-correlation structure of a time series. A lot of standard methods assume independent, e.g. non-correlated data. Concerning river discharge data this assumption is questionable (cf. Montanari et al. 2000 and Mudelsee 2007) and has to be verified before applying such methods. Auto-correlations describe the correlation between different points in time of a process. Therefore they are a kind of time dependence, as well as trends. River discharge may exhibit natural variability because of a lot of processes, such as evaporation, groundwater storage or the composition of the soil next to the river. Apparently these factors operate on small and large time scales. However, auto-correlations are a stochastic phenomenon, whereas trends are regarded as systemic or deterministic change. Thus, both phenomena have to be treated differently. Furthermore, the auto-correlation structure of river discharge typically is stationary, i.e. although the auto-correlations may cause the data to fluctuate away from its mean for a long time period, the data is always expected to return to its mean value.

It is our aim to separate the deterministic trend component from the stochastic variability. However, the mutual confusion of trend and auto-correlations is a known fact (see Beran 1994). Trend estimation methods may detect a "trend", which actually is produced by a stationary auto-correlated series and vice versa. To cope with this problem, in this work a method of trend analysis which takes care of auto-correlations is applied. We verify the methodology, compare it with established methods and apply it to empirical data in Southern Germany.

Furthermore, we assume that non-stationarities in the time series influence the evolution of extreme events (i.e. floods or droughts) in one or the other way. Therefore in the second part of this thesis a framework to adequately model non-stationary extreme values is presented. Thereby the auto-correlation structure of the extremes is not modelled explicitly, but it is removed. This is reliable, because auto-correlations play a much less important role when looking at extreme events. Trends in extremes are of special interest because of their heavy impacts on society. We incorporate them by allowing the parameters of a point process, which is fitted to the data, to vary with time. In this work we verify the extreme value assessment methodology for non-stationary extremes and apply it to empirical data in Southern Germany.

Hence the key questions of this thesis are:

- Are there trends in the mean of hydro-meteorological data?
- Are there trends in the extremes of hydro-meteorological data?
- Do these non-stationarities affect assessment measures commonly used by water management authorities?
- Do regional trend patterns emerge?

By developing a sound framework to address these questions, an essential requirement for the attribution of the detected trends to potential causes is achieved. Together with data and expert knowledge out of a variety of application fields, such as hydrology, climatology, history or soil sciences, the results of this thesis may serve to identify potential causes for trends, which is beyond the scope of this work.

The research results of this thesis are applicable in many ways. On the one hand, the findings of the assessment of trends and the auto-correlation structure in hydrological data are important for practitioners. The examination of these statistical characteristics provide indicators for water management authorities. They may use them to adjust measures for risk assessment, to tune flood protection systems, and to build up adequate protection constructions. The Bavarian Environmental Protection Agency for instance is engaged in several scientific projects to this purpose. This thesis originates from one of those, namely the German Ministry of Education and Research (BMBF) project for *scaling analysis of hydro-meteorological data* (Kropp and Schellnhuber 2006). On the other hand, the results of this dissertation are interesting in a scientific context. A better description of hydro-meteorological data is a valuable basis to interpret natural processes. Furthermore, an improved insight in the impacts of natural disasters (such as floods) on human society may be obtained by combining studies conducted in this thesis with an assessment of the development of society, which addresses the sensitivity and adaptive capacity of the exposed societal entities. Thus, the quantification of the vulnerability of a society, even in monetary terms, may be possible, at least from a "site-based", i.e. local, point of view (Turner et al. 2003 and Glaser and Stangl 2003).

This thesis is organised in four main parts. In Chap. 2 the methodological framework to assess trends of average and extreme data by means of time series analysis is sketched. Here the utilised data and its preparation are presented as well. Details of the assessment of trends in mean values and results of a study of basins in Southern Germany are given

in Chap. 3. We combine a trend estimate obtained by wavelet filtering and the class of stochastic fractional auto-correlated moving average (FARIMA) models to assess trends in a semi-parametric approach. A trend test framework is provided, which allows for data adaptive and flexible trend estimation and the consideration of short-term as well as long-term auto-correlation of the data. By doing so, we extend the trend test proposed by Craigmile et al. (2004) and therefore refer to it as Craigmile-Guttorp-Percival (CGP) trend test. In Sec. 3.5.2 we compare the CGP trend test to the Mann-Kendall trend test (cf. KLIWA 2003), which is common in hydrology but may be biased in the presence of auto-correlations (cf. Hamed and Rao 1998). In Chap. 4 a framework for examining non-stationary extreme values is formulated and findings of a study of the Danube River basin are presented and discussed. Common extreme value statistics assumes independent and identically distributed data (Embrechts et al. 1997). In the context of climate change and the availability of longer measurement records these assumptions become questionable. We examine possibly non-stationary extreme values with a point process approach. By doing so, a generalized Pareto distribution is fitted to excesses over a threshold, which allows for the elimination of auto-correlations in the extremes. The frequency of occurrence is modelled by a Poisson process. Trends in magnitude as well as in frequency of occurrence can be assessed and uncertainty measures are provided along with the results. Finally a summary is given and conclusions are drawn in Chap. 5.

CHAPTER 2

BASIC METHODS AND DATA

2.1 STOCHASTIC PROCESSES

2.1.1 Basic definitions

As outlined in Chap. 1, we use data analysis methods to retrieve information about hydrological processes. This methodology is not restricted to assess observed data, but is also capable to reproduce empirical data, which is useful either for prediction purposes or for simulation studies to assess uncertainty issues. In time series analysis *stochastic processes* are used to model the data. These processes are rather driven by randomness to model fluctuations of empirical data than by non-linearity (which mostly is utilised by dynamical models built up by e.g. differential equations). In the following we refer to standard concepts and definitions of stochastic processes as given, for example, in Priestley (1981). In Chaps. 3 and 4 these concepts are used to build up our trend assessment frameworks.

A *random variable* X is a function from a sample space S into the real numbers. With every random variable, we associate a function called the cumulative distribution function (cdf) of X, which we denote $F_X(x)$. It is defined by

$$F_X(x) = P_X(X \leq x) \qquad \text{for all } x \quad . \tag{2.1}$$

The *expected value* or *mean* of a random variable X, denoted by $E(X)$, is

$$E(X) = \mu = \int_{-\infty}^{\infty} x f_X(x) dx \quad , \tag{2.2}$$

where $f_X(x)$ is the density (pdf) of X. It is the first *moment* of X and the n^{th} centralised moment is defined by $\mu_n = E(X - \mu)^n$. The second moment is the *variance* of a random variable. The moments of a distribution are important characteristics of a distribution, but they do not need to exist.

A sequence of random variables in time $\{X_t\} \equiv X_1, X_2, \ldots$ is called a *stochastic process*, or random process. A sample of observations $x_1, x_2, \ldots, x_n$ can be interpreted as a realisation of a stochastic process. In this case, the values of these observations cannot be predicted precisely beforehand but probabilities can be specified for each of the different possible values at any particular time. These probabilities are determined by the marginal cdf of each random variable X_t.

Let $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ denote the joint distribution of $X_1, X_2, \ldots, X_n$. A stochastic process $\{X_t\}$ is *completely stationary* in case

$$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_1, x_2, \ldots, x_n) \equiv F_{X_{(t_1+k)},X_{(t_2+k)},\ldots,X_{(t_n+k)}}(x_1, x_2, \ldots, x_n) \tag{2.3}$$

holds for any admissible $t_1, t_2, \ldots, t_n \in \mathbb{R}$ and any k.

This basically means that for any set of time points $t_1, t_2, \ldots, t_n$ the joint probability distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ must remain unaltered if each time point is shifted by the same amount k. Complete stationarity implies independence from time for all moments (in case they exist). This requirement can be relaxed to *stationarity up to an order m*. Here it is only demanded that the main features of the distributions of X_{t_1} and $X_{(t_1+k)}$ are similar, i.e. that their moments up to order m have to be the same. Definitions of *weak stationarity* only assume time independence of moments up to order 2 (for further details see Priestley 1981). The distribution of a normal distribution is determined by its first two moments. Thus, for a Gaussian time series, weak stationary is equivalent to strict stationary.

The *auto-correlation function* $\rho(\tau)$ of weak stationary processes depends only on the differences $\tau = t_i - t_j$, which results in

$$\rho(\tau) = \frac{E([X_t - E(X_t)][X_{t+\tau} - E(X_t)])}{E(X_t^2) - [E(X_t)]^2} = \frac{E(X_t X_{t+\tau}) - [E(X_t)]^2}{\mathrm{VAR}(X_t)} \quad . \tag{2.4}$$

2.1.2 *Autoregressive moving average processes*

Autoregressive moving average (ARMA) processes are linear stochastic processes which model the auto-correlation structure of a data series. In our setting, the domain over which these random functions are defined, is a time interval. The auto-correlation structure can be seen as the "memory" of a time series, e.g. due to auto-correlation events in the past influence events in the present or future despite of the concept of randomness, which drives stochastic processes. The intensity of this memory depends on the influence of factors such as weather or soil conditions. Precipitation, for example, often does not have a memory at all and therefore is said to be *purely stochastic*. Hipel and McLeod (1994) present an overview for the physical justification of using ARMA models to represent river flow data. Autoregressive moving average (ARMA) processes now are well-known discrete parameter models incorporating auto-correlations of the random variable itself (AR) and a noise part (MA). Assuming $\mu = E(X_t) = 0$ an ARMA(p, q) process is defined by

$$\phi(B)X_t = \psi(B)\epsilon_t \tag{2.5}$$

with B denoting the back-shift operator $\mathrm{BX_t} = \mathrm{X_{t-1}}$ and ϵ_t being independent and identically distributed (iid) normal random variables with zero expectation and variance σ_ϵ^2 and

$$\begin{aligned} \phi(z) &= 1 - \sum_{i=1}^{p} \phi_i z^i \quad , \\ \psi(z) &= 1 + \sum_{j=1}^{q} \psi_j z^j \end{aligned} \tag{2.6}$$

are the autoregressive (AR) and moving average (MA) polynomials of order p and q, respectively.

An ARMA(p,q) process is stationary in case all solutions of $\phi(z) = 0$ lie outside the unit circle and $\phi(z)$ and $\psi(z)$ do not have common roots[1]. It is *causal* if all solutions of $\psi(z) = 0$ lie outside the unit circle.

For ARMA processes, the asymptotic decay of the correlations is exponential in the sense that there exists an upper bound for the auto-correlation function $\rho(k)$, i.e.,

$$|\rho(k)| \leq ba^k \quad , \tag{2.7}$$

where $0 < b < \infty$, $0 < a < 1$. $|a| < 1$ holds, therefore we have $\sum_{k=0}^{\infty} \rho(k)$ = constant $< \infty$. Stochastic processes with this kind of correlation structure are called *short-range correlated* processes. For further details see, e.g., Box and Jenkins (1976).

2.1.3 Fractional ARIMA processes

An autoregressive integrated moving average (ARIMA) process (Box and Jenkins 1976) is obtained by integrating an ARMA process. Thus, if Eq. (2.5) holds for the δth difference $(1-B)^{\delta} X_t$, then X_t is called an ARIMA(p,δ,q) process for δ being integer values.

Fractional ARIMA (FARIMA) processes (Granger and Joyeux 1980) are an extention of these well studied processes incorporating *long-range dependence* or *long memory* or *long-term correlation*. Long-range dependence is qualitatively different from the short-ranged AR or MA dependence. Auto-correlations of long-range dependent data decay slower than exponential and it is not possible to determine a specific time lag so that we find correlations of larger lags getting negligible. To be more specific, a process has long-range dependence or long-range correlation if the auto-correlation function $\rho(k)$ decays algebraically in the limit of large time lags k:

$$\lim_{k \to \infty} \frac{\rho(k)}{ck^{-\beta}} = 1 \quad , \tag{2.8}$$

with $\beta \in (0,1)$ and $c > 0$ being a finite constant. This implies that, contrary to the short-range dependent case, $\sum_{k=-\infty}^{\infty} \rho(k) = \infty$ holds, that is the correlations are not summable.

Assuming $\mu = E(X_t) = 0$, a FARIMA(p,δ,q) process is defined by

$$\phi(B)(1-B)^{\delta} X_t = \psi(B)\epsilon_t \quad , \tag{2.9}$$

with B, ϵ_t, $\phi(z)$ and $\psi(z)$ as introduced in Eq.(2.6) and $\delta \in \mathbb{R}$ being the fractional difference or long-memory parameter. A FARIMA(p,δ,q) process is causal and stationary if $\delta < 0.5$ and all solutions of $\phi(z) = 0$ and $\psi(z) = 0$ are outside the unit circle. A variety of non-stationary long-memory processes have stationary backwards differences, i.e. they can be made stationary by differencing. Thus, to obtain a stationary process out of a non-stationary one, where the non-stationarity is caused by the long-memory parameter, X_t must be differenced δ_i times, thereby δ_i is the integer part of δ. Also a random walk process can be made stationary by differencing (Percival and Walden 2000). A FARIMA

[1] Roots are the solutions of a polynomial set equal to zero.

process exhibits long memory for $0 < \delta < 0.5$. FARIMA(p, δ, q) models with $\delta < 0$ are called intermediate memory or "overdifferenced". In practice, this case is rarely encountered (Beran 1994).

In case $\delta \in \mathbb{Z}_+$, that is the ARIMA case, $(1-B)^\delta$ can be calculated by

$$(1-B)^\delta = \sum_{k=0}^{\delta} \binom{\delta}{k} (-1)^k B^k \quad . \tag{2.10}$$

For any real number $\delta \in \mathbb{R}$, i.e. the FARIMA case, the difference operator can be expanded to an infinite power series

$$(1-B)^\delta == \sum_{k=0}^{\infty} \Big(\frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)} \Big) (-1)^k B^k \quad . \tag{2.11}$$

Here, $\Gamma(x)$ denotes the gamma function. This formula can be reduced to Eq. (2.10) for the ARIMA case (for negative integers the gamma function has poles, that is the binomial coefficient is zero if $k > \delta$ and δ is an integer). For more details see Beran (1994), Ooms and Doornik (1999), and Sibbertsen (1999). Further details concerning the autocovariance function $\rho(k)$ of a FARIMA process or its spectrum (the Fourier transform of $\rho(k)$) are given in appendix A-1.

Three of the most simple models out of the FARIMA(p, δ, q) class are:

1. FARIMA(1,0,0) or AR(1)
$$(1 - \phi_1 \mathrm{B}) X_t = \epsilon_t \quad , \tag{2.12}$$
2. FARIMA(0,δ,0) or FD(δ)
$$(1 - \mathrm{B})^\delta X_t = \epsilon_t \quad , \tag{2.13}$$
3. FARIMA(1,δ,0)
$$(1 - \phi_1 \mathrm{B})(1 - \mathrm{B})^\delta X_t = \epsilon_t \quad . \tag{2.14}$$

The first model consists of only a short-range correlated component with parameter ϕ_1, the second model of a long-range correlated component only with long memory parameter δ. The third model combines the previous two in the sense that for $\delta = 0$ or $\phi_1 = 0$ it recovers the first or the second model, respectively.

The assessment of long-range dependence first became famous with the work of the hydrologist Hurst (1951), who was interested in modelling the storage capacity of reservoirs of the River Nile. By studying the flow of the Nile, he formulated a power law. In this formula the famous *Hurst coefficient* H is used. H is related to δ by $H = \delta + 0.5$ for stationary processes (for non-stationary long memory processes H is not defined).

In Fig. 2.1 as an example a realisation of a FD(δ) process with parameter $\delta = 0.4$ and an AR(1) process with parameter $\phi_1 = 0.8$ and their auto-correlation functions are depicted. Both processes have mean 0 and standard deviation 1. The AR(1) process exhibits an exponential decay and its auto-correlation function is summable (depicted in green in the right figure). On the other hand, $\rho(k)$ of the FD(δ) process decays much slower, algebraically.

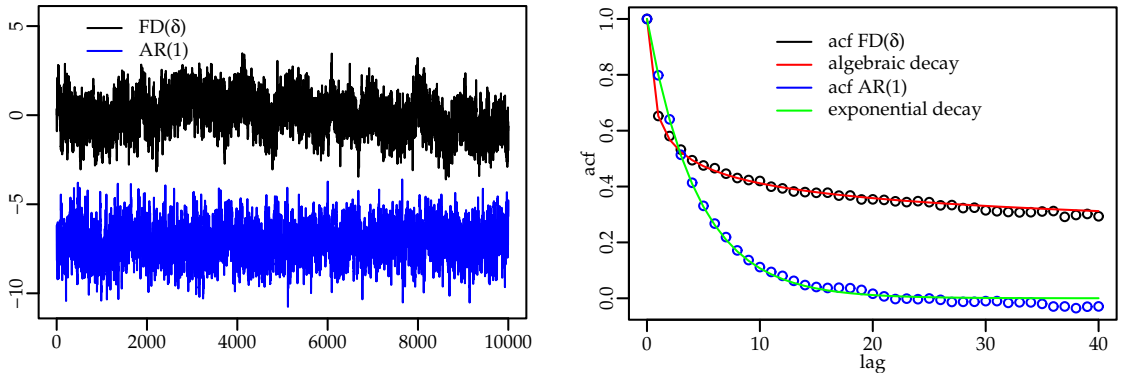


Figure 2.1: Example of a FD(δ) and AR(1) series and auto-correlation function. Left: Realisation of a FD(δ) process with $\delta = 0.4$ (black) and of an AR(1) process with $\phi_1 = 0.8$ (blue). The standard deviation of both processes is 1 and the mean zero, though the AR(1) process is shifted to become visible. **Right**: Auto-correlation function of the FD(δ) process (black) and the AR(1) process (blue). The auto-correlation function of the AR(1) process decays exponentially, i.e. $\rho(k) \leq b\phi_1^{|k|}$ (which can be rewritten as $\exp\{\ln(b) + |k|ln(\phi_1)\}$). This function is depicted in green. The auto-correlation of the FD(δ) process decays much slower, i.e. algebraically, according to $c|k|^{2\delta-1}$, which is shown in red. Hereby $b > 0$ and $c > 0$ are constants.

2.2 WAVELET ANALYSIS

Within the last decade wavelet methodology has had a significant impact in areas as image processing, statistics and differential equations. Wavelets in the sense used nowadays have been introduced by Goupillaud et al. (1984) as an alternative to local Fourier analysis. Wavelets are a special class of functions or sequences. The shape of these functions resembles a small wave, or *wavelet* in French-English notation. We use wavelet analysis to filter a time series, that is to split it in small-scale and large-scale parts.

A time series can be fully reconstructed with a linear combination of *wavelet filters*, which constitute the *wavelet transform*. This is analogous to a reconstruction by sinusoids in Fourier analysis. The members of the transform are generated from a *mother wavelet*, which is shifted in time and dilated. Each wavelet is non-zero only within a finite interval of time, which allows for the assessment of the variability of a time series not only local in scale (e.g. frequency) but also local in time. The mother wavelet is dilated when processing larger scales. Thus, the problem of under- or over-localisation, which occurs in case the windowed Fourier transformation is used, can be minimised. For more details see, e.g., Kaiser (1994), Vetterli and Kovacevic (1995), and Blatter (1998). The choice of the mother wavelet depends on the data to be analysed. Possible choices of wavelet and method lead to an orthogonal or non-orthogonal, a discrete or continuous or a real or a complex transform (or basis in linear algebra). An overview about the noteworthy aspects is given by Torrence and Compo (1998).

2.2.1 *Discrete wavelet transform*

The *discrete wavelet transform* (DWT) of $\{X_t\}$ is an orthonormal transform like the Fourier transform. The algorithm to compute the wavelet coefficients operates on dyadic time

series, that is the length of the time series must be a power of two. The transform is as well only computable for scales which are powers of two. This reduces redundancy (which occurs because of time and scale locality) and computing efforts. On the other hand, the examined scales cannot be chosen arbitrarily any more.

The DWT is given by a set of wavelet filters $\{h_{j,l}\}$, retrieved out of a mother wavelet. Here j denotes the levels $j = 1, ..., J$, which are associated with the scales $\tau_j \equiv 2^{j-1}$. l runs over time ($-\infty < l < \infty$), but the wavelet is transformed to a filter with the length of the time series n, i.e. it is periodised (for further details see Percival and Walden 2000). L_j denotes the *width* of the wavelet at level j. This refers to the fact that the wavelet filter is defined over an infinite time domain, but is only non-zero within the finite interval L_j. Thereby the width of the mother wavelet at scale 1 is $L \equiv L_1$. All wavelet filters on one particular scale do have the same length, i.e. all other wavelet filters on scale 1 are just shifted versions of the mother wavelet. The length of the wavelet filters of the higher levels $j > 1$ is derived from L by dilation: $L_j = (2^j - 1)(L - 1) + 1$.

Each wavelet filter has an associated scaling filter $\{g_{j,l}\}$, which is completely defined by the wavelet filter. This has technical reasons, which are immanent to the DWT: Wavelet filters here are used to split up the time series into its fractions at certain scales and a "rest". To define this rest, we need the scaling filter. On the first level, the scaling filter, for example, is constructed out of $h_{1,l}$ as follows: $g_{1,l} \equiv (-1)^{l+1} h_{1,L-1-l}$ (on the next levels, more complicated relationships exist). Let $n = 2^K$ be the length of the time series. Then the highest level J up to which we are able to compute the DWT is given by $1 \leq J \leq (K + 1)$.

Formally, a wavelet filter is a sequence that sums up to zero and is orthogonal to itself shifted by an even number of points. It also has unit *energy*, i.e. its squared Euclidean norm equals 1. In that sense the energy can be seen as an indicator of the variability of a stochastic process. For all j and all nonzero integers i we get:

$$\text{i)} \quad \sum_{l=-\infty}^{\infty} h_{j,l} = 0 \quad \text{and} \quad \text{ii)} \quad \sum_{l=-\infty}^{\infty} h_{j,l} h_{j,l+2i} = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.15}$$

where $h_{j,l} \equiv 0$ for $l < 0$ and $l \geq L_j$, i.e. the wavelet filters have a finite width L_j. Eq. (2.15 i) implies that any excursions above zero have to be cancelled out by excursions below zero, whereas Eq. (2.15 ii) signifies that generally the wavelet has to undergo some excursions away from zero (for more details see Percival 2002).

Let the time series $\{y_t\}$ be a realisation of a stochastic process with random variables Y_t , $t = 0, ..., n - 1$; n being dyadic[2]. Now let $\mathbf{Y}$ be the vector representation of $\{y_t\}$, that is $\mathbf{Y} = [y_0, \ldots, y_{n-1}]$. Applying, for example, the $LA(\cdot)$ wavelet transform (cf. Sec. 2.2.2) to a time series $\mathbf{Y}$ yields essentially a contrast between y_t and values before and after y_t at a fixed point in time t and a fixed scale, i.e. the length of the interval in which the contrast is measured. The $LA(\cdot)$ scaling filter yields a weighted average of the time series. Its length depends on the chosen scale. As an example in Fig. 2.2 a chirp signal[3] is depicted and the logarithm of its wavelet power spectrum for levels $j = 1, \ldots, 6$ (derived with a $LA(8)$ wavelet). The wavelet power spectrum consists of the squared norm of the wavelet coefficients and the logarithm is taken for better visibility. The intensity of the wavelet power goes to the smaller scales at the end of the time series, where its fluctuations become faster. It is also visible that the wavelet power spectrum smoothes the effects.

[2] That is n must be a power of two.

[3] A chirp signal is a sine wave whose frequency increases at a linear rate with time.

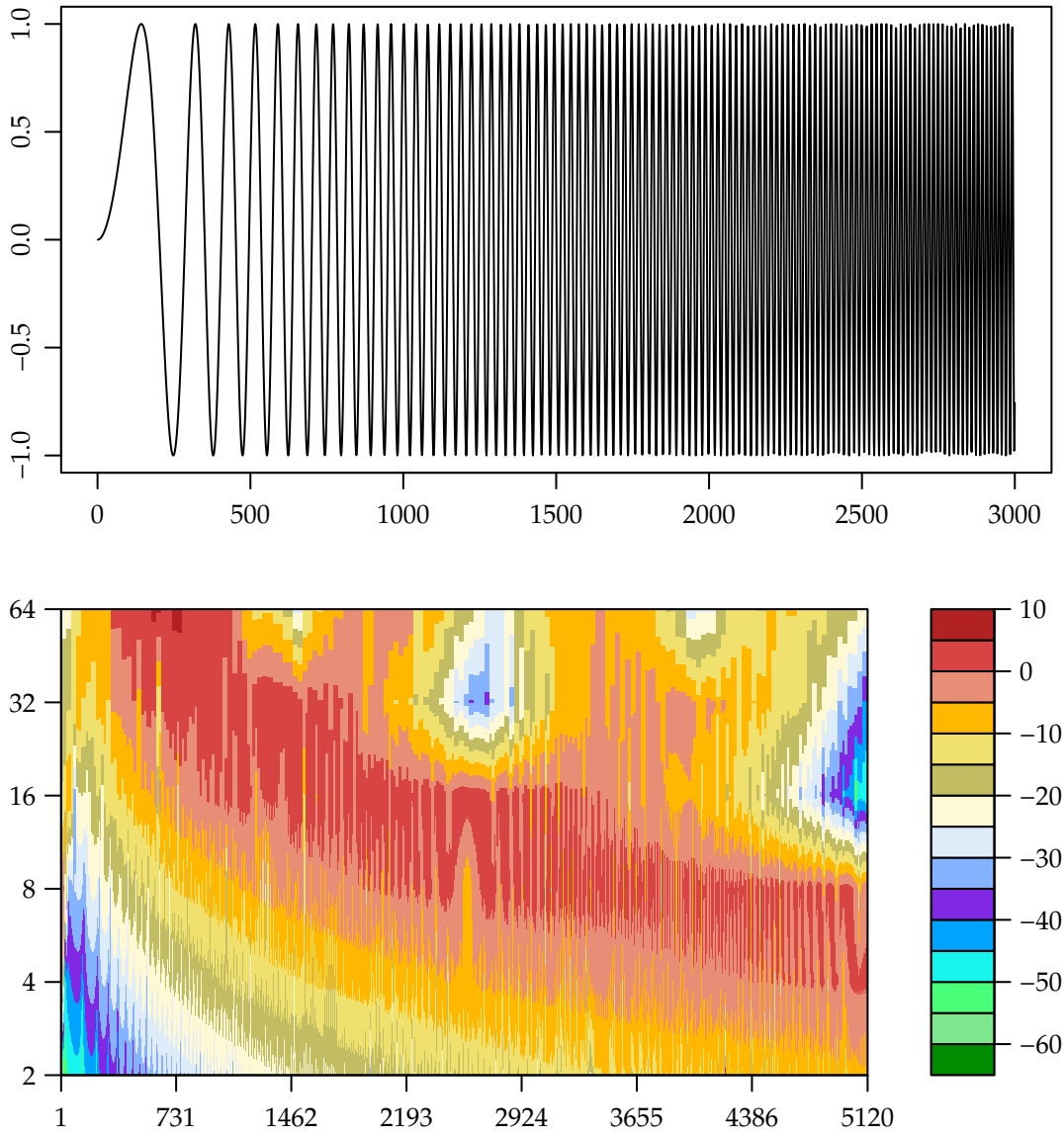


Figure 2.2: Chirp signal and wavelet power spectrum. Top: Chirp signal (only the first half of the time series is depicted for better visibility of its shape). **Bottom**: Logarithm of wavelet power spectrum of chirp signal for levels $j = 1, \ldots, 6$, derived with the $LA(8)$ wavelet.

By convolution of the wavelet filters with $\{y_t\}$, the wavelet coefficients $W_{j,k}$ and scaling coefficients $V_{j,k}$ for level $j = 0, ..., J$ and time $k = 0, .., n_j - 1$ are obtained (with the number of coefficients $n_j = n2^{-j}$). The larger the scales, the wider the applied filters. Thus, using an orthogonal transform implies that for larger scales fewer convolutions are

done, i.e. fewer wavelet coefficients are available:

$$W_{j,k} \equiv \sum_{l=0}^{L_j-1} h_{j,l} Y_{2^j(k+1)-1-l \bmod n} \quad ,$$
$$V_{j,k} \equiv \sum_{l=0}^{L_j-1} g_{j,l} Y_{2^j(k+1)-1-l \bmod n} \quad . \tag{2.16}$$

The calculation of the DWT is efficiently implemented by means of a pyramid algorithm (Mallat 1989). Formally, the DWT of $\{y_t\}$ is defined via matrix multiplication that yields vectors $\boldsymbol{W}_j \equiv [W_{j,0}, W_{j,1}, \ldots, W_{j,n_j-1}]^T$ and $\boldsymbol{V}_j \equiv [V_{j,0}, V_{j,1}, \ldots, V_{j,n_j-1}]^T$[4].

$\mathbf{W}_j$ is proportional to differences (changes) in values adjacent to y_t at scale τ_j, which gives the size of the interval around y_t, in which the changes are measured. $\mathbf{V}_j$ is proportional to the average of the data on scale τ_{j+1}.

Reconstruction

Let $\mathbf{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_J, \boldsymbol{V}_J]^T$. Then the time series, represented by the vector $\mathbf{Y} \equiv [Y_0, Y_1, ..., Y_{n-1}]^T$, is reconstructed as $\boldsymbol{Y} = \boldsymbol{\mathcal{W}}^T\mathbf{W}$, where the DWT matrix $\boldsymbol{\mathcal{W}}$ is a $n \times n$ matrix with rows constructed from the filters $h_{j,l}$ and $\boldsymbol{\mathcal{V}}$ is constructed from $g_{j,l}$, respectively. The exact form of $\boldsymbol{\mathcal{W}}$ is provided in Percival and Walden (2000). This reconstruction scheme can also be used to split up the time series into various *details* $\boldsymbol{\mathcal{D}}$ and a *smooth* $\boldsymbol{\mathcal{S}}$. The technique is realised as a pyramid algorithm. According to this algorithm $\boldsymbol{\mathcal{W}}$ consists of sub-matrices $\boldsymbol{\mathcal{W}}_1, \ldots, \boldsymbol{\mathcal{W}}_J$ and so does $\boldsymbol{\mathcal{V}}$. This defines a *multiresolution analysis* (MRA). For the first scale, e.g., one gets

$$\mathbf{Y} = \boldsymbol{\mathcal{W}}_1^T\mathbf{W}_1 + \boldsymbol{\mathcal{V}}_1^T\mathbf{V}_1 \equiv \boldsymbol{\mathcal{D}}_1 + \boldsymbol{\mathcal{S}}_1 \quad . \tag{2.17}$$

$\boldsymbol{\mathcal{D}}_1$ may be regarded as the variations of $\mathbf{Y}$ at the first scale and $\boldsymbol{\mathcal{S}}_1$ as the variations of $\mathbf{Y}$ up from scale one, i.e. a smoothed version of $\mathbf{Y}$. Applying the pyramid algorithm, $\mathbf{V}_1$ can be split up again, yielding $\mathbf{W}_2$ and $\mathbf{V}_2$ and so forth. Thus, $\mathbf{Y}$ may be reconstructed as

$$\mathbf{Y} = \sum_{j=1}^{J} \boldsymbol{\mathcal{D}}_j + \boldsymbol{\mathcal{S}}_J \quad . \tag{2.18}$$

As an example in Fig. 2.3 the MRA of the chirp signal in Fig. 2.2 is given for levels $j = 1, \ldots, 6$, that is scales from 2 time steps up to 64 time steps. We clearly see that the fluctuations of the signal, which we know to become faster with time, are distributed over all details. In S_6, the "rest", the fluctuation for scales larger than 64 units, is captured.

Maximum overlap discrete wavelet transform (MODWT)

The MODWT (maximum overlap DWT) is a modified version of the DWT. With the MODWT time series of non-dyadic length can be analysed. In contrast to the orthonormal DWT, the MODWT of level J is a highly redundant non-orthogonal transform yielding the column vectors $\widetilde{W}_1$,$\widetilde{W}_2$,....,$\widetilde{W}_J$ and $\widetilde{V}_J$, each of dimension n. Like the DWT, the MODWT is defined in terms of a pyramid algorithm. In this thesis the DWT and the MODWT are used. For more details see Percival and Walden (2000).

[4] $[.]^T$: notation for the transposed vector.

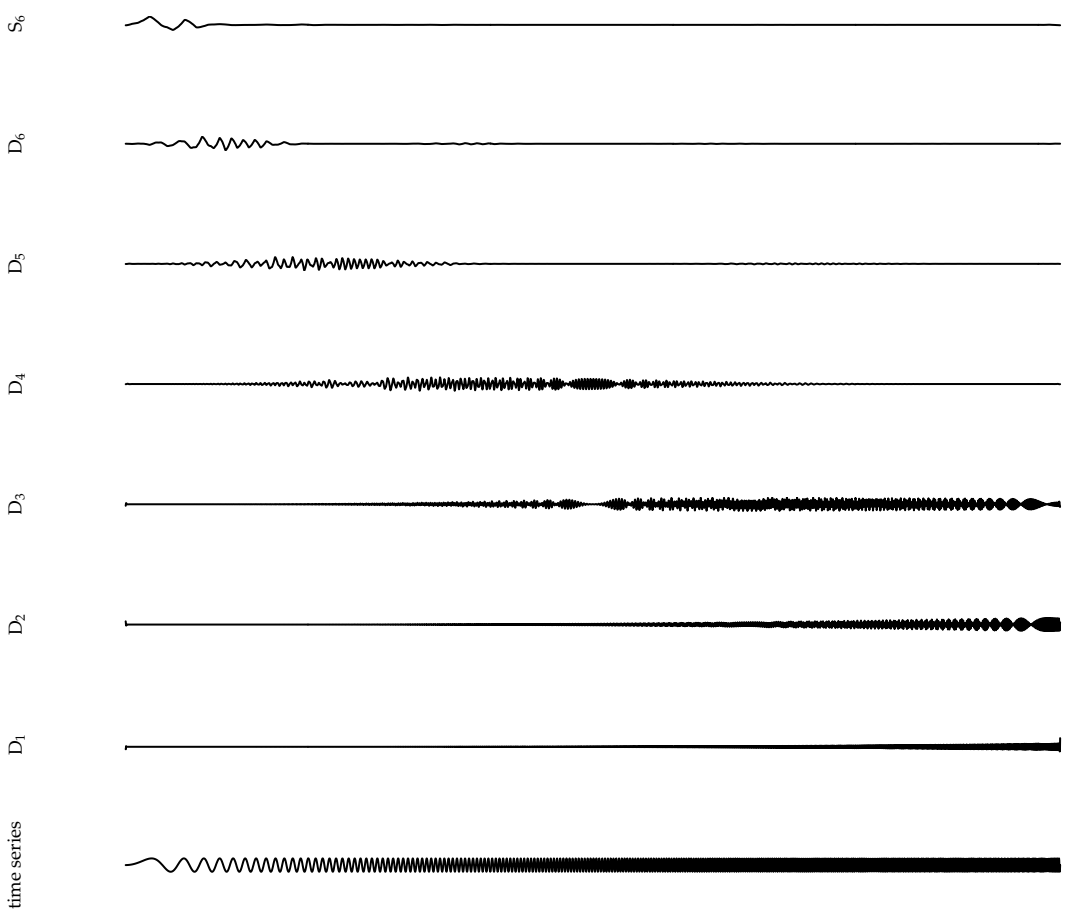


Figure 2.3: MRA of chirp signal. The chirp signal and its MRA for levels $j = 1, \ldots, 6$ are shown.

2.2.2 *Daubechies least asymmetric wavelet filters*

In this work a discrete and orthonormal transform is used, which is constructed from the Daubechies *least asymmetric* wavelet[5] ($LA(\cdot)$) (Daubechies 1992). The special characteristics of the classes of Daubechies wavelet filters is that they can be described in terms of (generalised) differences of adjacent (weighted) averages. Furthermore, the particularity of the class of the least asymmetric wavelet filters is that they get as close as possible to that of a linear phase filter. This practically means that events in a time series at time t_0 also are located at t_0 in the filtered series. They may be smoothed (depending on the width of the wavelet), but they are not shifted in time by filtering. This is important for interpreting the meaning of a filtered series in physical applications. Besides, the $LA(\cdot)$ wavelet filters filter out polynomials. We benefit from this feature when setting up the trend test procedure. The width of the mother wavelet determines the degree of accuracy we either want to achieve in the time or in the frequency domain. The choice of this width depends for example on the roughness of the data analysed. We conducted simulation studies on our empirical river discharge data and have found that the $LA(8)$ wavelet filter of width 8 (see Fig. 2.4) gives the most reliable results. Therefore we use it in all our

[5]The $LA(\cdot)$ is also called *symmlet* in literature.

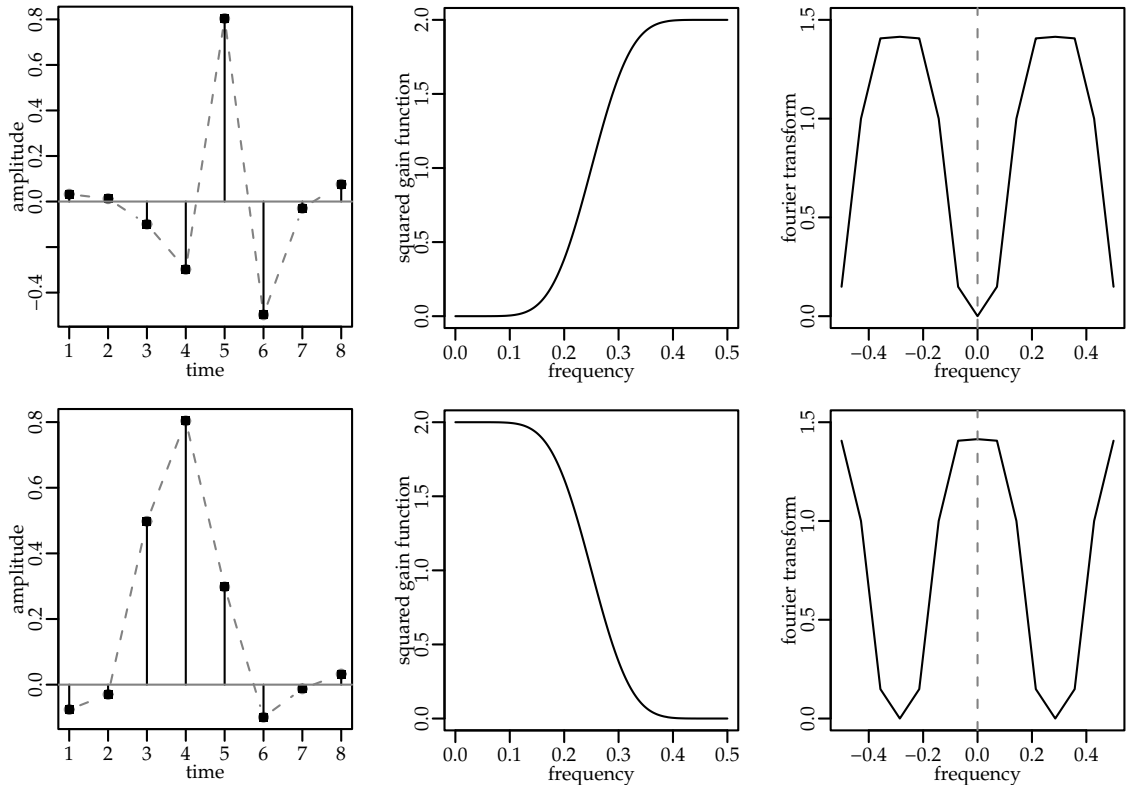


Figure 2.4: Daubechies least asymmetric wavelets. Top row: (**Left**) Non-zero part of the Daubechies least asymmetric wavelet filter of width 8 ($LA(8)$) in the time domain, first scale. (**Middle**) Squared gain function of the $LA(8)$. (**Right**) $LA(8)$ in the frequency domain. **Bottom row:** As upper row, but for the corresponding scaling filter.

analyses.

The application of the $LA(\cdot)$ wavelet filter to a time series essentially yields a contrast between an observation y_t at a fixed time point t and values before and after this observation. The $LA(\cdot)$ scaling filter yields a weighted average of the time series. The width of the smoothing depends on the scale τ_j.

As can be seen in Fig. 2.4 the wavelet filter $h_{1,l} \equiv h_l$ has a certain width in time and frequency and therefore comprises smoothing in both dimensions.

The frequency domain representation of a filter on the Fourier frequencies ω is given by its *transfer function*. Since in general these transfer functions are complex valued, it is convenient to consider their polar representation, namely

$$\text{(i) } G(\omega) = |G(\omega)|e^{i\theta^G(\omega)} \quad \text{and} \quad \text{(ii) } H(\omega) = |H(\omega)|e^{i\theta^H(\omega)} \quad , \tag{2.19}$$

for the scaling and wavelet filter respectively. In this connection $\mathcal{G}(\omega) = |G(\omega)|^2$ and $\mathcal{H}(\omega) = |H(\omega)|^2$ denote the *squared gain functions* and measure the amplitude of the filters, $\theta(\omega)$ is their *phase function*.

The squared gain functions are the same for all Daubechies wavelet filters. They are

given by

$$\mathcal{H}_{1,L}(\omega) = \left| \sum_{l=0}^{L-1} h_l e^{-i2\pi\omega l} \right|^2$$
$$= 2\sin^L\left(\frac{\pi\omega}{2}\right) \sum_{l=0}^{L/2-1} \begin{pmatrix} L/2-1+l \\ l \end{pmatrix} \cos^{2l}\left(\frac{\pi\omega}{2}\right) \quad , \tag{2.20}$$

on the first scale. The associated squared gain function for the scaling filter is retrieved by $\mathcal{G}_{1,L}(\omega) = \mathcal{H}_{1,L}(\pi - \omega)$. The squared gain functions are depicted in the middle figures of Fig. 2.4. Here noticeably the high frequencies are removed by the scaling filter, whereas the wavelet filter erases the low frequencies.

Then, for different phase functions $\theta(\omega)$, distinct Daubechies wavelet filter families are defined. By choosing the transfer function $G(\omega)$ in such a way that the phase function $\theta(\cdot)$ has the smallest maximum deviation from a *linear phase filter*[6], i.e. it is as close as possible to that of a linear phase filter, the definition of a $LA(\cdot)$ filter is completed.

2.2.3 *Wavelet variance*

In case a wavelet transform is orthonormal, the energy of a process $\mathbf{Y}$ is preserved, that is: $\|\mathbf{W}\|^2 = \|\mathbf{Y}\|^2$. In this connection $\|\mathbf{Y}\|^2 \equiv \sum_{t=0}^{n-1} y_t^2$ denotes the squared norm of the vector $\mathbf{Y}$. Let $\{Y_t\}$ be a stochastic process, then the wavelet coefficients $W_{j,t}$ define a new stochastic process that reflects variations in $\{Y_t\}$ on scale τ_j. Assuming that it exists, the *wavelet variance* is defined as the variance of all wavelet coefficients at a certain scale τ_j:

$$\nu_Y^2(\tau_j) \equiv \mathrm{VAR}(W_{j,t}) \quad \text{for } \tau_j = 2^{j-1} \quad , \tag{2.21}$$

for all t at that scale. In case $\nu_Y^2(\tau_j)$ exists for a process $\{Y_t\}$, and it is finite and independent of t, we get $\sigma_Y^2 = \sum_{j=1}^{\infty} \nu_Y^2(\tau_j)$ because of the energy preservation (with σ_Y^2 being the variance of $\{Y_t\}$ itself). The scale-based decomposition of σ_Y^2 is analogous to the frequency-based decomposition given by the Fourier power spectrum.

Plotting $\log(\nu_Y^2)$ against $\log(\tau_j)$ leads to the *global wavelet spectrum*. The global wavelet spectrum provides information about the variance of the process on a scale-by-scale basis. Therefore scales, which are important contributors to the process variance, can be identified by using the global wavelet spectrum (cf. (Torrence and Compo 1998)).

In case $\{Y_t\}$ is a pure self-similar process, its spectral density function (sdf) is approximated by $S_Y(\omega) \propto |\omega|^{\alpha}$. The wavelet coefficients at scale τ_j are associated with frequencies in the interval $[1/2^{j+1}, 1/2^j]$. So, by analysing the global wavelet spectrum, one gets an idea about the shape of the sdf. Using this relationship between the sdf and the wavelet variance also yields a power law for the wavelet variance for increments of self-similar processes, e.g. the FD(δ) process:

$$\nu_Y^2(\tau_j) \propto \tau_j^{-\alpha-1} \quad . \tag{2.22}$$

[6] $\theta(\cdot)$ is a linear phase filter, in case it is a linear function of ω. This excludes the possibility of wraps at $\pm\pi$. A linear phase filter just scales an incoming cosine signal and does not shift it.

By using Eq. (2.22) and the relation $\alpha = -2\delta$, the long-memory parameter δ can be estimated by fitting a straight line to the global wavelet spectrum. α is related to the Hurst coefficient H, i.e. $\alpha = 1 - 2H$ for the defined regions of H.

Daubechies least asymmetric wavelet filters of width L filter out polynomial trends up to an order r with $\frac{L}{2} \geq r + 1$. Therefore the long-memory parameter can also be estimated in case such a kind of trend is present in the data. Further details are provided in Percival and Walden (2000) or Stoev et al. (2005).

2.3 EXTREME VALUE STATISTICS

We give a short overview of extreme value theory to specify the basis of our extreme value assessment, outlined in Chap. 4. A more detailed presentation is provided by, e.g., Leadbetter et al. (1983), Embrechts et al. (1997), Coles (2001) and Katz et al. (2002). The statistics of extremes play an important role in water resources management, e.g. for calculating design values for flood protection buildings, such as dams or dykes. Extreme events are rare by definition. It is thus common in extreme value statistics to fit a theoretical distribution to these events and to derive results from this model. By doing so and by using appropriate parameter estimation techniques, an assessment of the uncertainty of derived statistical indicators, such as quantiles or occurrence rate, can be provided.

2.3.1 Models for block maxima

Let

$$M_m = \max\{X_1, \ldots, X_m\} \tag{2.23}$$

be the maximum drawn out of m random variables X_i, $i = 1, \ldots, m$, which have a common distribution function $F(\cdot)$. We denote maxima obtained in this way as *block maxima*.

A variety of distributions are known to represent these extremes of river run-off data in an adequate manner, e.g. the log-normal, the Pearson Type III, the gamma distribution (see, e.g., Apel et al. 2004, Pakosch 2004, Strupczewski et al. 2001 and Maidment 1993) or the generalized extreme value distribution, which we will discuss in further detail in the following.

Generalized extreme value distribution

The *generalized extreme value* distribution (GEV) has proven to be useful for a wide range of applications. It is theoretically well founded by the powerful Fisher-Tippett theorem (Fisher and Tippett 1928), which has been defined in full generality by Gnedenko (1943).

The Fisher-Tippett theorem states that for identically and independent (iid) random variables $\{X_i\}$ from any of a wide class of distributions M_m converges in distribution for interval size $m \to \infty$ towards the GEV distribution. This theorem also holds for weakly dependent $\{X_i\}$ (see, e.g., Embrechts et al. 1997).

The GEV distribution has a location parameter μ, a scale parameter σ and a shape parameter ξ. Depending on ξ it is divided into three subclasses. With $\xi < 0$ the Weibull

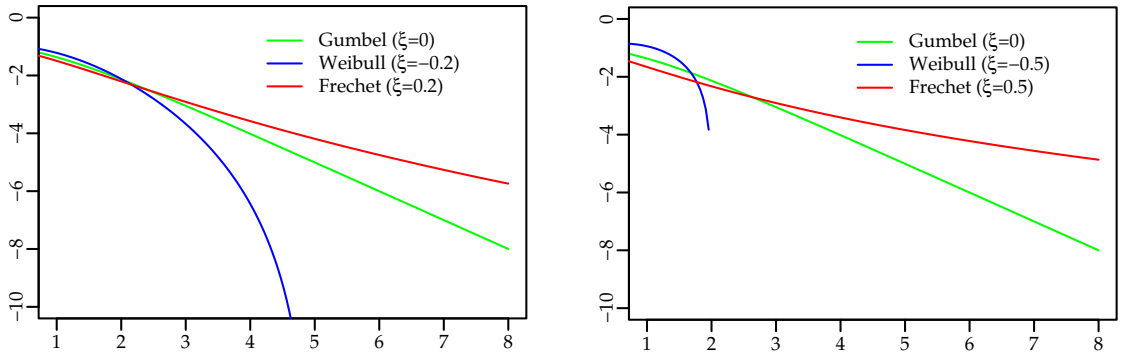


Figure 2.5: GEV distribution. Left: Tails of the Gumbel, Weibull and Frechet distribution with parameters $\mu = 0$ and $\sigma = 1$ (semi-logarithmic depiction). The shape parameter ξ is for the Weibull distribution -0.2 and for the Frechet distribution 0.2. **Right**: Tails of the same distributions are depicted, but with parameters $\xi = -0.5$ for the Weibull distribution and $\xi = 0.5$ for the Frechet distribution.

distribution is obtained, here the magnitude of the extremes is limited by a finite endpoint. For $\xi \to 0$ the occurrence of extreme values decays double exponentially with their magnitude (Gumbel distribution) and for $\xi > 0$ the distribution decays according to a power-law (Frechet distribution). In Fig. 2.5 these subclasses of the GEV are depicted with parameters $\mu = 0$, $\sigma = 1$ and various ξ values. In case the observations lie in the domain of attraction of the Frechet distribution, the extreme value distribution exhibits *heavy tails*. This implies a more frequent occurrence of extremes and an increased occurrence of arbitrarily large extreme events than in the Gumbel case. There is significant evidence that hydrologic variables exist whose extremes possess a heavy tailed distribution. An accordant listing with potential physical explanations, e.g. a certain influence of covariates, is given in Katz et al. (2002) and Katz et al. (2005).

Let M_m be a sequence of maxima drawn from iid random variables with a common distribution function $F(\cdot)$. In case there exist constants $\{a_m > 0\}$ and $\{b_m\}$ such that $P\{(M_m - b_m)/a_m \leq z\} \to \mathrm{GEV}(z)$ as $m \to \infty$, then, according to the Fisher-Tippett theorem, $\mathrm{GEV}(z)$ is a non-degenerate cumulative distribution function and given by

$$\mathrm{GEV}(z) = \exp\Big\{ -\Big[1 + \xi\Big(\frac{z-\mu}{\sigma}\Big)\Big]^{-1/\xi}\Big\} \quad . \tag{2.24}$$

The $\mathrm{GEV}(\cdot)$ is defined on $\{z : 1 + \xi(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

In the limit ($\lim \xi \to 0$) the Gumbel distribution is obtained

$$\mathrm{GEV}(z) = \exp\Big[-\exp\Big\{ -\Big(\frac{z-\mu}{\sigma}\Big)\Big\}\Big], \qquad -\infty < z < \infty \quad . \tag{2.25}$$

The rate of convergence towards the $\mathrm{GEV}(\cdot)$ depends on the marginal distribution $F(\cdot)$ of the X_i. The exponential distribution and the normal distribution, for example, both lie in the domain of attraction of the Gumbel distribution. Nevertheless, the asymptotics is reached much slower for block maxima drawn out of normal distributed data, i.e. a larger interval size m is necessary than for marginal exponential distributed data to obtain adequate results (Zhang et al. 2004; Embrechts et al. 1997).

The use of the Gumbel distribution is very common in hydrology. Several methods have been established to determine whether the simple Gumbel case is sufficient to describe the data or whether the more complex class of GEV distributions has to be used. There exists, e.g., the test of Hosking (see Küchenhoff and Thamerus 2004) or the deviance may be utilised as test statistics (cf. Coles 2002 and Sec. 2.5.2). It is advantageous to start with the more general GEV approach and to reduce this class of distributions to the Gumbel model, in case this is feasible (cf. Cox et al. 2002). The estimation of the GEV parameters, especially the estimation of the shape parameter ξ, is afflicted with uncertainty. Fisher and Tippett (1928) have already stated that for a finite block size m, the extreme values from a stationary Gaussian process are described best by a Weibull distribution, and not by the asymptotically correct Gumbel distribution. Furthermore, choosing the more simple Gumbel model results in much smaller confidence intervals for the estimated parameters. This might result in the underestimation of the uncertainty and therefore misinterpretation (see Coles et al. 2002, for an example). Therefore, in this thesis we first fit the more general GEV distribution to the extremes and, in case a model selection criterion suggests so, we reduce the model class to the more simple Gumbel case.

One restriction for the Fisher-Tippett theorem is that it relies on independent data. Nevertheless, a wide variety of maxima of classes of dependent, but stationary data, also converge towards the GEV distribution - in case the dependence is not too strong. More concretely, a mixing condition of asymptotic independence has to be fulfilled, which limits the long-range dependence being present in the maxima and thus ensures that extreme events are near-independent if they are sufficiently distant in time. The maxima of dependent but stationary data converge to a GEV distribution with other parameters than their independent counterparts. The relation between both parameter sets is expressed via the *extremal index*. Thus, let $X_1, X_2, \ldots$ be a stationary and dependent process, where the condition of asymptotic independence is fulfilled. This may be expressed by the $D(u_n)$ condition, which is satisfied for all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$ with $j_1 - i_p > l$,

$$\begin{aligned} &|P(X_{i_1} \leq u_m, \ldots, X_{i_p} \leq u_m, X_{j_1} \leq u_m, \ldots, X_{j_q} \leq u_m) - \\ &P(X_{i_1} \leq u_m, \ldots, X_{i_p} \leq u_m) P(X_{j_1} \leq u_m, \ldots, X_{j_q} \leq u_m)| \leq \alpha(m,l) \quad , \end{aligned} \tag{2.26}$$

where $\alpha(m,l) \to 0$ for some sequence l_m such that $l_m/m \to 0$ as $m \to \infty$.

Furthermore, let $X_1^*, X_2^*, \ldots$ be a sequence of independent variables with the same marginal distribution $F(\cdot)$ as $X_1, X_2, \ldots$. Define their maxima as $M_m = \max\{X_1, \ldots, X_m\}$ and $M_m^* = \max\{X_1^*, \ldots, X_m^*\}$. Then, in case the $D(u_n)$ condition is satisfied for a specific sequence of thresholds u_m that increase with m, we get according to the Fisher-Tippett theorem, cf. Eq. (2.24), $P\{(M_m^* - b_m)/a_m \leq z\} \to \mathrm{GEV}_1(z)$ for normalising sequences $\{a_m > 0\}$ and $\{b_m\}$ as $m \to \infty$. This holds, if and only if $P\{(M_m - b_m)/a_m \leq z\} \to \mathrm{GEV}_2(z)$, where

$$\mathrm{GEV}_2(z) = \mathrm{GEV}_1^{\theta}(z) \tag{2.27}$$

for a constant θ such that $0 < \theta \leq 1$. This extremal index θ may be interpreted as the reciprocal of the mean cluster size of the dependent maxima (Hsing et al. 1988) and can be calculated in several ways (cf. Ferro 2003). The FARIMA class of auto-correlated data also converges towards the GEV for a moderate δ value (see, e.g., Embrechts et al. 1997 and Davison and Ramesh 2000). Therefore auto-correlations in the data can be considered by using the extremal index. Similar results hold for the GPD and the point process approach (see Sec. 2.3.2).

However, the relation described in Eq. (2.27) is an asymptotic result. In case of short samples or a small interval size m are given, auto-correlations in the data result in broader confidence intervals for the estimated GEV parameters and related values, that is quantiles (Rust et al. 2007).

2.3.2 Threshold excess models

Extracting maxima out of intervals of a fixed length may seem to be a rather artificial way to find the extreme events of a time series. A more intuitive way is to define a threshold u and treat every observation above this threshold as an extreme event. By this means, *excesses* Y_t for a sequence of random variables $\{X_t\}$ are given by $Y_t = X_t - u$, conditional on $X_t > u$.

Peaks over threshold

The *peaks over threshold* (POT) technique models the occurrence of peaks of an iid sequence of random variables $\{X_i\}$ above a threshold u by a Poisson process with rate λ. To be able to do so, the excess values must be independent and have a common distribution function. Then λ only depends on the length of the observed time span. Negative extremes, below a threshold, are examined accordingly. Non-stationary extremes can be modelled by using an inhomogeneous Poisson process with non-stationary occurrence rate λ_t. For this non-stationarity of λ, various shapes may be assumed, see, e.g., Mudelsee et al. (2004). A theoretically well-founded approach suggests the use of the generalized Pareto distribution to model the magnitude of the excesses. This also determines the type of non-stationarity of λ (cf. Leadbetter 1991) and results in point processes to model extremes (see further below). Various authors denote such a more specific point process as POT model (e.g. Parent and Bernier 2003).

Generalized Pareto distribution

The *generalized Pareto distribution* (GPD) is used to model the magnitude of the excesses, not the frequency of occurrence. Let $X_1, X_2, \ldots$ be a sequence of iid random variables with a common distribution function $F(\cdot)$ and let the block maxima $M_m = \max\{X_1, \ldots, X_m\}$ be distributed according to a GEV distribution. Then, for large enough u the distribution function of the excesses Y, conditional on $X > u$, is given by the GPD with parameters ψ and $\tilde{\xi}$:

$$\mathrm{GPD}(y) = 1 - \left(1 + \frac{\tilde{\xi} y}{\psi}\right)^{-\frac{1}{\tilde{\xi}}} \quad . \tag{2.28}$$

The GPD$(\cdot)$ is defined on $\{y : y > 0 \text{ and } (1 + \tilde{\xi} y/\psi) > 0\}$.

For $u \to \infty$, the parameters of the GPD are uniquely determined by those of the associated GEV distribution (see, e.g., Coles 2001). This relation is given by

$$\psi = \sigma + \xi(u - \mu) \quad \text{and} \quad \tilde{\xi} = \xi \quad . \tag{2.29}$$

The accuracy of the estimated GPD parameters depends on an adequately chosen threshold u.

Regarding auto-correlations, consider a marginal distribution $F(\cdot)$, such that a limiting GPD distribution for the maxima M_m of a sequence of iid random variables $\{X_i^*\}$ exists. Then maxima out of a stationary but dependent sequence of random variables $\{X_i\}$, which have the same marginal distribution $F(\cdot)$, often also converge towards the GPD distribution. Here "often" refers to the type of dependence, which still has to fulfil certain criteria, i.e. the maxima have to be asymptotically independent. As outlined in Sec. 2.3.1, these criteria are achieved for long-range dependent FARIMA processes with a moderate value for δ. The limiting distribution of the excesses of the stationary but dependent series is of the same qualitative type as the one derived from the corresponding iid random variables (with respect to ξ). However, the rescaling coefficients a_m and b_m are different. The relation between the parameter vectors of the dependent and the independent case can be formulated by the extremal index (cf. Eq. (2.27)). For further details see Leadbetter (1991) or Cox et al. (2002).

Methods relying on threshold excesses possess the important advantage that autocorrelations can be considered to a certain extend, namely by eliminating them with a declustering algorithm (see Sec. 2.8.2 for more details). By doing so the correlation structure is not explicitly modelled, as in the FARIMA approach (see Sec. 2.1.3). The data are merely thinned out to derive an independent series. This new series converges approximately in the same way as its dependent counterpart.

Point processes

The *point process* (PP) approach comprises the POT approach to model the occurrence of extreme events in time and the Fisher-Tippett framework to determine the magnitude of the excesses by a theoretical distribution. In this way, the occurrence in time as well as the magnitude of the extreme events is modelled using a point process. The point process, which is fitted to the threshold excesses, is realised as a two-dimensional inhomogeneous Poisson process. This implies that, contrary to the POT approach, λ_t has a particular form, determined by the GEV (or GPD) parameters.

The GEV as well as the GPD distribution can be described as special cases of the point process representation (Katz et al. 2002). By means of the relation between GPD and GEV parameters (see Eq. (2.29)), the point process approach is capable to combine advantages of the block maxima and the POT approach: Point process parameter estimates obtained by MLE are based on threshold excesses. Therefore all advantages of the threshold approach can be exploited, that is the consideration of correlations by declustering. Furthermore, the set of threshold excesses often contains the "more extreme" extremes of a time series than the set of annual maxima (see Sec 2.3.3), which leads to more accurate inferences. Moreover σ is invariant to the threshold choice, in case the point process model parameterisation is done in terms of GEV parameters (see, e.g., Katz et al. 2005). A profound overview over point processes is given, for example, in Snyder (1975) or Guttorp (1995).

A point process on a set $\mathcal{A}$ is a stochastic rule for the occurrence of point events. The counting of events occurring in time is a simple example for point processes. Let the *counting process* $N(A)$ be a non-negative integer-valued random variable on a region A of a set $\mathcal{A}$ such that $N(A)$ denotes the number of points in A. Specifying the probability distribution of each of the $N(A)$ in a consistent way determines the characteristics of the

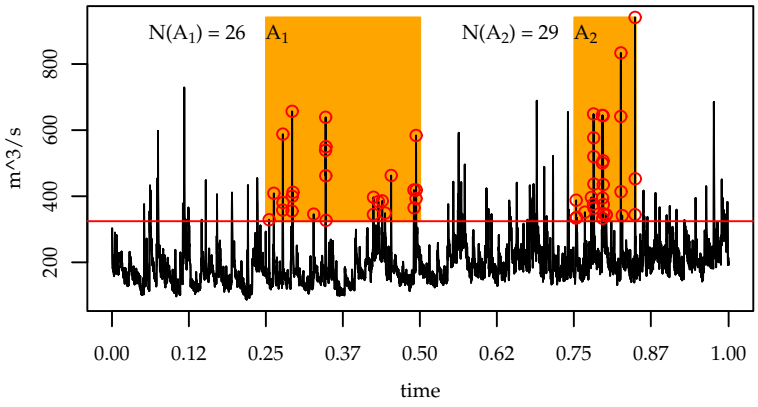


Figure 2.6: Counting process example. Daily discharge of the Inn River at Eschelbach is shown in black (winter season, not declustered) with threshold $u = 324\text{m}^3/\text{s}$ (red, determined with MRL plot and fit over a range of thresholds). The counting processes $N(A)$ count the number of events in subsets of $\mathcal{A} = (0,1) \times [u, \infty]$. Exemplarily the subsets $A_1 \subset \mathcal{A}$ and $A_2 \subset \mathcal{A}$ are shown (orange).

point process. The *intensity measure* $\Lambda(A) = E\{N(A)\}$ of the point process gives the expected number of points in any subset $A \subset \mathcal{A}$. In case the derivative λ of $\Lambda(A)$ exists, we define

$$\Lambda(A) = \int_A \lambda(x)dx \quad . \tag{2.30}$$

λ is the *intensity (density) function* or *rate* of the process.

Now let $\mathcal{A} = [t_1, t_2] \subset \mathbb{R}$ and $\lambda > 0$ be a constant rate. Then the *canonical point process* is the one-dimensional homogeneous Poisson process $\text{Poi}\{\Lambda(A)\}$ defined on this set. The number of points k in A follow a Poisson distribution and we get

$$P(N([0,t_2]) - N([0,t_1]) = k) = \frac{e^{-\lambda(t_2-t_1)}(\lambda(t_2-t_1))^k}{k!} \quad \text{for} \quad k = 0,1,2,\ldots \quad , \tag{2.31}$$

that is $\Lambda(A) = \lambda(t_2 - t_1))$ for all $\mathcal{A} = [t_1, t_2] \subset \mathbb{R}$. The homogeneous Poisson process can be generalised to a model for points that occur randomly in time, but at a time varying rate λ_t.

To apply the point process theory to extreme values, let $\{X_t\}$ be a series of independent and identically distributed random variables with common distribution $F(\cdot)$ such that the block maxima $M_m = \max\{X_1, \ldots, X_m\}$ are distributed according to the GEV for a sufficiently large m (see Sec. 2.3.1). Then define a sequence of point processes N_m on $\mathbb{R}^2$ by $N_m = \{(i/(m+1), (X_i - b_m)/a_m) : i = 1, ..., m\}$ (in this way the time axis is mapped to $(0,1)$). According to the Fisher-Tippett theorem, a convergence law can be obtained for the point processes: For a sufficiently large threshold u, N_m converges for $m \to \infty$ on regions of the form $A := (t_1, t_2) \times [u, \infty)$ with $[t_1, t_2] \subset [0,1]$ to a *homogeneous marked Poisson* process with intensity measure

$$\Lambda(A) = (t_2 - t_1)\Big[1 + \xi\Big(\frac{z-\mu}{\sigma}\Big)\Big]^{-1/\xi} \tag{2.32}$$

on $A = [t_1, t_2] \times [z, z_+]$. Here z_+ is the upper endpoint of the GEV distribution of M_m. A *marked* Poisson process denotes a process, which models occurrence time and position of the events. In Fig. 2.6 exemplarily two subsets A_1 and A_2 of $\mathcal{A} = (0,1) \times [u, \infty]$ are shown for the Inn River at Eschelbach.

The point process counts events in the region above the threshold. This implies that the event $\{(M_m - b_m)/a_m \leq z\}$ is equivalent to the event $N_m(A) = 0$. To obtain maximum likelihood estimates (MLE) of the GEV parameters via the point process approach, the intensity measure given in Eq. (2.32) is inserted in the likelihood function of a Poisson process (see Eq. (2.55)).

It is convenient to factorise $\Lambda(A)$ as two independent Poisson processes, i.e. $\Lambda(A) = \Lambda_1([t_1, t_2]) \times \Lambda_2([z, \infty])$ with

$$\Lambda_1([t_1, t_2]) = (t_2 - t_1) \qquad \text{and} \qquad \Lambda_2([z, \infty)) = \left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi} \quad . \tag{2.33}$$

In this way $\text{Poi}(\Lambda_1([t_1, t_2]))$ gives the random times at which $X_i > u$. With $\text{Poi}(\Lambda_2([z, \infty]))$ the rescaled sizes $(X_i - b_m)/a_m$ of these X_i are calculated, respectively. $\text{Poi}(\Lambda_1([t_1, t_2]))$ is a homogenous Poisson process with intensity

$$\lambda_1(t) \equiv \lambda = \{1 + \xi(u - \mu)/\sigma]\}^{-1/\xi} \quad , \tag{2.34}$$

for $\{1 + \xi(u - \mu)/\sigma]\} > 0$. Apparently, the time rate is related to the GEV parameter vector $\boldsymbol{\Theta} = (\mu, \sigma, \xi)$.

The sizes of the excesses follow an inhomogeneous Poisson process with intensity

$$\lambda_2(z) = \frac{1}{\sigma}\left(1 + \xi\frac{z-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1} \tag{2.35}$$

for $z > u$.

It is also possible to express the point process in terms of the GPD parameters ψ and ξ and the time rate λ. By using the relation between the GPD parameter ψ and the GEV parameter σ (cf. Eq. (2.29)) and the factorisation of two independent Poisson processes to a point process (given in Eq. (2.33)) and by applying Eq. (2.35), we are able to express the point process part, which models the magnitude of the extremes, in terms of the GPD:

$$P\{(X_i - b_n)/a_n > z | (X_i - b_n)/a_n > u\} = \frac{\Lambda_2[z, \infty)}{\Lambda_2[u, \infty)} = \left[1 + \xi\left(\frac{z-u}{\psi}\right)\right]^{-1/\xi} \quad . \tag{2.36}$$

This is equivalent to the distribution of the GPD, i.e. Eq. (2.28). Now the point process is described using the parameter vector $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$. The point process likelihood to estimate these parameters is given in Eq. (2.54).

Regarding auto-correlations, the same limiting behaviour holds as for the GPD case, i.e. stationary but dependent excesses converge towards the same limiting distribution (with a different parameter vector) as their stationary counterparts with the same marginal distribution $F(\cdot)$, in case they are near-independent if they are sufficiently distant in time (cf. Sec. 2.3.2 and Beirlant et al. 2004). The relation between both parameter vectors is given by the extremal index (cf. Eq.(2.27)). Auto-correlations furthermore may be considered using non-stationarity: Exceedances of long-term correlated processes are approximately Poisson distributed, but with a slowly varying rate between time periods (Cox et al. 2002). Like for the other threshold excess models it is also possible to eliminate auto-correlations by declustering (see Sec. 2.8.2).

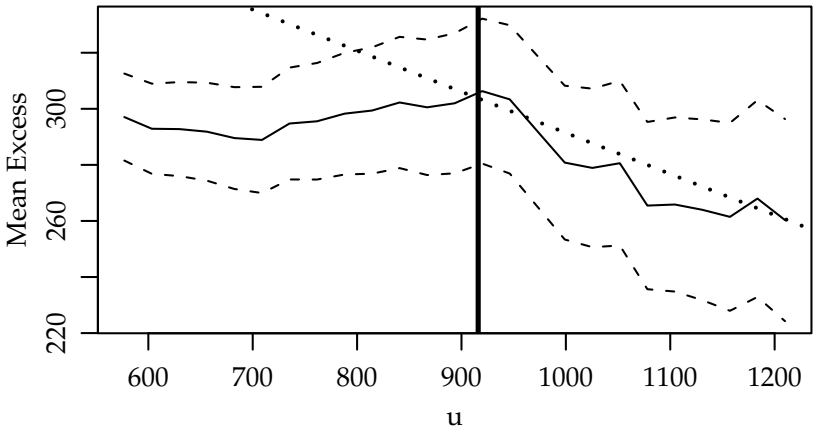


Figure 2.7: Threshold selection. Mean Residual Life Plot. Mean Residual Life plot for the Danube River at Pfelling with confidence intervals based on approximate normality of the sample means. The dotted line indicates the slope of the estimated theoretical mean excess for $u = 916\text{m}^3/\text{s}$, which is marked with the solid, vertical line.

Threshold selection

To be able to fit threshold excess models a reliable threshold u has to be selected. On the one hand, a high threshold is preferred, such that it is reliable to assume a limiting distribution for the excesses. On the other hand, a sufficient number of excesses must be available to estimate the model with a reasonable amount of uncertainty. Therefore a tradeoff between bias and variance has to be found. To operationalise this choice to a certain extent, we apply two heuristic methods. These methods are theoretically well founded, as outlined below, and represent the state-of-the art of threshold selection approaches used in extreme value statistics (cf. Coles 2001). In our analyses of empirical data (see 4.7 and Tab. A-4) we are capable to make a reliable selection in every case by using these methods. The credibility of the choice has been verified by analysing the threshold excesses obtained (see Sec.2.3.3).

The *mean residual life (MRL) plot* is a non-parametric method to determine the threshold. In this plot the sample mean of the threshold excesses for different u is depicted. This sample mean serves as estimate for the expected value of the excesses which holds in case the excesses are GPD distributed. The expected value of the GPD for a certain threshold u_0 exists for $\xi < 1$ and is given by

$$E(X - u_0 | X > u_0) = \frac{\psi_{u_0}}{1 - \xi} \quad . \tag{2.37}$$

Eq. (2.37) should also hold for thresholds $u > u_0$, in case the asymptotics is reached for u_0. From Eq. (2.29) we get $\psi_u = \psi_{u_0} + \xi(u - u_0)$ and together with Eq. (2.37) this yields

$$E(X - u | X > u) = \frac{\psi_{u_0} + \xi u_0}{1 - \xi} + \frac{\xi}{1 - \xi} u \quad , \tag{2.38}$$

i.e. a linear relationship between $E(X - u|X > u)$ and u. Thus we look for the lowest u_0 such that for $u > u_0$ the MRL plot changes linearly in u. u_0 is then regarded as lowest threshold for which the asymptotics is reached. In Fig. 2.7 an example for an MRL plot for discharge of the Danube River at Pfelling is given. Evaluating the plot by following the proposed method we get $u_0 \approx 916\text{m}^3/\text{s}$.

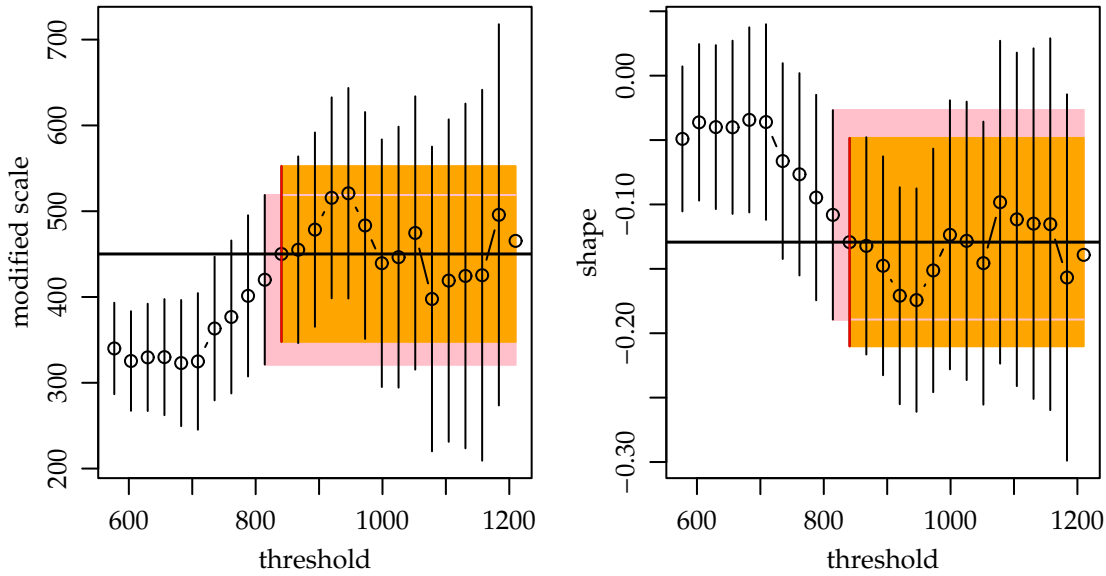


Figure 2.8: Threshold selection. GPD fit over range of thresholds for the Danube River at Pfelling. GPD parameter estimates for a range of thresholds, which are chosen such that about 25% to 3% of the data exceed them. The modified scale parameter $\psi^* = \psi - \xi u$ (**left**) and the shape parameter ξ (**right**) are shown. The parameter estimates for the thresholds (black circles) are depicted with relating 95% confidence intervals (black vertical lines). The lowest threshold, for which the confidence intervals for the parameter estimates include all parameter estimates for higher thresholds, is at approximately $840\text{m}^3/\text{s}$ (CI's are depicted in orange and the respective parameter estimates are accented with horizontal black lines). For comparison reasons the CI for the estimates of the nearest lower threshold are depicted in pink.

Of course the MRL plot is also obtainable by using a point process with GEV parametrisation (see Eq. (2.55)). Here the transformation between GEV and GPD parameters given in Eq. (2.29) can be utilised. In case the time series is too short or noisy, the interpretation of the MRL plot is ambiguous.

As second method, we therefore apply a parametric technique to verify the threshold selection. For this purpose a GPD is fitted for a range of thresholds. Then the parameter estimates and their confidence intervals are compared. We expect the fitted GPD distribution parameters to be stable in case the asymptotic behaviour is reached. More specifically, ξ should stay approximately the same for all thresholds u above a threshold u_0 in case the GPD is a reasonable model. According to Eq. (2.29), the GPD scale parameter changes with u, i.e. $\psi_u = \psi_{u_0} + \xi(u - u_0)$. Therefore we analyse a modified scale parameter $\psi^* = \psi_u - \xi u$ which is also expected to stay constant for all $u > u_0$.

u_0 is chosen as the lowest threshold such that the confidence intervals of the parameter estimates cover all parameter estimates for higher thresholds. As an example, the parameter estimates and corresponding 95% confidence intervals are shown for a range of thresholds in Fig. 2.8 for the Danube River at Pfelling. From this analysis we derive $u_0 \approx 840\text{m}^3/\text{s}$. It is visible that the according confidence intervals for the parameter estimate cover all parameter estimates for higher thresholds, i.e. they lie within the orange rectangle. The confidence intervals for the nearest lower threshold would not include the following parameter estimates, e.g. some estimates lie outside the pink rectangle.

The same methodology certainly is applicable by fitting a range of point processes with GEV parametrisation to the extremes.

For all our studies on empirical data we use both methods to be able to cross-check and to derive reliable results. In the given example we chose the lower threshold of $u = 840\text{m}^3/\text{s}$, which is indicated by the fit over several ranges of thresholds (Fig. 2.8).

2.3.3 *Comparison of block maxima and threshold excesses approach*

In practice threshold excesses as well as block maxima are used to assess extreme values. Both sets therefore should represent the tails of river discharge data reasonably well. The comparison of a GPD fit to threshold excesses and a GEV fit to block maxima does not lead to a concluding decision about which technique has to be preferred: Cox et al. (2002) compare the relative efficiency between the two extreme value approaches by numerical investigations. They find that the goodness of the technique heavily depends on the chosen interval size m and the parameters. The threshold approach always gains efficiency in case the threshold is chosen in a way that more extreme values are available than with the block maxima approach, that is more information can be evaluated. Thus, the choice of block maxima or threshold excesses depends on the characteristics of the empirical data analysed.

For the empirical series used in this thesis, we compared the block maxima obtained with an interval size of $m = 365$ days and the threshold excesses obtained by selecting a threshold according to Sec. 2.3.2. In all cases we find that the set of declustered threshold excesses includes more extreme events of our data than the set of block maxima. As an example, in Fig. 2.9 annual maxima (black) and threshold excesses (red) for $u = 116\text{m}^3/\text{s}$ and $u = 8\text{m}^3/\text{s}$ are depicted for the Iller River at Kempten and Vils River at Pfrontenried, respectively. The thresholds are derived with the methodology outlined in Sec 2.3.2. Before the comparison, the threshold excesses are declustered by using a cluster size of 3 (4) days, which means the set of excesses is thinned out to eliminate correlations (for further details see Sec. 2.8.2). Therefore 54 out of 101 threshold excesses remain for the Iller at Kempten, i.e. nearly as many as the 60 block maxima obtained from the 60 years of measurement. In the case of the Vils River at Pfrontenried 100 extreme values out of 295 remain. In Fig. 2.9 both sets of extremes are compared. The maxima are sorted and for better visibility we start with the most extreme event. Naturally both techniques are able to select the highest values in the time series. Nevertheless, the clustering in the data obviously causes the block maxima technique not to find all highest extreme values available: The block maxima series decays faster than the threshold excesses series and also falls below the threshold. We find a similar situation as depicted in Fig. 2.9 for all analysed gauges, even when a threshold had to be chosen such that the number of threshold excesses is equal or lower than the number of block maxima (which is exemplified in the left hand side of Fig. 2.9). This is caused by the auto-correlations present in the data. Because of these findings, we use a threshold approach to assess extreme events in all our analyses.

2.3.4 *Non-stationary extreme value models*

Conventional extreme value analysis needs the assumption of independent and identically distributed observations, which implies iid extreme values. Both assumptions can

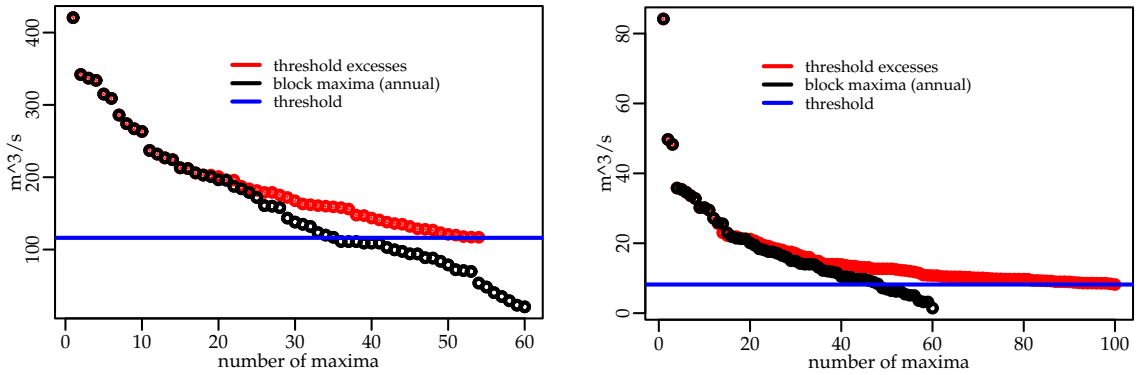


Figure 2.9: Comparison of block maxima and threshold excesses. Left: Sorted annual block maxima (black) and threshold excesses (red) above the threshold $u = 116\text{m}^3/\text{s}$ (blue) for discharge of the Grosse Vils River at Vilsbiburg. **Right**: Sorted annual block maxima (black) and threshold excesses (red) above the threshold $u = 8\text{m}^3/\text{s}$ (blue) for discharge of the Vils River at Pfrontenried.

be relaxed simultaneously: Certain forms of auto-correlations can be handled by declustering (see Sec. 2.8.2) and non-stationary extreme values can be modelled by allowing the parameters of the models to vary with time. All models for extremes, which are presented in this work, can be extended in this way, so they can be used to model non-stationary extreme values.

In all analyses we use a non-stationary point process model with GPD parametrisation, therefore here we present only extensions for this type of model (cf. Sec. 4.7).

As outlined in Sec. 2.3.2, we model the point process by means of two independent Poisson processes, which can be transferred to the combination of a Poisson process and the GPD distribution: $Poi\{\Lambda(A)\}$ models the frequency of occurrence of the extremes. Dependent on the number of occurred extremes the magnitude of the excesses is modelled by using a GPD distribution (according to Eq. (2.36)). Non-stationarity is introduced by allowing ψ, ξ and λ to vary with time, i.e. by using a non-stationary GPD(ψ_t, ξ_t) distribution and an inhomogeneous Poisson process $Poi(\lambda_t)$. We estimate the parameters of the point process by maximising the likelihood (see Sec. 2.4.2). By this means, it would also be possible to introduce a time-varying threshold u_t.

Estimation of the shape parameter

The framework presented is capable to allow for the time-dependence of all three parameters λ, ψ and ξ. However, the variation of ξ around 0 lacks a theoretical interpretation. According to the Fisher-Tippett theorem (see Sec. 2.3.1), different marginal distributions of the random variables $X_1, \ldots, X_n$ have different domains of attraction of the limiting extreme value distribution, that is Weibull $(\xi < 0)$, Gumbel $(\lim \xi \to 0)$ or Frechet $(\xi > 0)$. In case the best fitted model includes a ξ_t which changes sign, this therefore may indicate the usage of too few observations and therefore too large errors of the estimates or a too low threshold, that is the asymptotics has not been reached yet. Another explanation would indeed be the change of the marginal distribution of the $X_1, \ldots, X_n$. Stochastic

models, which incorporate such forms of non-stationarity, are beyond the scope of this work.

We chose to only use models with constant ξ in all our analyses. This is in line with the work of other authors, such as Coles (2001), Cox et al. (2002), and Pakosch (2004). It seems practicable, because even if the assumption of a constant ξ actually is not fulfilled for a specific empirical time series, non-stationary models with a fixed ξ might still be adequate. This is because small time variations of the shape parameter can be emulated by time-varying location and scale parameters (Katz et al. 2002 and Zhang et al. 2004), which is transferable to time-varying rate and scale parameters in case of the GPD parametrisation of the point process model. The goodness-of-fit procedures described in Sec. 2.6 provide information about the feasibility of this reduction of the model class.

In all cases where we find that a non-stationary model suits the extremes best, its estimate of the shape parameter ξ is smaller than the corresponding shape parameter estimate of the stationary model. Our findings are in line with Morrison and Smith (2002), who show that a mixture of Gumbel distributions may cause the estimation of a heavy tailed distribution in case a stationary model is assumed. This implies that in case only stationary models are fitted to the data, falsely a heavy tailed distribution may be assumed and therefore the probability of occurrence of severe floods may be overestimated.

Non-stationary rate and scale parameter

To model extremes, we use point process models with a GPD representation and possibly varying rate λ_t and scale ψ_t. According to the needs of the modeller, all sorts of shapes may be assumed, i.e. linear,

$$\psi_t = \beta_0 + \beta_1 t \quad , \tag{2.39}$$

or exponential,

$$\lambda_t = \exp(\beta_0 + \beta_1 t) \quad . \tag{2.40}$$

More complex structures are also thinkable, e.g., a change-point model,

$$\psi_t = \begin{cases} \psi_1 & \text{for } t \leq t_0 \\ \psi_0 & \text{for } t > t_0 \end{cases} \quad , \tag{2.41}$$

or the modelling of seasonal dependencies (cf. Coles 2001). We use polynomial and exponential shapes to model the non-stationarity of λ_t and ψ_t, which is outlined in more detail in Sec. 4.4.

Return level estimation

Quantiles of extremes, that is *return levels*, are of crucial importance in water management since they are the basis for design flood probability calculations. For the GEV distribution the return level z_p is calculated as

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-p)\}^{-\xi}] & \text{for} \quad \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for} \quad \xi = 0 \end{cases} \quad , \tag{2.42}$$

where (for $m = 365$ days) z_p is expected to be exceeded by the annual maximum in any particular year with probability p. A less precise formulation is that z_p is exceeded on

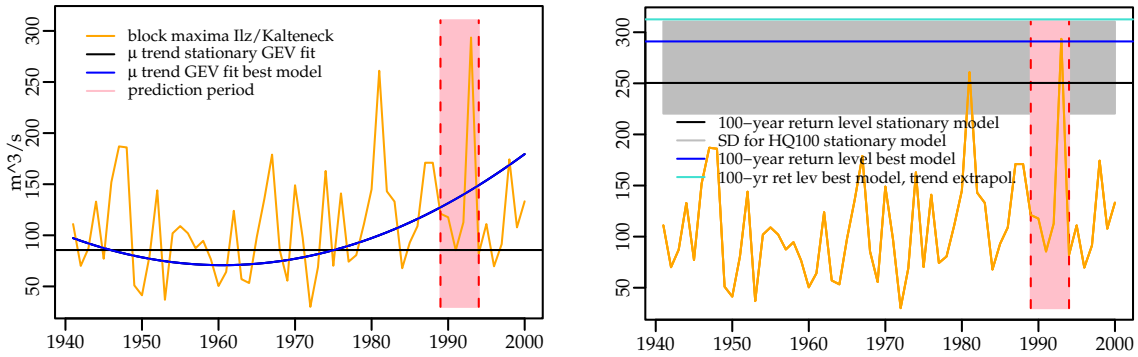


Figure 2.10: Return level calculation. Left: Annual block maxima of the Ilz River at Kalteneck (orange). GEV models are fitted to the data from 1941 to 1989. The evolution of the location parameter μ is shown for the stationary model (black) and the best model according to the deviance statistic (blue, non-stationary model with quadratic trend in μ and constant ψ and ξ). The prediction period from 1989 to 1994 covers 5 years (pink). **Right**: The 100-year return levels for the stationary GEV model (black), the non-stationary GEV model (blue) and the non-stationary GEV model with trend extrapolated within the prediction period (turquoise) are depicted.

average once every $1/p$ years. This notion deals with periods of time, i.e. $1/p$ is the *return period*. It is possible to consider auto-correlations in the data when estimating the return level by using the extremal index. Further details are given in Embrechts et al. (1997). In Fig. 2.10 the annual block maxima of the Ilz River at Kalteneck are shown. A stationary GEV model is fitted to the data of the period 1941 to 1989, that is we pretend not to know the subsequent measurements. Its estimate for the parameter μ is shown in black on the left hand side. The best fitting model assumes a quadratic trend in the parameter μ, which is added in blue, and no trend for the shape and scale parameter. The best fitting model is selected using the deviance statistic, as described in Sec. 2.5.2. The 100-year return level estimate $\hat{z}_{0.01}$, is depicted in black for the stationary case on the right hand side. Its standard deviation is given in grey, calculated by the profile likelihood (cf. Coles 2001, for details).

In the non-stationary case, for every observed time point t a different GEV parameter vector $\boldsymbol{\Theta}$ is possible. Nevertheless, a return level can still be provided by choosing a fixed time point t_0. The return level is then calculated such that it is valid from t_0. To do so, $\boldsymbol{\Theta}(t_0)$ is inserted in Eq. (2.42). In this setting, an estimated trend in the extreme values is considered but not extrapolated into the future. Several authors proceed in this manner (see, e.g., Bárdossy and Pakosch 2005). In Fig. 2.10 we set $t_0 = 1989$ and the according return level estimate $\hat{z}_{0.01}$ considering the trend in the past is shown in blue. Fig. 2.10 clearly shows that the incorporation of non-stationarity in extreme value analysis leads to different results of important assessment measures, such as the return level. Therefore it is a crucial task to model non-stationary extreme values with processes which allow for non-stationarity. The trend shape determines, whether its influence lowers or augments the return level. Results for the Danube River basin are presented in Sec. 4.7.

In case an estimated trend is extrapolated into the future, which is feasible in the fully parametric approach presented here, we have to define a "prediction period". Then all return levels $(z_p^{t_0}, \ldots, z_p^{t_k})$ within this prediction period $[t_0, t_k]$ can be calculated, according to a specified return period $1/p$. The highest return level out of this range of return

levels now is chosen to get an estimator for the return level within this prediction period, which is crossed with probability p once in a year. This method follows the precautionary principle. In Fig. 2.10 the prediction period is set to a length of 5 years, i.e. from 1989 to 1994. The estimate $\hat{z}_{0.01}$ considering the extrapolated trend within the prediction period is shown in turquoise.

In case a probability for the whole prediction period shall be given, we can use the assumption of independence of the extreme events. The probability p' that z_p is exceeded within a prediction period of length $(k+1)$ is given by

$$p' \approx 1 - \{P(X < z_p)\}^{k+1} = 1 - (1-p)^{k+1} \quad . \tag{2.43}$$

We allow for non-stationarity and therefore may have a different z_p for every time point within the prediction period. Thus, we calculate the range of quantiles $(z_p^{t_0}, \ldots, z_p^{t_k})$ fulfilling Eq. (2.43) and again choose the highest out of them. For a prediction period of 100 years and 95% flood protection (i.e. $p' = 0.05$), for example, we have to compute $z_{0.0005}$ over the whole prediction period and choose the highest return level out of this sequence as reference. Regarding the example in Fig. 2.10, the return level, which is crossed only once within the prediction period of 5 years with a probability of 99%, is given by $372\mathrm{m}^3/\mathrm{s}$, which is not depicted.

The calculation of the return level is transferable to the point process approach. In the stationary case we can easily stick to the common unit for return periods, i.e. *years*, though the unit for threshold excesses rather is the *number of observations per time unit*. Let p be the probability that z_p is exceeded once in a year. We calculate p by using the GPD distribution, which depends on the occurrence of the extreme event. This is determined by λ, i.e.

$$p = P\Big\{\frac{(X_i - b_n)}{a_n} > z \Big| \frac{(X_i - b_n)}{a_n} > u\Big\} P\Big\{\frac{(X_i - b_n)}{a_n} > u\Big\} \quad . \tag{2.44}$$

By using Eq. (2.36) and setting $z = z_p$ we get

$$z_p = \begin{cases} u + \frac{\psi}{\xi}[(\lambda/p)^{\xi} - 1] & \text{for} \quad \xi \neq 0 \\ u + \psi \log(\lambda/p) & \text{for} \quad \xi = 0 \end{cases} \quad . \tag{2.45}$$

Thereby we either may use the parameter vector $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$, which is estimated on the whole time series of length n. Then λ denotes the number of excesses for this whole time period. We may also use the adjusted scale and shape parameters ψ and ξ of the GPD, which are linked to the GEV parameters for annual maxima by Eq.(2.29). In this case λ denotes the number of excesses per year to make Eq. (2.45) valid.

In the non-stationary case all three parameters may be time dependent. Therefore we have to determine a fixed time point t_0. Then the return level at this time point can be calculated in the same ways as outlined for the return level obtained by using a non-stationary GEV model. Although the occurrence of extreme events may vary from time series to time series when using threshold excesses, any time point t_0 may be chosen: By using the assumed trend shape with its estimated parameter values, the non-stationary parameters of the extreme value distribution are at hand for each t in $1, \ldots, n$ through interpolation. In all our point process analyses, we estimate the occurrence rate λ and the GPD parameters (which are adjusted to GEV parameters for annual maxima), and insert this parameter vector in Eq. (2.45).

Concerning uncertainty analysis, we approximate confidence intervals for the return level estimates (see Eq. (2.45)) by using the *delta method*. As outlined in Secs. 2.4.2 and 2.4.3, we estimate all parameters by maximising the likelihood and benefit from the consistency and efficiency properties of these estimators. When applying the delta method, we assume a multivariate normal distribution for the parameter estimates and calculate the error propagation by using the variance-covariance matrix of the parameter estimates (cf. Coles 2001).

By the delta method, we get

$$\mathrm{VAR}(\hat{z}_p) \approx \bigtriangledown z_p^T \Sigma \bigtriangledown z_p \quad , \tag{2.46}$$

where Σ is the variance-covariance matrix of $(\hat{\lambda}, \hat{\psi}, \hat{\xi})$. COV($\hat{\lambda}$,$\hat{\psi}$) and COV($\hat{\lambda}$,$\hat{\xi}$) are zero, because we use two independent Poisson processes to construct the point process, as described in Sec. 2.3.2. For $\xi \neq 0$ we get

$$\begin{aligned} \bigtriangledown z_p^T &= \left[\frac{\partial z_p}{\partial \lambda}, \frac{\partial z_p}{\partial \psi}, \frac{\partial z_p}{\partial \xi}\right] \\ &= \left[\frac{\psi}{p^\xi}\lambda^{\xi-1}, \xi^{-1}\left\{\left(\frac{\lambda}{p}\right)^\xi - 1\right\}, \frac{\psi}{\xi}\left(\frac{\lambda}{p}\right)^\xi \ln\left(\frac{\lambda}{p}\right) - \frac{\psi}{\xi^2}\left\{\left(\frac{\lambda}{p}\right)^\xi - 1\right\}\right] \quad , \end{aligned} \tag{2.47}$$

whereas for $\xi = 0$, z_p depends only on ψ and λ and we get

$$\bigtriangledown z_p^T = \left[\frac{\partial z_p}{\partial \lambda}, \frac{\partial z_p}{\partial \psi}\right] = \left[\frac{\psi}{\lambda}, \ln\left(\frac{\lambda}{p}\right)\right] \quad . \tag{2.48}$$

In the non-stationary case, $\mathbf{\Theta}_t = (\lambda_t, \psi_t, \xi)$ itself is time-dependent. We provide the functional form of this dependence, therefore we are able to calculate the derivative of z_p by using the chain rule. For example for $\lambda_t = \exp\{a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4\}$ and the a_i being constants we get,

$$\bigtriangledown \lambda^T = \left[\frac{\partial \lambda}{\partial a_0}, \frac{\partial \lambda}{\partial a_1}, \frac{\partial \lambda}{\partial a_2}, \frac{\partial \lambda}{\partial a_3}, \frac{\partial \lambda}{\partial a_4}\right] = [\lambda, \lambda t, \lambda t^2, \lambda t^3, \lambda t^4] \quad , \tag{2.49}$$

and for $\psi_t = b_0 + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4$ with the b_i being constants, we get

$$\bigtriangledown \psi^T = \left[\frac{\partial \psi}{\partial b_0}, \frac{\partial \psi}{\partial b_1}, \frac{\partial \psi}{\partial b_2}, \frac{\partial \psi}{\partial b_3}, \frac{\partial \psi}{\partial b_4}\right] = [1, t, t^2, t^3, t^4] \quad . \tag{2.50}$$

For $\lambda_t = \exp\{a_0 + a_1 t\}$ and $\xi \neq 0$ the derivative $\partial z_p / \partial a_0$, for example, is given by

$$\frac{\partial z_p}{\partial a_0} = \frac{\partial z_p}{\partial \lambda}\frac{\partial \lambda}{\partial a_0} = \frac{\psi}{p^\xi}\lambda^\xi \quad , \tag{2.51}$$

and in Eq. 2.46 we have to use the variance-covariance matrix Σ of the estimates of a_i and b_i and ξ.

2.4 PARAMETER ESTIMATION

2.4.1 FARIMA process parameter estimation

The parameters of the FARIMA(p,δ,q) models, which describe the auto-correlation structure of the data, are estimated via *Whittle's approximation* to the Maximum-Likelihood

estimator (MLE) (see, e.g., Sibbertsen 1999). The latter is based on minimising

$$Q(\mathbf{\Theta}) = \sum_j \frac{\mathcal{P}(\omega_j)}{\mathcal{S}(\mathbf{\Theta};\omega_j)} \quad , \tag{2.52}$$

where $\mathcal{P}(\omega_j)$ denotes the periodogram of the data at the Fourier frequencies $\omega_j = 2\pi j/N$ and $\mathcal{S}(\mathbf{\Theta};\omega_j)$ the spectral density of the FARIMA process with parameter vector $\mathbf{\Theta}$ comprising $(\phi_1, ..., \phi_p, \delta, \psi_1, ..., \psi_q)$. At the minimum of $Q(\widehat{\mathbf{\Theta}})$, $\widehat{\mathbf{\Theta}}$ is an estimate for the model parameters. The standard deviation provided for the parameters is based on the assumption of a Gaussian process and is calculated from the asymptotic distribution of the estimators in the limit of long records.

An extensive discussion on modelling long-range dependence also with respect to FARIMA(p, δ, q) models and the Whittle estimator can be found in Beran (1994).

2.4.2 *Point process parameter estimation*

In extreme value analysis, which is related to the Fisher-Tippett theorem (see Sec. 2.3.1), the estimation of the model parameters is especially important, because the type of extreme value distribution and therefore the behaviour of the most extreme events depends on the magnitude of the shape parameter ξ.

To estimate the parameters for extreme value models (see Sec. 2.3) we use the maximum likelihood approach. This method is well established for parameter estimation for probabilistic models and very well described in the standard literature (cf. DeGroot 1984 and Cox and Hinkley 1994).

The performance of maximum likelihood estimation of parameters of extreme value models may be very erratic for small samples, say $n \leq 25$ (cf. Coles and Dixon (1999) and Katz et al. (2002)). Such small-sample behaviour can not be ruled out on a theoretical basis, because the justification for ML arises from its asymptotic properties (i.e. for large samples). Therefore probability weighted moment estimators are superior for small samples in estimating upper quantiles. Nevertheless, we restrict our parameter estimation method to ML in this work. We always use empirical data with at least 50 extreme values at hand. Besides, the ML technique is preferable because of its universal applicability, for example to non-stationary data. Furthermore, the ML method automatically provides approximate confidence intervals for parameter estimates, either via the information matrix or by means of the profile likelihood (for details see, e.g., Coles 2001).

To derive the maximum likelihood of a non-stationary point process approximation to extreme values let $X_1, \ldots, X_n$ be iid random variables and let δ_i be an indicator for X_i lying above some given threshold u. Moreover, let $P(X > u) = \lambda/n$, so λ is the rate. The likelihood contribution for an X_i falling below u is thus $1 - \lambda/n$, and for an X_i above u it is λ/n times the GPD approximation with parameters (ψ, ξ), because we actually model excesses, i.e. the conditional probabilities $P(X > y + u | X > u)$.

The density $\mathrm{gpd}(\cdot)$ of $\mathrm{GPD}(y)$ (see Eq. (2.28)) is given by

$$\frac{d\mathrm{GPD}(y)}{dy} = \psi_i^{-1}\left[1 + \xi_i\left(\frac{x_i - u}{\psi_i}\right)\right]^{-\frac{1}{\xi_i}-1} \quad . \tag{2.53}$$

So the likelihood function for sufficiently large u and n is, apart from a scaling factor,

$$\begin{aligned}\mathcal{L}(\lambda,\psi,\xi;x_1,\ldots,x_n) &= \{P(X_i<u)\}^{1-\delta_i}\{P(X_i>u)P(X_i=x|X_i>u)\}^{\delta_i}\\ &\propto \prod_{i=1}^{n}\Big[\Big\{1-\frac{\lambda_i}{n}\Big\}^{1-\delta_i}\Big\{\frac{\lambda_i}{n}\psi_i^{-1}\Big[1+\xi_i\Big(\frac{x_i-u}{\psi_i}\Big)\Big]^{-\frac{1}{\xi_i}-1}\Big\}^{\delta_i}\Big] \quad (2.54)\end{aligned}$$

This likelihood is expressible in terms of the GEV parameters (cf. Coles 2001), i.e. we use $\lambda_i = \{1+\xi_i(u-\mu_i)/\sigma_i\}^{-1/\xi_i}$, approximate the number of not exceeding the threshold as n and approximate $1-\lambda/n$ with $\exp\{-\lambda/n\}$ by utilising $\ln(1-z)\approx -z$. Thus we get

$$\begin{aligned}\mathcal{L}(\lambda,\psi,\xi;x_1,\ldots,x_n) &= \mathcal{L}(\mu,\sigma,\xi;x_1,\ldots,x_n)\\ \propto \exp\Big\{-\frac{1}{n_{\text{py}}}\sum_i\Big[1+\xi_i\frac{(u-\mu_i)}{\sigma_i}\Big]^{-\frac{1}{\xi_i}}\Big\} &\ \prod_{i=1}^{n}\Big\{\sigma_i^{-1}\Big[1+\xi_i\Big(\frac{x_i-\mu_i}{\sigma_i}\Big)\Big]^{-\frac{1}{\xi_i}-1}\Big\}^{\delta_i} \quad (2.55)\end{aligned}$$

n_{py} is the number of observations per year. By scaling with this factor, the parameter estimates (μ,σ,ξ) instantly correspond to the GEV parameters of the annual maxima distribution.

Maximising the likelihood given in Eq. (2.55) leads to estimates for the parameters (μ,σ,ξ) whereas maximising the likelihood given in Eq. (2.54) gives parameter estimates for (λ,ψ,ξ) of the limiting intensity measure $\Lambda(\cdot)$. The likelihood approach provides an innovative way to incorporate non-stationarity in the extreme value model by allowing for non-stationary parameters. This may even include the use of covariates, that is the dependence from other variables in the system.

In all studies presented in this thesis, we use the point process approach in terms of the GPD parameters ψ and ξ and the time rate λ. By using Eq.(2.54) we are able to proceed in two steps: We separately maximise the likelihood for an inhomogeneous Poisson process with λ_t to model the occurrence of extreme events in time and the likelihood of a GPD$(\cdot)$ to model the magnitude of the extremes.

2.4.3 *Consistency and efficiency*

To check the validity of a parameter estimator, or to choose an estimator from a range of possible estimators, several statistical evaluation procedures have been developed in mathematics (cf. Casella and Berger 2002). The most basic concepts are presented in the following together with a classification of the parameter estimators used in this thesis.

A *consistent* estimator converges to the true value as the sample size n becomes infinite, that is bias[7] and variance of the estimator get zero. Let $\{\widehat{T}^n_{t^*}\}$ be a sequence of estimators at the fixed time point t^* and increasing sample size n. Then $\{\widehat{T}^n_{t^*}\}$ is a consistent sequence of estimators for T_{t^*} if

$$\lim_{n\to\infty}\text{VAR}_{T_{t^*}}(\widehat{T}^n_{t^*}) = 0 \quad \text{, and} \quad (2.56)$$

$$\lim_{n\to\infty}\text{BIAS}_{T_{t^*}}(\widehat{T}^n_{t^*}) = 0 \quad , \quad (2.57)$$

for every T_{t^*} in the parameter space.

[7] The bias of an estimator $\hat{\theta}$ is given by $E(\hat{\theta})-\theta$.

The limit of the variance as large sample measure is known to be problematic. So often the estimator is measured against its *asymptotic variance*. The asymptotic variance or variance of the limiting distribution of the estimator $\widehat{T}^n_{t^*}$ is defined as the variance parameter σ^2 of a normal distribution with mean zero in case

$$\lim_{n\to\infty} k_n(\widehat{T}^n_{t^*} - \tau(T_{t^*})) \to N(0, \sigma^2) \tag{2.58}$$

in distribution and $\{k_n\}$ being a sequence of constants and $\tau(\cdot)$ a continuous function. In case an estimator is *asymptotically normal* it is also consistent (Casella and Berger 2002).

Whereas consistency expresses the asymptotic accuracy of an estimator, efficiency assures the existence of an upper bound for its asymptotic variance. An estimator is *asymptotically efficient* in case Eq. (2.58) holds, i.e. its asymptotic variance exists, and if it satisfies the Cramér-Rao Lower Bound (see DeGroot 1984 and Casella and Berger 2002 for further details). Then the estimator is the best unbiased estimator in terms of a small variance.

Thus let $X_1, X_2, \ldots$ be iid random variables with density $f(X|T_{t^*})$. An estimator $\widehat{T}^n_{t^*}$ is asymptotically efficient for a parameter T_{t^*}, if $\sqrt{n}[\widehat{T}^n_{t^*} - \tau(T_{t^*})] \to N(0, v(T_{t^*}))$ in distribution and its asymptotic variance $v(T_{t^*})$ satisfies the Cramér-Rao Lower Bound, that is

$$v(T_{t^*}) = \frac{[\tau'(T_{t^*})]^2}{E_{T_{t^*}}\left((\frac{\partial}{\partial T_{t^*}} \log f(X|T_{t^*}))^2\right)} \quad . \tag{2.59}$$

$E_{T_{t^*}}\left((\frac{\partial}{\partial T_{t^*}} \log f(X|T_{t^*}))^2\right)$ is the *expected Fisher information* of the sample, $\tau'(\cdot)$ denotes the derivative of τ.

It is typically the case that ML estimates for iid random variables are efficient, asymptotically normal and consistent. This holds also for iid random variables with a non-normal distribution, in case this distribution fulfils certain regularity conditions (see, e.g., Casella and Berger 2002).

Yajima (1985) and Dahlhaus (1989) show that ML estimators for Gaussian processes with long-term correlations (which fulfil a few assumptions) are also consistent, asymptotically normal and efficient. This result holds also for FARIMA processes. The estimator has a rate of convergence of $\sqrt{n}$.

Whittle's approximation of the ML estimator, which we use to estimate parameters of the FARIMA processes, has the same asymptotic properties as the exact ML estimator (Sibbertsen 1999). It therefore is efficient, asymptotically normal and consistent for Gaussian processes. Nevertheless, the Whittle estimator loses efficiency in case the data is non-Gaussian (Giraitis and Robinson 2001).

Regarding extreme value statistics, we do not use the Whittle approximation, but the exact maximum likelihood estimators. This is manageable, because we do not have to assess many data points. It is also useful, because the distribution of extreme values commonly is non-Gaussian. The endpoints of the GEV distribution are functions of its parameter values: $\mu - \sigma/\xi$ is an upper end-point of the distribution when $\xi < 0$, and a lower end-point when $\xi > 0$. This means that the usual regularity conditions are violated and the standard asymptotic likelihood results are not automatically applicable, but depend on ξ. According to Smith (1985) we get:

- when $\xi > -0.5$ the MLEs show the usual asymptotic properties,
- when $-1 < \xi < -0.5$ the MLEs exist, but do not have the standard asymptotic properties
- when $\xi < -1$ the MLEs are unlikely to exist.

2.5 MODEL SELECTION

The dynamics of a system, recorded in finite time series, cannot be reconstructed unambiguously. This manifests in the possibility of fitting different stochastic models reasonably well to a time series (Beran 1994). Thus, the basic idea here is to select a suitable model from a set of competing models, which best describes the data – according to some criterion. In this work all parameter estimation is performed by using maximum likelihood or approximations to this (see Sec. 2.4). It can be used as well as selection criterion when choosing between models of the same complexity. To select between models with a different grade of complexity we use methods which evaluate the maximum likelihood. One always ponders between complex models which fit very well to the data but bear the risk of over-fitting, and simple models which possibly represent the data not adequately. The selection criteria reflect this choice by balancing likelihood and number of parameters of the models.

2.5.1 Akaike information criterion

The *Akaike information criterion* is defined by

$$\mathrm{AIC}_f(\kappa) = -2\log \mathcal{L}(\boldsymbol{\Theta}; Y) + f\kappa \quad , \tag{2.60}$$

where $\mathcal{L}(\boldsymbol{\Theta}; Y)$ denotes the likelihood of the parameter vector $\boldsymbol{\Theta}$. The factor f weighs the influence of the penalty term due to the number of parameters κ. If f is set to 2 we get the Akaike Information Criterion (AIC) (Akaike 1973; Akaike 1979), for $f = \log n$ the Bayesian Information Criterion (BIC) (Schwarz 1978), and for $f = 2c \log\log n$ the Hannan-Quinn Information Criterion (HIC), in which we use $c = 1.0001$. The model to select is the one with the smallest AIC_f. For FARIMA(p, δ,0) models BIC and HIC are shown to yield consistent estimators for the autoregressive order p (Beran et al. 1998). For some low order FARIMA(p, δ, q) models Bisaglia (2002) performed simulation studies and recommends BIC and HIC as selection criteria. In this thesis the HIC is used as selection criterion between FARIMA models, see Chap. 3.

2.5.2 Deviance statistic

The *deviance statistic* D relies on the likelihood to compare models (see, e.g., Coles 2001, Nelder and Wedderburn 1972, and Davison and Hinkley 1997). It corresponds to the residual sum of squares in the analysis of a linear regression model. The test based on this statistic is also called *likelihood-ratio test*. The deviance statistic is formulated as follows:

$$D = 2\{l_1(M_1) - l_2(M_2)\} \quad . \tag{2.61}$$

Here $l_i(M_i)$ denotes the maximised log likelihood of model M_i and M_2 is nested in M_1. Under suitable regularity conditions and for a large sample size its distribution follows a χ^2 distribution with k degrees of freedom, where k is the difference in parameters between the two models (Coles 2001). Several authors report good results when comparing non-nested, but reasonable similar models (Zhang et al. 2004 and Katz et al. 2005).

In extreme value statistics the deviance statistic is not only used to choose the type of distribution of the extreme values (e.g. Weibull, Gumbel or Frechet in case of the GEV), but also whether a trend is present in the extreme values or not (see Sec. 4.6). Strupczewski et al. (2001) stress that the distribution function has a greater effect on the AIC value than a trend model. This result probably is transferable to the deviance statistic, because both rely on ML.

In this work the deviance statistic is used to choose between non-stationary extreme value distributions and processes that is the GEV, GPD and PP.

2.6 MODEL VALIDATION – GOODNESS-OF-FIT

With goodness-of-fit procedures we test, whether a given time series is compatible with being a realisation of a model, e.g. a FARIMA model, or an extreme value distribution, or a point process. Thereby a degree of confidence has to be specified to conduct the test. A lot of analytical procedures assume a certain class of models or distributions from which the best is selected via some criterion (see Sec. 2.5). It is not possible to define this class in a way that all potential models, which might be suitable to represent the time series, are included. Therefore the best model out of this class should satisfy a goodness-of-fit test. In the following, the standard goodness-of-fit tests used in this work are discussed briefly.

2.6.1 FARIMA goodness-of-fit test

A goodness-of-fit test for FARIMA(p, δ, q) models formulated in the spectral domain is proposed by Beran (1992). The test statistic

$$T_B = \frac{A}{B^2} \quad , \tag{2.62}$$

with

$$\begin{aligned} A &= \frac{4\pi}{n} \sum_{j=1} \left(\frac{\mathcal{P}(\omega_j)}{\mathcal{S}(\omega_j; \boldsymbol{\Theta})} \right)^2 \quad \text{, and} \\ B &= \frac{4\pi}{n} \sum_{j=1} \frac{\mathcal{P}(\omega_j)}{\mathcal{S}(\omega_j; \boldsymbol{\Theta})} \quad , \end{aligned} \tag{2.63}$$

is asymptotically normal with mean $\mu_{T_B} = \pi^{-1}$ and variance $\sigma^2_{T_B} = 2\pi^{-2}n^{-1}$. This test compares the periodogram $\mathcal{P}(\omega)$ of the empirical time series with the spectral density function $\mathcal{S}(\omega_j; \boldsymbol{\Theta})$ of the fitted model. $\boldsymbol{\Theta}$ denotes the parameter vector of the fitted FARIMA model. It has been shown that this test is equivalent to the Portmanteau test (see Percival et al. 2001). The Portmanteau test is based on the idea that for the correct model

the residuals (in the time domain) should approximately be a realisation of a white noise process.

The smallest significance level for which the null hypothesis "H_0: The empirical data are compatible with being a realisation of the model" is falsely rejected, is denoted by

$$\alpha_{crit} = 1 - \Phi(T'_B) \quad \text{for} \quad T'_B = \frac{T_B - \mu_{T_B}}{\sigma_{T_B}} \quad , \tag{2.64}$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable.

In case a trend is present in the data the proposed goodness-of-fit test is biased: Depending on the shape of the trend the periodogram is likely to be distorted at low frequencies. A possible workaround is to perform the test on the global wavelet spectrum of the $LA(8)$ wavelet (see Sec. 2.2.3). There a possible polynomial trend component (up to a certain order) is removed. Unfortunately, regarding our empirical time series, the DWT provides too few data points to do so.

2.6.2 *Poisson process goodness-of-fit test*

To provide a goodness-of-fit test for inhomogeneous Poisson processes $\text{Poi}(\lambda_t)$, which we use to model the frequency of the occurrence of excesses $Y = X - u$ over a threshold u, we exploit the properties of the Poisson process. For a homogeneous Poisson process $\lambda = \text{MEAN}(\text{Poi}(\lambda)) = \text{VAR}(\text{Poi}(\lambda))$ holds and

$$p_{\text{Poi}} = \frac{\text{VAR}(\text{Poi}(\lambda))}{\text{MEAN}(\text{Poi}(\lambda))} \tag{2.65}$$

should be near to 1. Now we use p_{Poi} as test statistic and generate its distribution by simulation. We generate R series from the possibly inhomogeneous Poisson process, and calculate the corresponding p^*_{Poi} from them, which gives the distribution.

We derive the empirical $p_{\text{Poi(emp)}}$ by first transforming the empirical time series X into a series X_F of zeros and ones, thereby $X_F = 1$ indicates the occurrence of an excess, i.e. $X > u$ and $X_F = 0$ indicates $X <= u$. Then we get $p_{\text{Poi(emp)}} = \text{VAR}(X_F)/\text{MEAN}(X_F)$.

Let the significance level α be 0.05. This is a two-sided test, that is we then agree on the empirical time series being a realisation of the Poisson distribution in case more than $0.025R$ of the p^*_{Poi} lie above and below $p_{\text{Poi(emp)}}$.

2.6.3 *Kolmogorov-Smirnov test*

The *Kolmogorov-Smirnov (KS) statistic* measures the difference of two probability distributions. We use it to determine whether a probability distribution differs from a hypothesised distribution. The KS test is based on finite samples. It has the advantage that no assumption about the distribution of data is needed, i.e. it is non-parametric and distribution free (DeGroot 1984).

The null hypothesis states that a random sample $X_1, \ldots, X_n$ stems from a known continuous cumulative distribution $F(x)$ (with known parameters). The one-sample KS test

compares the empirical distribution function $\hat{F}(x)$ of the sample with this theoretical cdf. Its KS statistic is given by

$$\begin{aligned} K &= \sup_{-\infty<x<\infty} |\hat{F}(x) - F(x)| \\ &= \max\Big\{ \max_{1\le j\le n} \Big\{ \frac{j}{n} - F(X_{(j)}) \Big\}, \max_{1\le j\le n} \Big\{ F(X_{(j)}) - \frac{(j-1)}{n} \Big\} \Big\} \quad , \end{aligned} \tag{2.66}$$

where $\{X_{(j)}\}$ is the order statistics[8] of $\{X_j\}$. The statistic $\sqrt{n}K$ rapidly converges to its limiting distribution as $n \to \infty$. If H_0 is true this limiting distribution is given by

$$\lim_{n\to\infty} P(\sqrt{n}K \le t) = 1 - 2\sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2t^2} \quad . \tag{2.67}$$

The quantiles of this distribution commonly are listed in tables (see, e.g., Sachs 1991). A test for significance of the null hypothesis is established by assuming a certain significance level and verifying whether $\sqrt{n}K$ passes the corresponding quantile.

It is known that the Kolmogorov-Smirnov test has low power for the typically skewed distributions of extreme values (Strupczewski et al. 2001). Besides that, the KS test is sensitive to auto-correlations in the data (Dutilleul and Legendre 1992). Perhaps the most serious limitation is that the distribution must be fully specified. This implies that Eq. (2.67) is no longer valid in case the parameters are estimated from the data. Therefore we provide a test distribution by simulation. We use the KS test to evaluate the goodness-of-fit of the GPD distribution. Let $X_1, X_2, \ldots$ be a series of iid random variables with block maxima M_m converging to the GEV for $m \to \infty$. Then the distribution of the according threshold excesses $Y = X - u$, conditional on $X > u$, is given by the GPD (see Eq. (2.28)). To normalise Y_i, we transform it to

$$\tilde{Y}_i = -\log\Big\{1 + \xi\Big(\frac{Y_i - u}{\psi}\Big)\Big\}^{-1/\xi} \quad \text{for} \quad i = 1, \ldots, n \quad . \tag{2.68}$$

$\tilde{Y}_i$ is distributed according to the exponential distribution, $\tilde{Y}_i \sim 1 - \exp^{-\tilde{y}_i}$. The goodness-of-fit of a stationary GPD model is tested by assessing whether the transformed $\tilde{Y}_i$ can be regarded as a realisation of an exponential distribution using the KS test. This procedure easily is transferred to non-stationary GPD models. Then Eq. (2.68) with non-stationary parameters is applied to transform the excesses and the resulting $\tilde{Y}_i$ are as well compared to the exponential distribution. To simulate the distribution of the KS test statistic, we generate an ensemble of R time series from the GPD distribution, which is compared with the empirical excesses. Then we fit the same GPD model to the simulated series, conduct the KS test and keep the p-value at a level of $\alpha = 0.05$. This ensemble of simulated p-values provides the distribution.

2.6.4 *Probability plot and quantile plot*

The *probability plot* (pp-plot) and the quantile plot (qq-plot) are widespread tools to visually assess the validity of an estimated distribution function $\hat{F}(\cdot)$ by comparing it to the empirical distribution function of the data. Given an ordered sample of independent

[8] The order statistics of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order.

observations $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ and an estimate of the distribution function $\hat{F}(\cdot)$, a probability plot consists of the points

$$\left\{ \left(\hat{F}(x_{(i)}), \frac{i}{n+1} \right) : i = 1, \ldots, n \right\} \quad . \tag{2.69}$$

The points of the pp-plot have to be located close to the unit diagonal in case $\hat{F}(\cdot)$ is a good model for the data. As for a simple non-parametric estimate of the empirical density function, a probability of $1/(n+1)$ is attached to each of the x_i. With the pp-plot it is verified whether the estimated parametric distribution function matches this non-parametric estimate.

A *quantile plot* is given by

$$\left\{ \left(\hat{F}^{-1}\left(\frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \ldots, n \right\} \quad . \tag{2.70}$$

$x_{(i)}$ and $\hat{F}^{-1}(i/(n+1))$. Both provide estimates for the $i/(n+1)$ quantile of the distribution $F(\cdot)$. Therefore the quantile plot should consist of points close to the unit diagonal in case $\hat{F(\cdot)}$ is a good estimate for $F(\cdot)$.

2.7 DATA

In this thesis, measurements of daily mean river run-off (measured in cubic meters per second [m^3/s]) are used. The area of investigation is located in Southern Germany comprising the catchments[9] of the Danube River and the Neckar River[10].

We selected this area on behalf of the department for water management of the Bavarian Environmental Protection Agency, which wants to use further our methodology and results. Long measurements exist for this region, the longest from 1901 to 2003. From the Neckar basin 10 data records are used, which all cover the period from 1940 to 2003. The length of these records is depicted in Fig. 2.11 (red). The sub-catchments near the Danube River basin provide about 90 records for our analysis purposes. The time periods which are covered by these series are depicted on the right hand side of Fig. 2.11. Out of them, we chose the time period between 1941 and 2000 (which is highlighted with an orange hatched area) to perform comparative trend analysis studies. For this purpose 50 series out of the 90, which jointly cover this period, were selected.

The sizes of the analysed catchments in the Neckar River basin range from 6km^2 to 3 995km^2, as depicted in Fig. 2.12. Thereby most of the catchments have a size between 100km^2 to 500km^2. For the Danube River the 90 series analysed are located in catchment areas with sizes from 91km^2 to 143 600km^2, as shown in Fig. 2.13. For a better classification the size of the 70 smallest catchments is also depicted on the right hand side of Fig. 2.13. Most of the river catchments have a size between 50km^2 and 1 200km^2. For a geographical coordination, the location of the stations is shown in Fig. 2.14. In several cases where we were mainly interested in the large-scale evolvement of the time series

[9] River catchments are regarded as the domain of attraction of water at a given geographical position.

[10] The data has been made available by the department for water management of the Bavarian Environmental Protection Agency (*http://www.bayern.de/lfw/*), which is concerned with topics such as the protection of the environment, geology and water management of Bavaria.

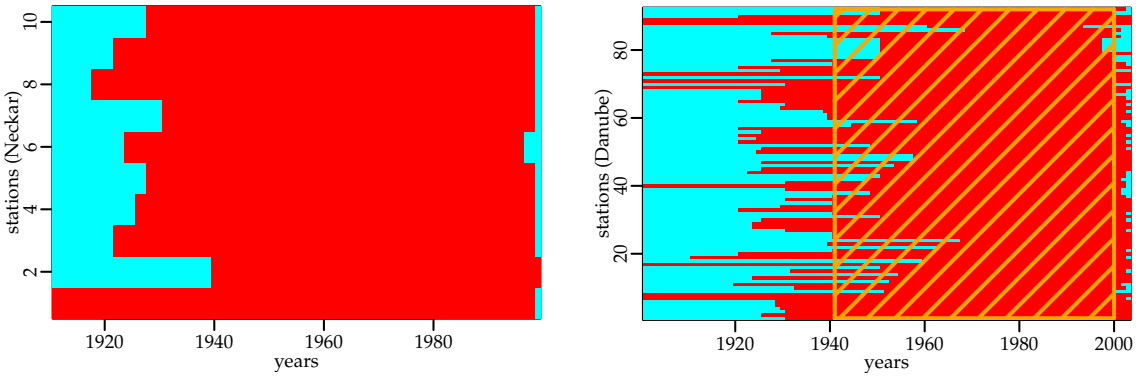


Figure 2.11: Length of time series. Left: Time periods covered by examined time series in the Neckar basin (red). **Right**: Time periods covered by examined time series in the Danube basin (red). The hatched area (orange) highlights the period of 60 years between 1941 and 2000, which is used to perform comparative analysis with about 50 stations out of the 90 depicted.

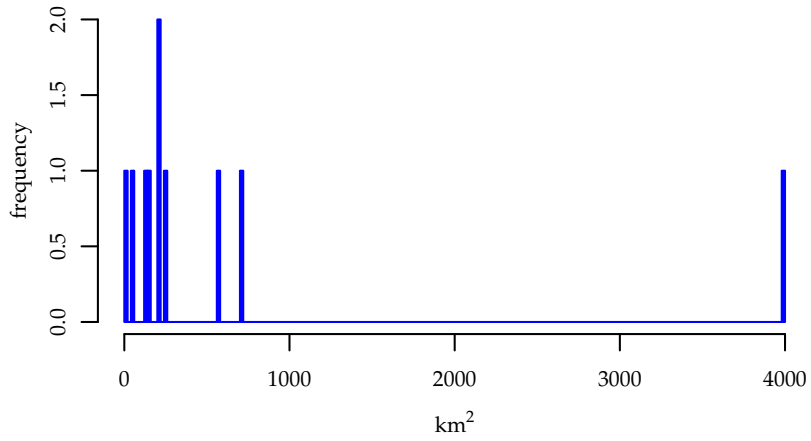


Figure 2.12: Size of Neckar River sub-catchments. Histogram of the sizes of catchments analysed near the Neckar River (in km^2).

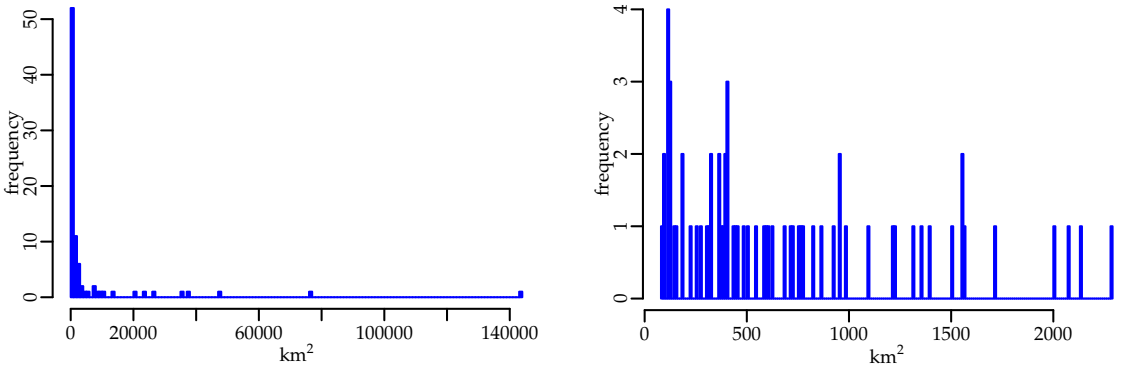


Figure 2.13: Size of Danube River sub-catchments. Left: Histogram of the sizes of river catchments near the Danube River related to the 90 gauges analysed in this thesis (in km^2). **Right**: Histogram of the sizes of the smallest 70 catchments.

we aggregated the daily mean discharge data to monthly means, monthly maxima or annual maxima and then analysed these series. To provide reliable results, measurement records with too many or too long gaps were disregarded. In case there were only a few daily measurements missing, they were replaced by the average of the according days in other years obtained from the remaining series. In case aggregated data was used, that is monthly maxima, the gaps are allowed to be up to 10 days long.

The regions have encountered various floods during the last decades, i.e. in May 1999, August 2002 and August 2005. This goes along with a change in the precipitation pattern in this area. Heavy winter precipitation events become more frequent in Bavaria and Baden-Württemberg (Dietzer et al. 2001) and are anticipated to become even more frequent in the future (Schröter et al. 2005). Local water management authorities expect an increase in heavy floods due to climate change. As a consequence, a *climate change factor* of 1.15 was introduced in Bavaria in 2006. Due to this factor, every design value for river discharge is expanded about 15% (BlfW 2006).

The region provides a good data basis for the aims of this thesis regarding completeness and statistical characteristics, such as trends and auto-correlation structure various analysis purposes. It has already been chosen for trend assessment studies by other projects such as KLIWA (2003).

2.8 PREPROCESSING THE DATA

2.8.1 Seasonal effects

Hydrological data underlies the annual cycle, that is seasonal effects. This deterministic, periodic non-stationarity has to be treated differently than the stochastic part of the time series. This holds in particular if analysis methods require independent and identical distributed (iid) data. These seasonal effects can either be incorporated by seasonal models (see, e.g., Lohre et al. 2003 and Montanari et al. 2000), neutralised by the classification of the data or removed by deseasonalisation. For our goals it served best to analyse seasonal extreme values and deseasonalised mean values, as outlined below.

Data classification into seasons

We eliminate seasonality in our extreme value data by dividing them into seasons. The division of data into seasons or even shorter time periods, that is the analysis of only a sub-sample of the data series is performed for several purposes. Changes within the year may be investigated, e.g. an earlier start of the snowmelt can be tracked by examining the change in winter floods (see, e.g., Mudelsee et al. 2004) or winter precipitation (Robson 2002). Furthermore, a different impact of various processes on hydrometeorological data may be assumed for different seasons. Therefore the evolution of a special season may be in the focus. By looking only at this season, effects due to the influence of other seasons are avoided.

We analyse the winter season (December, January and February). This season is especially interesting for trend detection issues, several studies suggest a change in winter

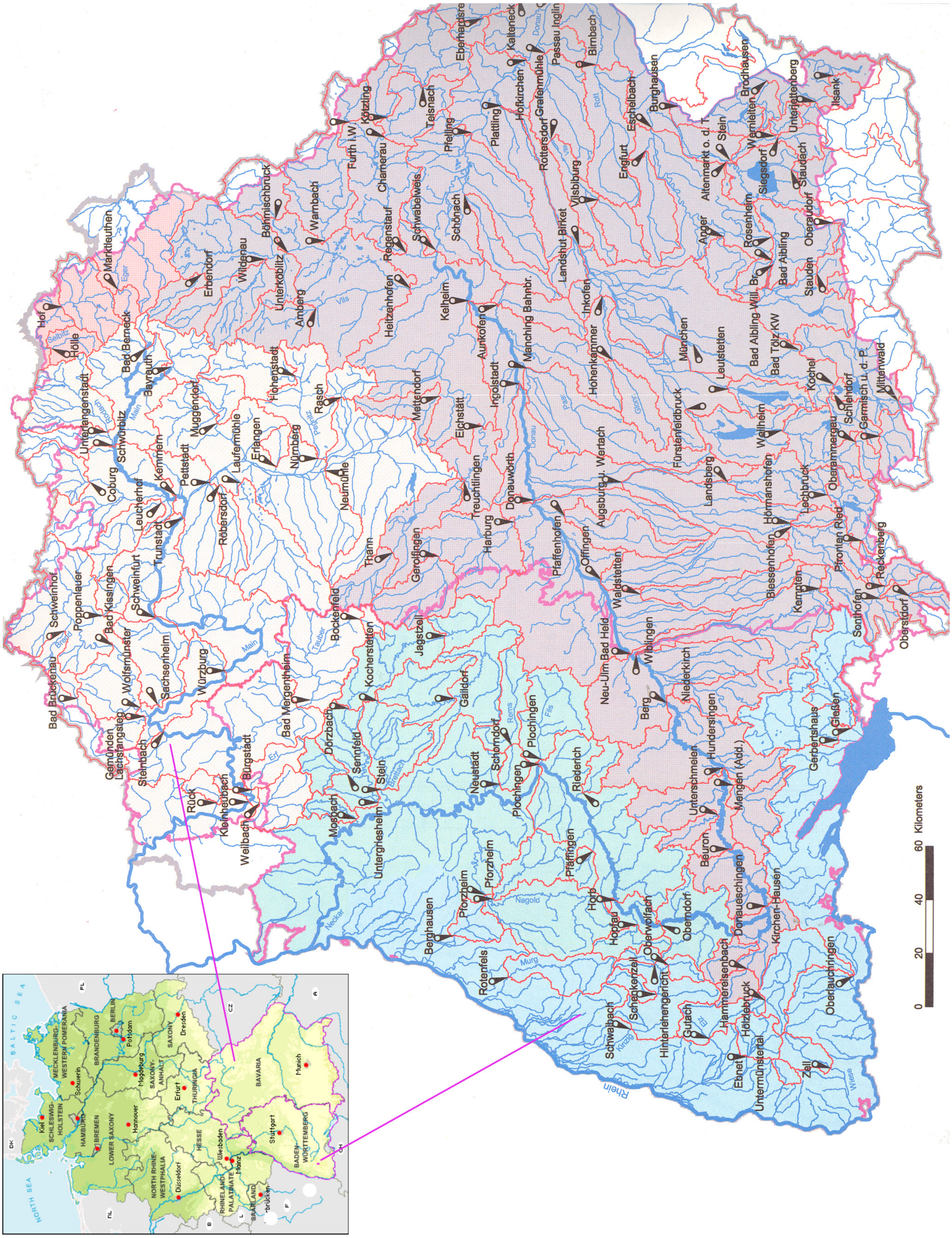


Figure 2.14: Gauges in Bavaria and Baden-Württemberg. The Danube River basin is depicted in grey and the Neckar River basin in turquoise (*source: KLIWA (2003), page 89*).

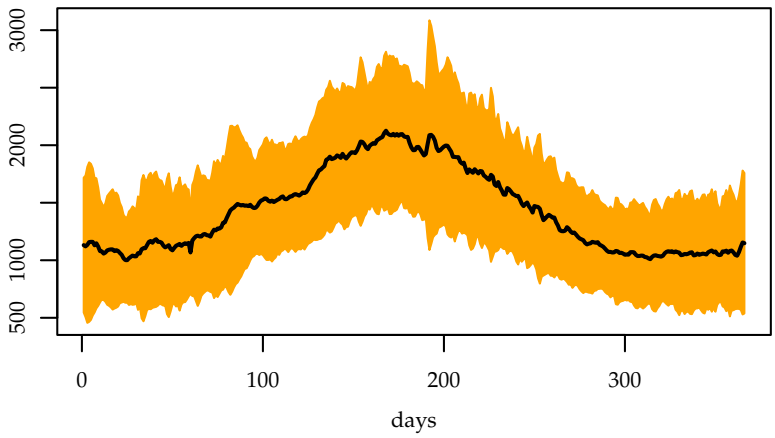


Figure 2.15: Reference year. Reference year of daily measurements from the Danube River at Achleiten (1901-2003). The mean of the year is depicted in black. To deseasonalise the data, it is subtracted from the data. Normalised data are divided by its standard deviation, which is shown in orange.

floods (Kundzewicz and Schellnhuber 2004). Furthermore the assumption of stationarity within this season is justified, which we tested in a simulation study by comparing the inter-seasonal variation of all four seasons. Defining the winter season as the period from December to February is done by several authors (see, e.g., Jacobeit et al. 2003). The classification into seasons even is no obstacle to finding an extreme value model, which fits the extreme value data for the whole year: By estimating the parameters of the models using the likelihood function, it is possible to re-merge the models found for the different seasons and to determine a best model for the whole year in terms of maximum likelihood (Coles 2001).

Deseasonalisation

To *deseasonalise* the mean values, a *reference year* is created by calculating the average and variance of each day of the year over all years. Then the strong periodicities of the annual cycle are approximately removed from the daily measurements by subtracting the average. We furthermore *normalise* the data, that is we divide each day by the standard deviation of the respective day in the reference year. By doing so, the seasonal periodicity in the variance is removed, see, e.g., Hipel and McLeod (1994). FARIMA(p, δ, q) models do not reflect time-dependent variance (heteroscedasticity), therefore the results for the series normalised in this way are expected to be more reliable. An example for the reference year of the Danube River discharge measured at the gauge Achleiten is given in Fig. 2.15.

2.8.2 Declustering

A lot of data evaluation methods assume independent data. Therefore we model auto-correlations in the mean values explicitly with FARIMA(p, δ, q) models. Regarding extreme value analysis, modelling the auto-correlation structure is difficult due to the lack of data. We therefore eliminate the auto-correlations in the extremes via so-called *declustering* (Embrechts et al. 1997; Chavez-Demoulin 1999).

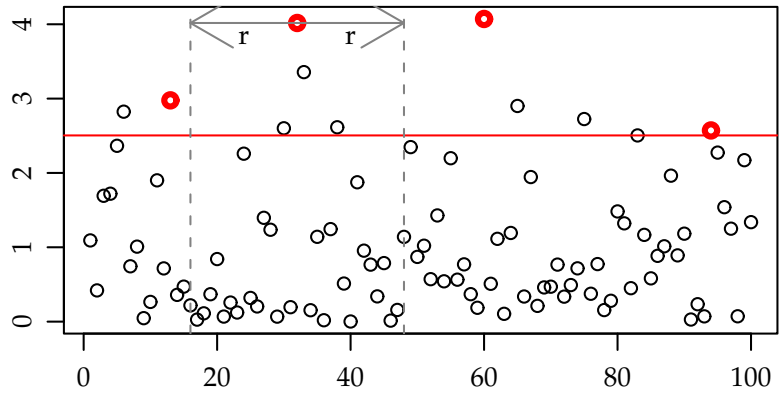


Figure 2.16: Declustering algorithm. Simulation data from a GPD distribution is shown with a threshold u, which is crossed by 10% of the data (red). The excesses of clusters with cluster size $2r = 20$ are highlighted in red.

Declustering can be done for extreme values which are defined as exceedances over a certain threshold u. To do so, the *cluster size* $2 * r$ has to be found (Ferro 2003). It is assumed that observations which lie further apart in time than this cluster size are independent. Actually long-range dependent data with a moderate fractional difference parameter δ can be decorrelated in this way (Embrechts et al. 1997). As depicted in Fig. 2.16 the exceedances are grouped into the same cluster if they are separated by fewer than r values below u. For the analysis, we keep only the highest exceedance of each cluster, i.e. the red data points in Fig. 2.16. In practice, we compare the result of the declustering algorithm for various cluster sizes and keep the smallest r, for which we see that the number of remaining extremes is nearly the same as for larger r. For the discharge analysed, we find r values between 3 days and 6 days to be sufficient. We furthermore test the resulting set of extreme values for independence.

CHAPTER 3

TRENDS IN MEAN VALUES CONSIDERING AUTO-CORRELATION

3.1 TREND DETECTION APPROACHES

Different potential driving forces have to be considered when analysing hydrological or meteorological time series. The distinction of such components is still a matter of research. The approach that we will introduce herein may be used for this purpose by making the rather simple assumption that systemic changes can be represented by a deterministic, slowly changing trend component, whereas natural variability is modelled by a stochastic process (see, e.g., Bloomfield 1992 and Smith 1993). At any rate it is desirable to distinguish between these two components in a time series, because a trend is expected to continue in the future whereas natural variability changes direction in a more or less unpredictable manner. In case a deterministic trend component is identified in the data, its origin and interpretation is yet another important task, which is domiciled in different application areas, such as hydrology or meteorology.

A variety of frameworks to assess trends under consideration of auto-correlations have been established. In all approaches the trend component and the stochastic part, which reflects the natural variability present in the data, are combined additively. The techniques are distinguished by the way the trend is estimated and by the assumptions made for the auto-correlation structure, i.e. the class of stochastic models which are used to model the auto-correlations. Bloomfield (1992), Woodward and Gray (1995) and others evaluate trends in global temperature by combining a trend and an autoregressive integrated moving average (ARIMA) component in a fully parametric approach, that is the assumed trend shape is also derived via parameterised functions. For several decades the examination of long-term correlations has drawn more and more attention. This sort of auto-correlation structure, which persists over very long (theoretically infinite) time periods, has been found to be present in river run-off data (e.g. Hurst et al. 1965, Lawrence and Kottegoda 1977 and Montanari et al. 1997). The effect of long-term correlated processes on trend estimation is well described (c.f Beran 1994). Also parts of hydrological statistics have already been adapted to this long-term correlation phenomenon (Koutsoyiannis 2003). Smith (1993) utilises the standard least squares estimator to test for trends under consideration of long-term memory. Sibbertsen (1999) uses S-estimators to assess a trend for data with long-term correlated errors. Percival and Walden (2000) and Craigmile et al. (2004) employ wavelets to test for the significance of a trend under assumption of a long-term correlated fractionally differenced FD(δ) or a short-term correlated autoregressive AR(1) model. All these methods possess advantages and disadvantages concerning a correct estimation of the trend and the provision of confidence

intervals for this estimate. We contribute to the evaluation of trends in the mean by combining a trend estimate obtained by wavelet filtering and the class of fractional ARIMA (FARIMA) models to assess trends in a semi-parametric approach.

3.2 TREND DEFINITION

There exists a wide range of trend definitions, depending on the methods used, the purpose, and the area of application. Therefore no universal specification is sketched here, but rather a definition is given, which enables us to set up a trend assessment framework.

Definition 3.1 (Trend)**.** A *trend* is regarded as a continuous slowly varying change in the mean of a time series, which has an impact at large scales.

Craigmile et al. (2004), Percival and Walden (2000), and Bloomfield (1992), for example, share this view of a trend as fluctuations on large scales. In this connection a trend denotes a non-stationarity in time which may also consist of periodicities, e.g. the annual cycle, for example. We do not restrict our trend definition to trends with an increasing or decreasing tendency. This notion especially suits river discharge data, where non-monotonic trends are rather common. Change points in the data are not included in this definition and therefore have to be removed before the analysis.

We furthermore assume that every observed time series $\{y_t : t = 0, ..., n-1\}$ can be modelled as a linear composition of a deterministic trend component $\{T_t\}$ plus a realisation of a stochastic process $\{X_t\}$:

$$Y_t = T_t + X_t \qquad t = 0, \ldots, n-1 \quad . \tag{3.1}$$

3.2.1 Stochastic versus deterministic trends

As presented in Secs. 2.1.2 and 2.1.3, stochastic auto-correlated processes such as ARMA or FARIMA processes may exhibit *stochastic trends*, i.e. they are non-stationary stochastic processes and do not satisfy the conditions for strict stationarity (cf. Eq. (2.3)) or weak stationarity in the case of FARIMA processes having a Gaussian noise term. Non-stationary stochastic processes have time dependent moments. The *random walk* is a prominent example for a stochastic process exhibiting a stochastic trend. Its variance is time dependent.

Contrary to this, a deterministic trend is assumed to be due to non-stochastic mechanisms. Therefore it is modelled independently of the stochastic portion of the analysed dynamics. The importance of the detection of deterministic trends partly lies in the unequal treatment of deterministic and stochastic non-stationarities. Realisations of a non-stationary stochastic FD(δ) process, for example, are transferred to a stationary model by (multiple) differentiation (see Sec. 2.1.3 for further details). On the other hand, time series with a deterministic trend component become stationary by subtracting the trend component. The disregard of this difference has more or less severe consequences (see Rinne and Specht 2002): A deterministic trend which is not detected may bias the parameter fits in a way that a non-stationary stochastic model is wrongly chosen as the

one fitting best. The Dickey-Fuller test is designed to distinguish between non-stationary short-term correlated models (unit-root-models) and deterministic trends. The performance of the Dickey-Fuller test is reduced when it is applied to long-term correlated data (Hassler and Wolters 1994).

Within the climate community, there exist several approaches to model temperature time series by using stochastic non-stationary models. Woodward and Gray (1993) and Woodward and Gray (1995), for example, find an ARIMA(9,1,0) process which possesses roots lying on the unit circle (see Sec. 2.1.3 for further details) as best fitting model for the global temperature data reconstruction provided by Hansen and Lebedeff (1987) and Hansen and Lebedeff (1988). They favour this process over a combination of a stationary ARMA model with a linear deterministic trend component. Smith (2001) in contrast argues in a process oriented way saying that (a) the assumption of a non-stationary stochastic process and (b) the possibility of arbitrary far shifts from their starting values over a long time period are questionable assumptions for climatic variables.

The incorporation of non-stationary stochastic models in our approach is theoretically possible. Nevertheless, we have always found stationary stochastic models to fit best to the analysed empirical data (for results see Sec. 3.5). This is indicated by the parameter estimates (see Secs. 2.1.2 and 2.1.3). Thus, in this work mean values of river discharge are either described by a stationary stochastic model or by a stationary stochastic model with a deterministic trend component and the model selection can be reduced to these cases.

3.2.2 Trends and auto-correlation

Smoothly varying trends are difficult to distinguish from natural variability on large scales caused, e.g., by auto-correlation. Bhattacharya et al. (1983), Taqqu et al. (1995) and Bunde et al. (2002) address relevant pitfalls and provide possible ways of distinction. This exchangeability implies that the evaluation of a possible deterministic trend implicitly needs assumptions about the natural variability: Trends of small magnitudes will less often be considered as significant in a system which is known to be highly variable anyway. This fact is reflected by the selection of a stochastic model capturing the natural variability. The choice has an important influence on the confidence interval estimated for the trend parameters and therefore on the significance of the trend estimate (see Sec. 3.3.4 for further details).

Particularly long-term correlation, e.g. long-range dependence, causes long excursions from the mean (Beran 1994). Therefore the detection of a deterministic trend in a long-range correlated series is even more challenging than in a short-range correlated series (for a definition of these terms cf. Sec. 2.1). As noted by Diebold and Inoue (2001), in certain circumstances deterministic trends and long memory (long-term correlations) are 'effectively different labels for the same phenomenon, in which case attempts to label one as "true" and the other as "spurious" are of dubious value'. An example for long-term correlated data exhibiting spurious trends is given in Fig. 3.1. One could easily be tempted to assume a deterministic trend component in the data series of the right hand side. But this series is a realisation of a stationary process, i.e. has mean zero – though with heavy long-term correlations. The simulated series are quite long with 10 000 data points. Thus, spurious trends are not only a feature of short time series.

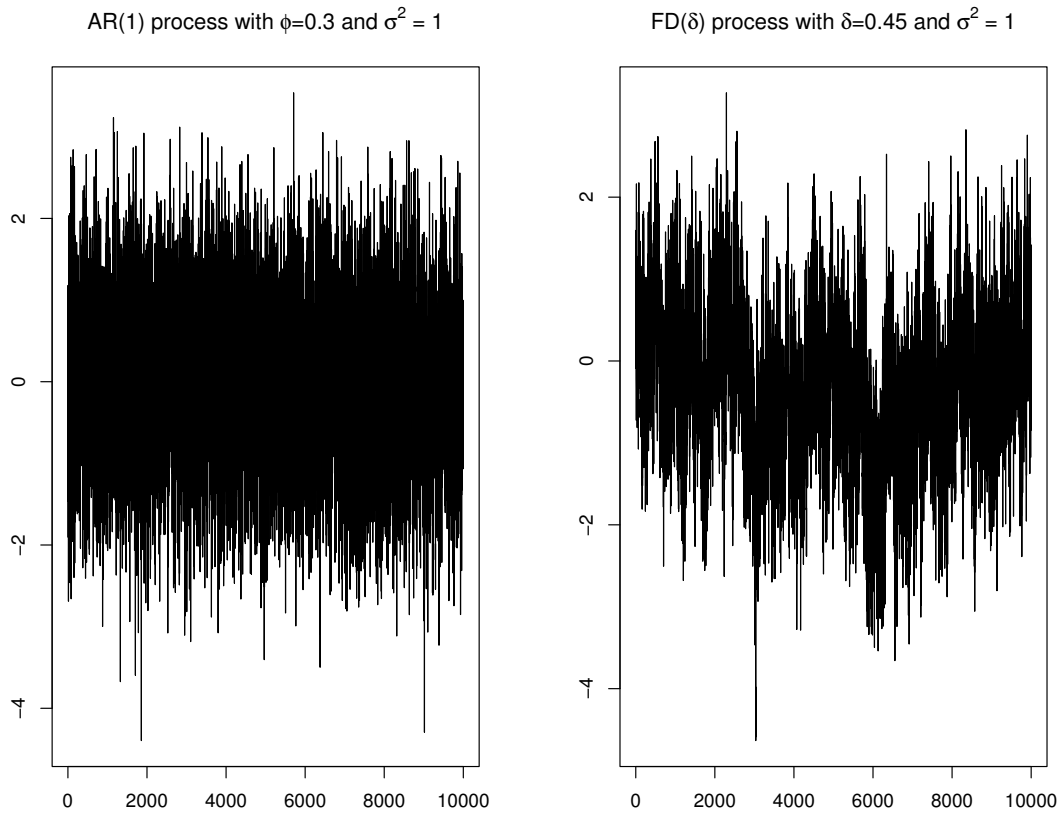


Figure 3.1: Comparison of short- and long-range correlation. Realisations of a short-range correlated AR(1) process (left hand side) with parameter $\phi_1 = 0.3$ and a long-range correlated FD(δ) process (right hand side) with parameter $\delta = 0.45$. Both stochastic processes are stationary. However, in the realisation of the long-range correlated process one might suspect a trend.

Furthermore, the distinction between deterministic trend and short memory on the one hand and long memory on the other, is still a matter of research. Giraitis et al. (2001) clarify the difficulties of R/S-type test statistics[1] to discriminate between long memory and short memory combined with specific forms of non-stationarity. The estimation of the long-range parameter using heuristic methods may also be biased by short-term correlations. Rust (2007) discusses various methods to estimate the Hurst coefficient and to detect long-range correlations (see also 2.1.3) and mentions potential problems.

Several authors report long-term memory being present in river discharge records (see, e.g., Hurst 1951, Lawrence and Kottegoda 1977, Montanari et al. 1997, Beran 1994, Ooms and Franses 2001 and Mudelsee 2007). Therefore we assess trends by modelling the correlation structure of the data with a stochastic fractional ARIMA process, which is able to reproduce short-term as well as long-term correlated data (see Sec. 2.1.3 for further details). The FARIMA models as well as short-range correlated ARIMA processes are fitted to the data and the best model is chosen with the HIC selection criterion, which

[1] Rescaled adjusted range statistic which is utilised by Hurst (1951) to identify the ideal storage capacity of the Nile River reservoirs. He identifies also a power law, so the R/S statistic can be used to estimate the Hurst coefficient.

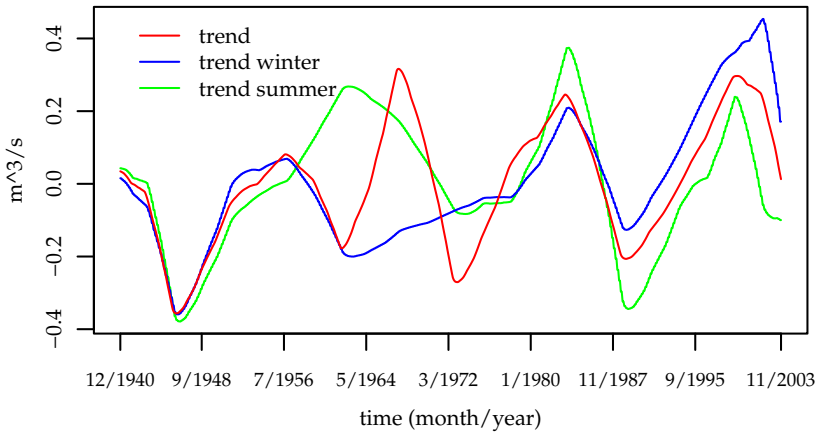


Figure 3.2: Trend shape of the Dreisam River at Ebnet. Trend estimates for river discharge anomalies (monthly means) of the Dreisam River at Ebnet (red) and for the winter season (blue) and the summer season (green).

relies on the likelihood (cf. Sec. 2.5.1).

Long-range dependence and heavy tails

Long-range dependence is often mentioned in conjunction with heavy tailed distributions. At first sight there is no connection between both phenomena. Heavy tailed distributions, such as the Pareto distribution, exhibit an asymptotic hyperbolic shape - regardless of the distribution for small values of the random variable - i.e. $P(X > x) \sim x^{-a}$ for $x \to \infty$, $0 < a < 2$. These distributions are also known as power-law distributions. Even one single observation in time can be described as a realisation of a random variable with a specific distribution. Therefore the definition of a distribution does not rely on time, but is somehow static. Long-range dependence on the other hand (as introduced in Sec. 2.1.3), always deals with the connection between several observations (or random variables) in time, namely the auto-correlation structure. Both phenomena need not occur together. They can, for example, in case several time dependencies interfere with each other. Then the resulting data may exhibit heavy tails (cf. Katz et al. 2005). It is moreover possible to connect long-range dependence with the theory of point processes, which are used to model the occurrence and magnitude of extreme values (Daley and Vesilo 1997).

3.2.3 *Trend shape and trend detection methods*

Apart from the fundamental question whether a trend is present in the data or not, the trend shape is very important. In the first place, the interpretation of this shape is useful for drawing conclusions about the behaviour of the analysed system. Thereby the direction of the trend, i.e. increasing or decreasing, is one possible outcome. Other research questions, such as the retrieval of periodicities or other large scale patterns, are conceivable. As an example in Fig. 3.2 the estimated trend of discharge anomalies for the Dreisam River at Ebnet is depicted (red). Furthermore trend estimates for the winter season (blue) and summer season (green) are shown. The river discharge trends exhibit decadal periodicity.

However, we model a system which is expected to be composed of the deterministic trend and a stochastic part representing natural variability. Therefore we regard the

analysed time series just as one realisation of a stochastic system and, as pointed out in Sec. 3.3.4, the variability of this system and the reliability of the trend estimate heavily depend on the inherent correlation structure of the system. Hence the inclusion of confidence intervals is very important for the interpretation of the trend shape and the method chosen for trend estimation should provide parameters which are applicable for uncertainty assessment and suffice mathematical standards of estimation (i.e. consistency and efficiency conditions). Last but not least, the trend estimate determines the part of the time series which is left to be interpreted as stochastic variability. Thus the result of a test for a trend heavily depends on the shape of the trend estimate. This holds for the framework presented here, as well as for standard regression analysis, e.g. choosing a straight line as trend estimate (see Beran 1994 for further details).

One way to estimate trends is to proceed in a fully parametric way. This means that a certain trend shape, e.g. polynomial or exponential, has to be assumed. These subjective assumptions are at best driven by expert knowledge about the processes of the observed system. In case of lack of information, the assumption of rather flexible trend shapes, such as polynomials, is practical. Since in this methodology the trend shape is known, the trend can easily be extrapolated into the future. Fully parametric approaches therefore are very helpful in forecasting. Nevertheless, the class of assumed trend forms has an influence on the trend test, which only tests for the existence of these particular trends. Hence the trend assumptions have to be reliable. Fully parametric approaches offer a setting where the trend and the stochastic model are estimated simultaneously (see, e.g., Smith 1993). In this way the mutual dependence between both characteristics, the trend and the complexity of the stochastic part, is fully considered. On the other hand, here the parameter estimates for both phenomena reflect this mutual dependence. Furthermore, the stochastic model and the trend estimate as a whole must not get too complex. Otherwise models with a lot of parameters are used and thereby the usual optimisation problems arise. Therefore rather simple trend assumptions, such as straight lines, are usually made when using a fully parametric approach. For demonstrations of this methodology see, e.g., Bloomfield (1992) and Woodward and Gray (1995).

In a semi-parametric framework the trend estimation is data driven, and a priori knowledge about the type of trend (e.g. linear, exponential, etc.) is not necessary. Therefore, this approach is especially useful for assessing time series like river run-off, which are expected to have complicated and unpredictable trend forms. Using this technique, at first the trend is usually estimated with a non-parametric method, i.e. via wavelets, kernels, splines or by simply smoothing the time series. Then, in a second step, the parametric stochastic part, e.g., a FARIMA model, is estimated. Since this procedure involves several steps, the appropriateness of the trend estimate will always be questionable, although estimators based on semi-parametric approaches show properties like consistency or efficiency. For applications in the hydro-meteorological context see, for example, Sibbertsen (1999) and Ramesh and Davison (2002).

We chose to estimate the trends using wavelets in a semi-parametric approach. The methodology is based on Craigmile et al. (2004). We extended it by using a broader class of stochastic models (FARIMA), which enables us to represent the correlation structure of empirical data more flexibly. For specific details see Kallache et al. (2005).

3.3 TREND ESTIMATE

We assume that the observed time series can be modelled additively by a deterministic trend component $\{T_t\}$ plus a realisation of a stochastic process $\{X_t\}$: $Y_t = T_t + X_t$ for $t = 0, \ldots, n-1$. By proceeding in this way, auto-correlations in the time series are modelled in X_t. In the proposed method, the influence of the variability of the system, that is the structure of X_t, on the uncertainty of the trend estimate is taken into account.

To estimate the trend we chose a Daubechies least asymmetric wavelet of width 8. As outlined in Sec. 2.2.2, the LA(8) wavelet possesses several advantages which are applicable in trend assessment. Furthermore, the wavelet of width 8 exhibits good reconstruction properties for a variety of time series tested, therefore we use it for all our discharge data. The trend is estimated using the DWT (see Sec. 2.2.1). By doing so, the data vector $\boldsymbol{Y}$ is decomposed into a component $\widehat{\boldsymbol{T}}$, representing the variability on large scales, and a component $\widehat{\boldsymbol{X}}$ for small scales:

$$\boldsymbol{Y} = \boldsymbol{\mathcal{W}}^T\mathbf{W} = \boldsymbol{\mathcal{W}}^T\mathbf{A}\mathbf{W} + \boldsymbol{\mathcal{W}}^T(\mathbf{I}_n - \mathbf{A})\mathbf{W} = \widehat{\boldsymbol{T}} + \widehat{\boldsymbol{X}} \quad . \tag{3.2}$$

$\mathbf{I}_n$ is the $n \times n$ identity matrix, $\boldsymbol{\mathcal{W}}$ are the wavelet coefficients and $\mathbf{W}$ is the matrix of the wavelet transform. $\mathbf{A}$ contains all non-boundary wavelet coefficients, whereas $(\mathbf{I}_n - \mathbf{A})$ contains all scaling and boundary coefficients. Daubechies wavelet filter of length L eliminate a polynomial trend of order K in the non-boundary wavelet coefficients if $K \leq L/2 - 1$. Since a reconstruction $\boldsymbol{Y} = \boldsymbol{\mathcal{W}}^T\mathbf{W}$ gives back the original time series, the trend $\{T_t\}$ is captured by the boundary wavelet coefficients and the scaling coefficients. The inverse wavelet transform of these coefficients $\widehat{\boldsymbol{T}}$ can be used as an estimate for the trend. $\widehat{\boldsymbol{T}}$ not only contains the deterministic trend component but also the stochastic variability on large scales and in the following is referred to as the *trend estimate*.

3.3.1 Separating scale

As presented in Sec. 3.3, the time series is decomposed in its fluctuations according to different scales to estimate the trend. So one has to define this separating scale between "trend" and fluctuations on small scales. Most reliable results are obtained by selecting the separating scale l_s between $\widehat{\boldsymbol{T}}$ and $\widehat{\boldsymbol{X}}$ as the largest scale with a non-zero number of wavelet coefficients not affected by boundary conditions (Craigmile et al. 2004). Thereby the choice of the separating scale depends on the number of analysed data points n. Furthermore, the separating scale should be chosen such that the definition of trend as fluctuations on large scales is met. As pointed out in Sec. 2.2.2 applying a Daubechies scaling filter is a kind of weighted smoothing procedure. The choice of l_s determines the amount of variation of the time series being present in the trend estimate. Considering both aspects and using the $LA(8)$ wavelet filter a separating scale of $l_s = 6$ is found to meet our requirements best. All investigations presented in the following are carried out with this l_s. It corresponds to a separating scale interval of 32 to 64 months when dealing with monthly measurements. Thus the estimated trends reflect variations of the time series which are larger than 5 years. The choice of the separating scale is important for the trend test. It may be compared to the selection of a window size when smoothing data to get a trend estimate. Percival and Walden (2000) show that shifting the separating scale only one unit might change the trend test results.

3.3.2 Boundary conditions

Wavelet coefficients are always a certain aggregation of the variations of the time series around a time point t_0. The larger the analysed scale, the more points get aggregated, i.e. the more wavelet coefficients are affected by the boundary conditions. To be able to calculate these wavelet coefficients some kind of assumption for the elongation of the time series at both edges has to be made. One alternative is to assume periodic continuation, i.e. the end of the time series is identified with the beginning. Other approaches are to expect reflecting boundary conditions, where the time series is reflected at both edges, or to pad the series with the mean value at the edges (see Torrence and Compo 1998 for more details). In the following we use the latter alternative. As one goes to larger scales, padding decreases the amplitude of the wavelet and scaling coefficients as more zero values enter the analysis. The trend estimation and the calculation of its sigma intervals are done on this elongated padded series, whereas the test for trend is applied to the original empirical series and its fluctuation estimate without padded parts. When using wavelet filters with width $L > 2$ for the transform, the boundary coefficients cause a higher variability of the trend estimate at the endpoints of the time series and accordingly wider confidence intervals there.

3.3.3 Stochastic component

Having estimated the trend a stochastic model is applied which represents $\widehat{X}$ best. The class of FARIMA(p,δ,q) models (see Sec. 2.1 for further details) is regarded as being able to reproduce the auto-correlation structure of the data reasonably well. Thereby iid and normal distributed noise is assumed. Therefore the distribution of the analysed data should be approximately normal. Deviations from this assumption result in slightly biased parameter estimates for the FARIMA model. On the other hand, methods which explicitly consider non-normal distributed data are not robust against mis-specification of this distribution (Giraitis and Robinson 2001). To estimate the parameters we use the Whittle estimator as a fast and robust approximation of maximum likelihood estimators (see Sec. 2.4.1). It shows the consistency and efficiency properties described in Sec. 2.4.3.

The maximum likelihood parameter estimation procedure of the FARIMA(p,δ,q) processes assumes stationary data, but we use the methodology to estimate parameters of a non-stationary series. Therefore we investigated the bias of the parameter estimation induced by non-stationarity. Fig. 3.3 shows the results of a Monte Carlo simulation to assess the bias when estimating the $\delta = 0.3$ parameter due to a trend. We use simulations from a FARIMA(p,δ,q) process of length 1 000 and with $p, q = 1$ for this study. Trends with various shapes and strength (0 to 3 times the standard deviation of the process) are added to the data. The strength of the added trend component is indicated at the x-axis, whereas the type of the added trend is visualised with different colours, i.e. linear (black), parabolic (red) and quadratic (green). The parameters are estimated on the original data series (solid lines) and on the filtered series, where a potential trend is estimated using wavelets (as described in Sec. 3.3) and is removed from the series. The bias is visualised as deviance from the original value for δ, e.g. 0.3. As expected, the bias of the estimate of δ due to a trend increases with the trend magnitude when fitting the parameters on the original series. In case the parameter estimates for the FARIMA process are obtained from the filtered series, we get a reverse picture: The bias is about as large as the bias introduced through trend, but decreases with augmenting trend magnitude. Similar results

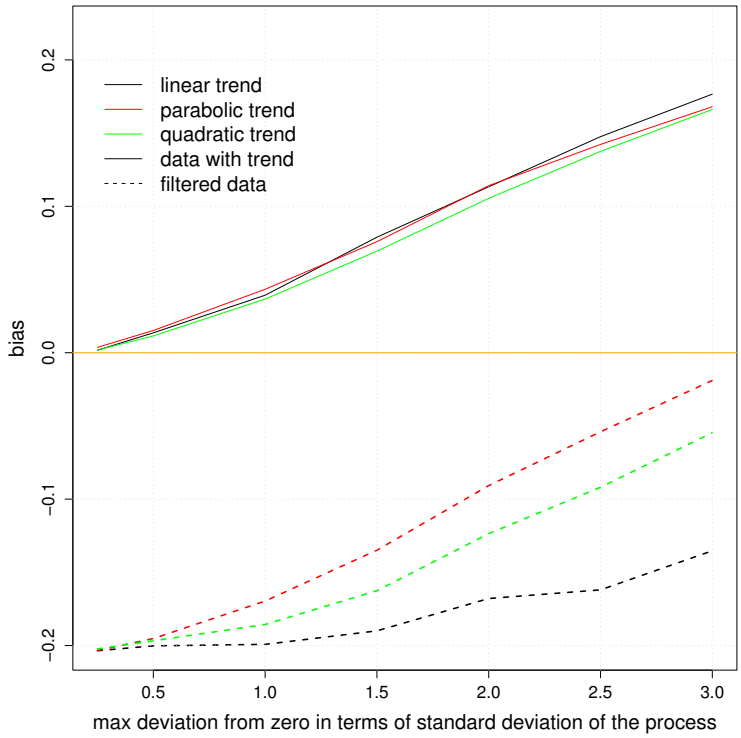


Figure 3.3: Parameter estimation under trend or filtering. Assessment of the bias of the estimate for the fractional difference parameter $\delta = 0.3$ of a FARIMA(p,δ,q) model with $p, q = 1$ due to a trend or to filtering. Various trend shapes, i.e. linear (black), parabolic (red) and quadratic (green) are added to simulated FARIMA series. The strength of the trend is augmented along the x-axis (0 to 3 times the standard deviation of the process). In case the fit is performed on the original data, the bias is depicted with solid lines, while dashed lines indicate a fit on the filtered series where a trend estimate (which is obtained using wavelets) is removed.

are obtained for the AR parameter $\phi_1 = 0.3$. The MA parameter $\psi = 0.3$ is not affected. Thus, the bias of the Whittle approximation to maximum likelihood estimation due to a possible trend cannot be reduced by wavelet filtering. In this work we therefore use the maximum likelihood estimates without any correction term. We consider this a reliable approach, because trends in empirical time series are normally much weaker than e.g. 3 times the standard deviation of the process itself. Therefore the bias we might get for the parameter estimates in case a trend is present in the data is small compared to the confidence intervals for the estimates.

A variety of models out of the FARIMA class is fitted to $\widehat{X}$ and the best one is selected via some information criterion (see Sec. 2.5.1). In the following we always use the HIC as selection criterion to select between different FARIMA models. A goodness-of-fit test (see Sec. 2.6.1) is applied to ensure the adequacy of the best model, afterwards the trend test (Sec. 3.4) is carried out.

Analysing empirical data, we find that the model chosen as best one does not have an influence on the shape of the trend, but it has an impact on the variability of the trend estimate and therefore on the test for significance of the trend: Standard errors for the trend parameters obtained under an assumption of long-range dependence may be considerably larger than those obtained under a short-range correlated autoregressive model (see Sec. 3.3.4 and 3.4 for further details). This result is also obtained by various other authors, e.g. Smith (1993), Sibbertsen (1999) and Bloomfield (1992). Beran (1994) infers furthermore that the broadened confidence interval for estimated trend parameters can lead to an acceptance of the hypothesis that there is no trend in case of long-term correlations, whereas this hypothesis is rejected in case of short-term correlations. Cohn and Lins (2005) compare several trend tests applied to the annual northern hemisphere mean temperature anomalies. They consider different types of auto-correlation,

and find also that the stochastic model chosen has an important influence on the trend test.

3.3.4 *Variance of the trend estimate*

The connection described in the previous section, between estimated trend $\widehat{\boldsymbol{T}}$ and assumed stochastic model $\boldsymbol{X}$, is made explicit in our approach by reorganising Eq. (3.2), namely

$$\widehat{\boldsymbol{T}} = \boldsymbol{\mathcal{W}}^T \mathbf{A} \boldsymbol{\mathcal{W}} \boldsymbol{Y} = \mathbf{R}\boldsymbol{T} + \mathbf{R}\boldsymbol{X} \quad , \tag{3.3}$$

with $\mathbf{R} = \boldsymbol{\mathcal{W}}^T \mathbf{A} \boldsymbol{\mathcal{W}}$.
Therefore $\mathrm{cov}(\widehat{\boldsymbol{T}}) = \mathbf{R}\mathrm{cov}(\boldsymbol{X})\mathbf{R}^T$ holds, i.e. the variance of the trend estimate involves the covariance of $\boldsymbol{X}$. So if $\boldsymbol{X}$ is a realisation of a stationary stochastic process, the auto-covariance sequence $\rho_{\boldsymbol{\Theta}}(k)$ of $\boldsymbol{X}$ can be used to calculate the variance of the trend

$$\mathrm{VAR}(\widehat{T}_t) = \sum_{k=-(n-1)}^{n-1} \rho_{\boldsymbol{\Theta}}(k) r_{t,k} \quad , \tag{3.4}$$

with $r_{t,k} = \sum_{i=0}^{n-1-|k|} R_{t,i} R_{t,i+|k|}$ and n being the length of the time series. $R_{t,i}$ denotes the (t,i) element of $\mathbf{R}$ (for further details see Craigmile et al. 2004). Details concerning the structure of the auto-covariance function $\rho(k)$ of a stationary FARIMA process are given in App. A-1. The variance of the trend estimate depends on several factors: The wavelet transform used, the chosen separating scale l_s, the number and magnitude of the parameters of the stochastic process which complements the trend, and finally the sample size n.

In our analysis the Daubechies LA(8) wavelet transform is chosen for all analyses (see Sec. 2.2). This means the width of the wavelet L is larger than 2 and so the variance of the trend estimate changes with time. The overall variability oscillates with time, dependent on the shape of the scaling filter. Higher variability at the endpoints of the time series is also induced due to the influence of the boundary wavelet coefficients.

Regarding the influence of the stochastic model, Eq. (3.4) implies that the variance of the trend estimate is proportional to the variance of the stochastic process σ^2. Furthermore, the choice of the stochastic model as well as its estimated parameters have an impact on the width of the confidence interval of the trend estimate. This is in line with findings on confidence intervals of estimates for other important statistic quantities. Koutsoyiannis (2003) and Rust et al. (2007) find confidence intervals of quantile estimates getting larger in case long-range correlation is present in the data and Sibbertsen (1999) reports the like for regression estimates. For an example see Fig. 3.4. Here fluctuations in the confidence interval due to the choice of the wavelet transform are averaged out and the confidence intervals are calculated for the padded series. The range of influence of the boundary conditions is not depicted, but it can be derived analytically (Percival and Walden 2000). Examining the character of the influence of the chosen model we find that under assumption of long-range dependence, the confidence intervals obtained for the trend estimate can be considerably larger than those obtained under a short-range correlated model. This important result has occurred repeatedly during our analysis of hydro-meteorological data.

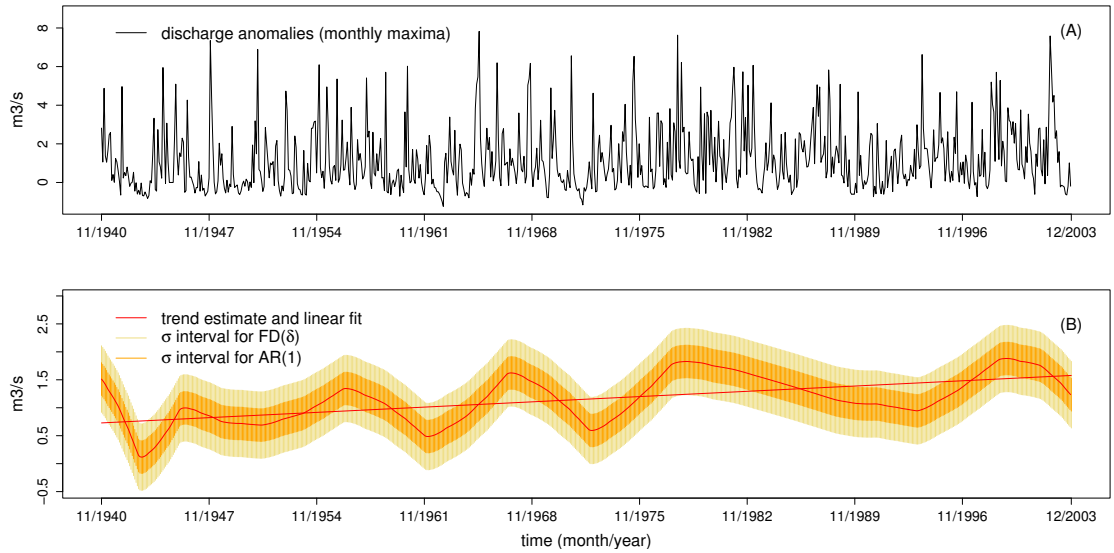


Figure 3.4: Wavelet trend estimate and confidence intervals for different stochastic models. (A) River discharge anomalies (monthly maxima) for the Erms River at Riederich and **(B)** trend estimate with one-sigma intervals for a FD(δ) model ($\delta = 0.252$, $\sigma^2 = 2.465$) and an AR(1) model ($\phi_1 = 0.328$, $\sigma^2 = 2.443$). Additionally, a straight-line fit is shown.

The long-range dependence parameter δ determines the decay of the auto-correlation function for large lags (Ooms and Doornik 1999). We therefore can conclude that for long stationary time series δ dominates the magnitude of the sum in Eq. (3.4), i.e. the sum of the auto-correlation function and therefore the size of the standard deviation of the trend estimate. Unfortunately, for short series no general rule can be provided to conduct from the FARIMA model type to the size of the confidence intervals for the trend estimate. This relation depends on the existence and magnitude of the parameter δ, the number and magnitude of the AR and MA parameters and on the sample size n.

3.3.5 *Consistency and efficiency*

The bias in estimating $\boldsymbol{T}$ with $\widehat{\boldsymbol{T}}$ is zero, in case a Daubechies wavelet filter of adequate length L is used such that $K \leq L/2 - 1$ holds. Here K denotes the order of the polynomial trend. This means

$$E(\widehat{\boldsymbol{T}} - \boldsymbol{T}) = \mathbf{0}_n \tag{3.5}$$

with $\mathbf{0}_n$ being a vector of n zeros (cf. Craigmile et al. 2004 for further details).

Furthermore, the trend estimator presented here converges to a Gaussian random variable (Craigmile et al. 2004), i.e.

$$\lim_{n\to\infty}(\widehat{\boldsymbol{T}} - \boldsymbol{T}) \to \varepsilon \quad ; \quad \varepsilon \sim N(\mathbf{0}, \mathbf{R}\mathrm{cov}(\boldsymbol{X})\mathbf{R}^T) \quad . \tag{3.6}$$

This means it is an asymptotically normal and consistent estimator (even in combination with a FARIMA model). Further details are given in Sec. 2.4.3.

Whereas the property of consistency is concerned with the asymptotic accuracy of an estimator efficiency assures the existence of an upper bound for the asymptotic variance of an estimator (cf. Sec. 2.4.3). This cannot be proved universally for $\widehat{\boldsymbol{T}}$, $\widehat{\boldsymbol{T}}$ is not an asymptotically efficient estimator.

There exist consistent and asymptotically efficient estimators for trends under consideration of long-term correlation. Sibbertsen (1999), for example, considers a regression model with error terms having long-term correlations. He shows that the LSE (least squares estimator) and BLUE (best linear unbiased estimator) are consistent and asymptotically efficient estimators. The only difference to models with independent error terms or error terms with short-term correlations is the rate of convergence. It depends on δ and the estimators converge more slowly. However, in this parametric regression model the trend shape is fixed. Sibbertsen (1999) discusses also a semi-parametric regression model with a twice differentiable continuous function as non-parametric part (presented by Beran and Ghosh (1998)). This estimator is unbiased, $\sqrt{n}$-consistent, and also not efficient.

3.3.6 *Pointwise confidence intervals*

By using the results sketched in the previous section standard $100(1-\alpha)\%$ pointwise confidence intervals are obtained for the trend estimate:

$$\widehat{T}_t \pm q_{1-\frac{\alpha}{2}} \sqrt{\mathrm{VAR}(\widehat{T}_t)} \qquad t = 0, \ldots, n-1 \quad , \tag{3.7}$$

where q_α is the α-quantile of the standard normal distribution. These pointwise confidence intervals are used in all following analyses.

Estimation of the parameters of the stochastic process and related uncertainty

In Sec. 3.3.5 the properties of the trend estimator are determined under assumption of known parameters of the stochastic process $\boldsymbol{X}$. In practice, we estimate the FARIMA process parameters and do not know the true value of $\mathrm{VAR}(\widehat{T}_t)$ for each t in $1, \ldots, n$.

Craigmile et al. (2004) investigate the properties of $\widehat{\mathrm{VAR}}(\widehat{T}_t)$ and the coverage probabilities for the confidence intervals for $\widehat{\boldsymbol{T}}$, which have been obtained using Eq. (3.7), i.e. not considering the difference between $\mathrm{VAR}(\widehat{T}_t)$ and $\widehat{\mathrm{VAR}}(\widehat{T}_t)$. This analysis is done for a $\mathrm{FD}(\delta)$ process with fixed δ. They find that conservative intervals are obtained for the white noise case ($\delta = 0$), whereas too narrow intervals are derived for positive δ.

3.4 TEST FOR SIGNIFICANCE OF THE TREND

A very common procedure to test for a trend is to fit a straight line to the data and test for the compatibility of the slope parameter with zero. This method is well extendable to the consideration of auto-correlations, but it does not allow for more complicated trend forms. Therefore it exhibits the drawbacks described in Sec. 3.2.3.

When allowing for more complicated trend shapes other test statistics have to be provided. For fully parametric approaches the deviance statistic may be used to test for trends (cf. Davison 2003 and Cohn and Lins 2005), see Sec. 2.5.2. Natural stochastic variability of the time series might also be a good measure to construct a trend test statistic. The comparison of variabilities, as, for example, performed with the F-test, has a

long tradition in statistics. Rybski et al. (2006) for instance analyse reconstructions of the northern hemisphere temperature anomalies and test whether a smoothed version of the time series, taken as trend, is compatible with the natural variability of a fitted stochastic FD(δ) model. They compare the process variability with the estimated trend separately for each time point, but without considering the multiple testing problem.

The test for trend used in this thesis compares the variability in the whole time series $\boldsymbol{Y}$ with the fluctuations on small scales $\widehat{\boldsymbol{X}}$. By doing so, the test statistic is composed out of aggregates of the data over time and no point by point comparison is necessary. In this way, the consideration of auto-correlations is possible in a sound manner. In the following we refer to this trend test as *Craigmile-Guttorp-Percival trend test*.

As outlined in the previous sections, it is assumed that the time series is composed by a deterministic and a stochastic (possibly auto-correlated) part. The test now is based on a hypothesis testing procedure: Two complementary hypotheses are defined to make a statement about the process $\boldsymbol{Y}$, that is $\boldsymbol{Y}$ is solely stochastic or it is composed out of a stochastic and a deterministic component. Without loss of generality, we assume $\boldsymbol{Y}$ has a zero mean. The null hypothesis H_0 is now: $T_t = 0 \; \forall t = 0, ..., n-1$ versus H_1: not H_0.

The goal of the hypothesis test is to decide, based on a realisation of the process, which of the two complementary hypotheses is true. For this purpose a rule is defined that specifies for which sample values the decision is made to accept H_0 as true and for which sample values H_0 is rejected and H_1 is accepted as true. More precisely the test statistic p_c is defined as

$$p_c = \frac{\|\boldsymbol{Y}\|^2}{\|\widehat{\boldsymbol{X}}\|^2} \quad , \tag{3.8}$$

where $\|.\|$ denotes the Euclidean norm. For $\{T_t\} \neq \{0\}$, p_c should be large.

To be able to define regions of rejection and acceptance of H_0 the distribution of the test statistic under assumption that H_0 is true, $F(p_c|H_0)$, has to be provided. In the special case that $\widehat{\boldsymbol{X}}$ is given by an FD(δ) process with restricted domain $\delta \in [0, 0.25]$ this distribution can be derived theoretically (see Craigmile et al. 2004). However, to obtain test results for the more general setting presented here we generate this distribution via Monte Carlo simulation. By this means, R realisations (here always $R = 4\,000$) of the best fitting stochastic model with the fitted parameters (but without trend component) are generated. Then the test statistic p_c^r is calculated for each realisation $r = 1, \ldots, R$, which results in an estimate for the distribution of p_c under the assumption that H_0 is true. Now a level of significance α has to be defined. H_0 is rejected at this level in case the test statistic derived out of the empirical time series $\boldsymbol{Y}$, $p_{c(\text{empirical})}$, exceeds the upper 100α%-quantile of the simulated p_c^r values.

An inherent assumption in this comparison of variances is that a possible trend significantly alters the ratio between the variance on small and large scales. By expecting p_c to be larger with trend than without trend, we assume that a possible trend adds to the stochastic variability and that therefore the overall variability on large scales becomes higher. The cases in which a deterministic trend in the data happens to be contrary to the inherent stochastic variability and therefore reduces the variability of $\boldsymbol{Y}$ are not considered here.

A known problem of hypothesis testing is that the distribution of p_c under assumption of H_1 has not to be taken into account to decide whether to accept or to reject H_0, at

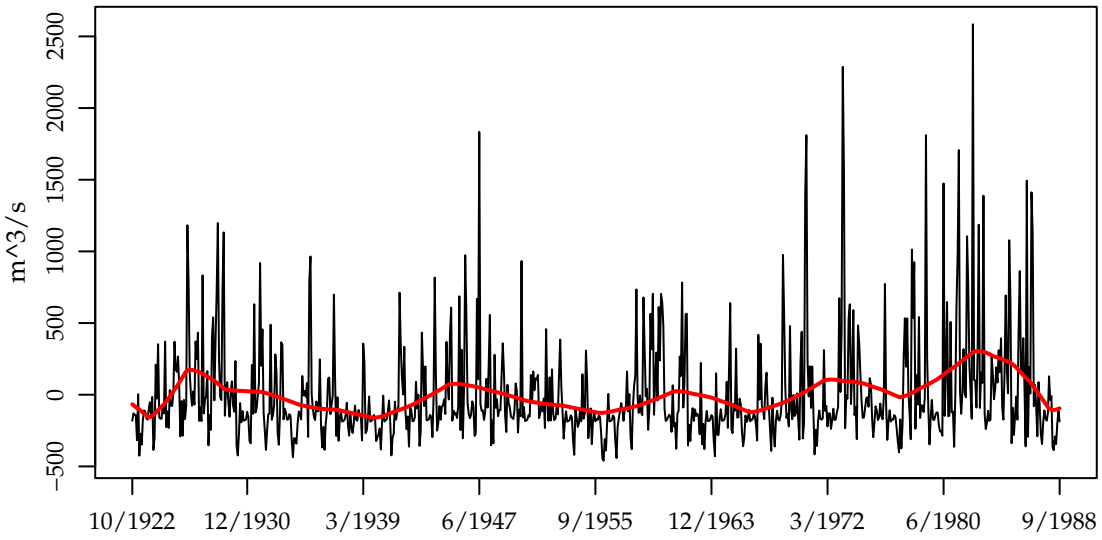


Figure 3.5: Wavelet trend estimate for the Fox River at Wayland. The anomalies of monthly mean discharge of the Fox River at Wayland is depicted for the years 1922 to 1988 (black). The annual cycle is removed. The trend estimated obtained by wavelet filtering is shown in red.

least for the hypothesis testing procedure described so far. The level α quantifies the Type I error, that is falsely rejecting H_0. But no estimation about the magnitude of the Type II error, i.e. falsely accepting H_0, is made. To evaluate this error and the capacity of the test to reject H_0, the power of the test is investigated in the following section.

3.4.1 Power of the trend test

Consider a hypothesis test with $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^c$, the complement. Suppose $\mathcal{R}$ denotes the rejection region for this test, i.e. for $\mathbf{x} \in \mathcal{R}$ the null hypothesis is rejected. Now the *power function* of a hypothesis test is given by $P_\theta(\mathbf{X} \in \mathcal{R})$. This is an adequate evaluation tool, because

$$P_\theta(\mathbf{X} \in \mathcal{R}) = \begin{cases} \text{probability of a Type I error} & \text{if } \theta \in \Theta_0 \\ \text{one minus the probability of a Type II error} & \text{if } \theta \in \Theta_0^c \end{cases}, \quad . \tag{3.9}$$

The ideal power function is 0 for all $\theta \in \Theta_0$ and 1 for all $\theta \in \Theta_0^c$. Except in trivial situations, this ideal power function cannot be attained. A small Type II error is a measure of how well a test is able to discriminate between H_0 and H_1. In our setting the probability distribution of the test statistic p_c under assumption of H_1, that is we have p_c calculated on a series with an arbitrary trend form, cannot be provided. There exist infinitely many trend shapes and we would have to consider them all to be able to calculate the power function. However, we have checked the ability of our trend test to reject H_0 for artificial data with certain types of trend added. In this way we get an impression about the capability of the trend test to discriminate between series with and without trend.

To determine the power of the trend test, a Monte Carlo study is performed on realisations of an AR(1) and a FD(δ) model with diverse parameter choices and deterministic

trends superimposed. This means the power is only tested for these special cases. The amplitude of the deterministic trend is increased in small steps from 0 to 0.4 times of the standard deviation of the process. Analysing linear trends, where the trend starts at $t = 0$ and the endpoint $t = n - 1$ is varied from 0 to 0.4σ, the performance of the trend test is quite good. It is slightly lower than the power of a standard linear regression test (cf. Craigmile et al. 2004). Furthermore, we want to analyse the performance of the test in an empirical setting. Therefore we added a trend component to the artificial time series, which has been derived from the Fox River at Wayland. The river is situated in North America and has been chosen for illustrational purposes. The trend component is depicted in red in Fig. 3.5.

The resulting power curves are depicted in Fig. 3.6. The major results are as follows:

(a) In case the artificial data exhibits a trend with a shape which as well might occur in nature, namely the trend estimated from the Fox River at Wayland, the power curves are fortunately nearly as good as for a linear trend, as depicted in Fig. 3.6 (A).
(b) The power of the trend test does not weaken significantly when a linear trend starts in a later part of the time series. This was tested for a starting point as late as $t = (5/6)n$ (see Fig. 3.6 (B)). The setting is chosen to examine whether the test is applicable in cases where about 200 years of measurements are available, but a possible trend in the values just starts within the last 30 years. This is useful for, e.g., meteorological time series, where an increase due to anthropogenic interference is expected to manifest only within the latter part of the series.
(c) The trend test is vulnerable against jumps in the data, which can be seen in Fig. 3.6 (C). This was tested by including a jump in the middle of the series with amplitude 0 to 0.4σ. Thereby a jump is not regarded as peak but rather as step function, that is a shift of the whole latter part of the time series. It shows that the occurrence of jump points should be excluded before the trend test analysis. To deal with this problem Craigmile et al. (2004) proposed an extension to the methodology: By adopting the wavelet coefficients to the jump points, the trend test can still be applied. However, the jump points have to be identified before this procedure can be conducted.
(d) As shown in Fig. 3.6 (D), the trend test is robust against changes in variance of the time series, which cancels out in Eq. (3.8).
(e) The power does not only depend on the choice of the stochastic model, but also on the magnitude of the parameters. For an FD(δ) model, i.e. it reaches a power of one faster for small δ, which can be seen in all figures of Fig. 3.6.

3.5 APPLICATION, RESULTS AND DISCUSSION

3.5.1 River Neckar basin

For this analysis daily river discharge records from several catchments near the Neckar River in Southern Germany are investigated. The 10 series were selected according to their length and completeness. The run-off data are examined for the jointly covered time period from November 1940 to December 2003. In all datasets the annual cycle in mean and variance is removed (see Sec. 2.8.1). Finally, the daily measurements are

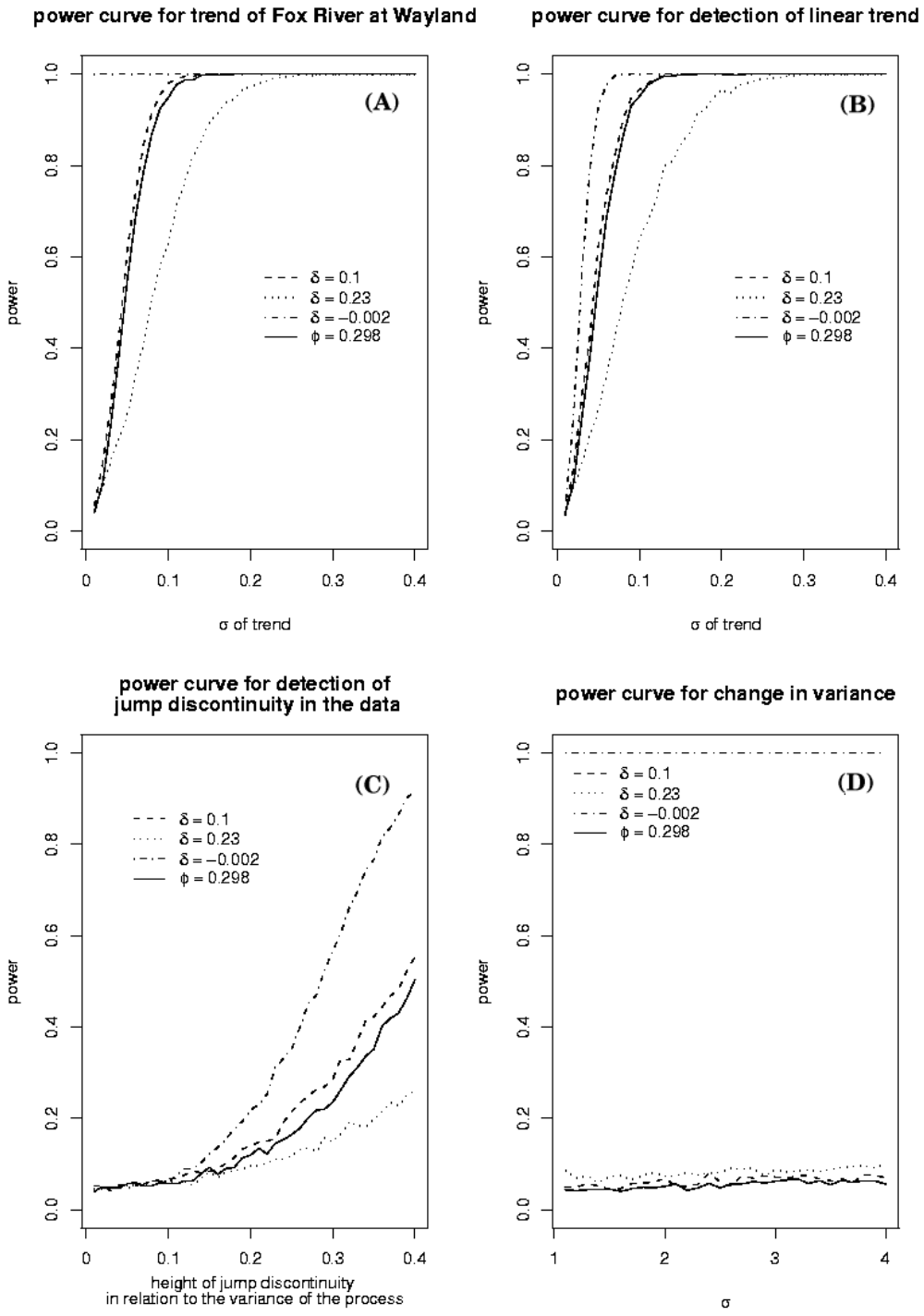


Figure 3.6: Power of the trend test. Power curves for wavelet based test for trend under model assumption of AR(1) with $\phi_1 = 0.298$ and FD(δ) with $\delta = -0.002$, $\delta = 0.1$ and $\delta = 0.23$. **(A)** A trend component which has been estimated from the Fox River at Wayland is added. The variance of this trend is increased such that its standard deviation reaches from 0σ (no trend) up to 0.4σ of the stochastic series. **(B)** A linear trend is added, starting in the last sixth of the series. **(C)** A jump discontinuity is incorporated in the middle of the data. **(D)** The second part of the time series has an increased variance, ranging from σ_1 to $4\sigma_1$, thereby σ_1 is the variance of the first part. The trend test is robust against this change and does not falsely detect a trend.

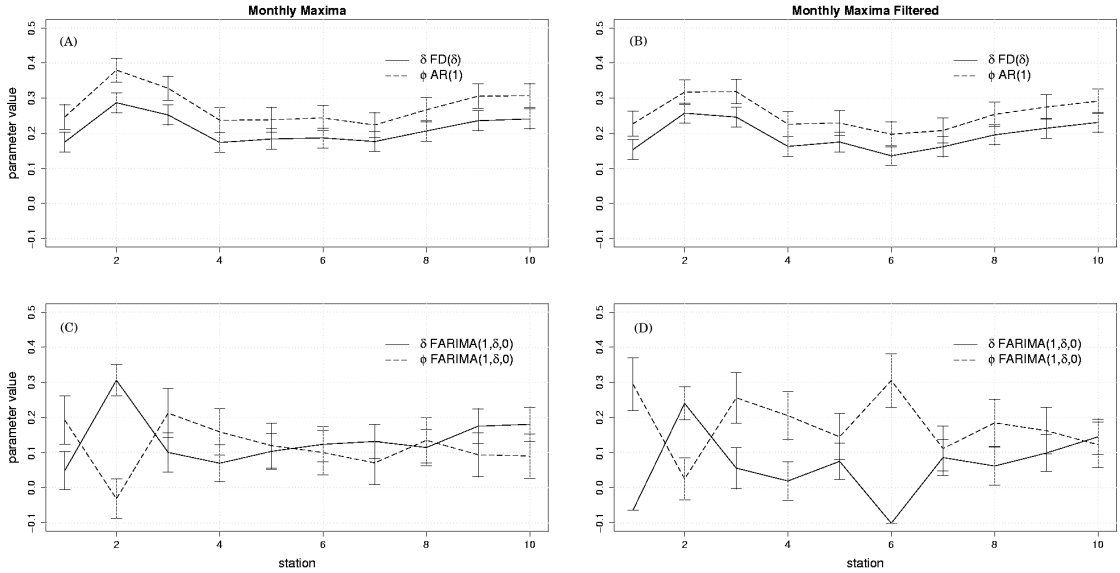


Figure 3.7: Influence of trend elimination. (A) Fitted parameters for the FD(δ) and the AR(1) model and **(C)** the FARIMA(1,δ,0) model for the monthly maximum data and the monthly maximum filtered data, respectively (**(B)** and (**(D)**). River Gauges: 1 Josbach/Hölzlebruck, 2 Ammer/Pfäffingen, 3 Erms/Riederich, 4 Glatt/Hopfau, 5 Wiese/Zell, 6 Dreisam/Ebnet, 7 Rems/Neustadt, 8 Fils/Plochingen, 9 Neckar/Horb, 10 Neckar/Plochingen.

aggregated to monthly mean and monthly maximum values. Records containing April to September and October to March data are used to study the summer and winter seasons separately and in the following are referred to as summer and winter data. Despite of the deseasonalisation, different trends in winter and summer data are expected, for example, because of changes in the precipitation patterns. This becomes apparent in the results (see Tab. 3.1 for results).

The trend $\widehat{\boldsymbol{T}}$ is estimated by means of wavelets. Stochastic models of the FARIMA class are fitted to the remaining fluctuations on small scales $\widehat{\boldsymbol{X}}$ using the Whittle estimator. We restrict our analysis to FARIMA(p,δ,q) models with $0 \leq \delta < 0.5$ and $p,q \leq 2$ ($q \leq 1$ for the winter and summer data due to the shortness of the time series). Filtering that is removing the wavelet trend estimate from the data may produce data which is represented best by a model with $\delta < 0$, because not only the trend component but also the stochastic variations on large scales have been removed there. Such occurrences suggest that the long-term parameter δ is redundant in these cases. Therefore we exclude models with an estimated $\delta < 0$, the counterpart model with $\delta = 0$ is considered in the analysis anyway.

To evaluate the influence of the elimination of the trend estimate $\widehat{\boldsymbol{T}}$ from the data, we compare the parameter fits $\widehat{\boldsymbol{\Theta}}'$ on the time series $\boldsymbol{Y}$ itself and the parameter fits $\widehat{\boldsymbol{\Theta}}''$ on the filtered series $\widehat{\boldsymbol{X}}$. In Fig. 3.7 estimates for the parameters δ and ϕ_1, obtained for the FD(δ), AR(1), and FARIMA(1,δ,0) models are shown for monthly maxima (original data and filtered data, the stations are ordered according to the increase in catchment size). The estimates for the two one-parametric models do not change qualitatively. However, the magnitude of ϕ_1 and δ is lower for the filtered data. For the more complex FARIMA(1,δ,0) model filtering causes the short-term correlation parameter ϕ_1 to become larger, whereas the long-term correlation parameter δ becomes smaller. In some cases it becomes compat-

ible with being zero. Therefore, as expected subtracting a possible trend estimate alters the description of the natural variability, i.e. the parameter estimates and possibly the choice of the best stochastic model.

As discussed in Sec. 3.3.3 a possible trend biases the parameter estimation $\widehat{\boldsymbol{\Theta}}'$ of a FARIMA model on $\boldsymbol{Y}$, i.e. under assumption of no trend. To account for the bias the parameter estimation is performed under consideration of the deterministic trend component, represented by the trend estimate $\widehat{\boldsymbol{T}}$, i.e. $\widehat{\boldsymbol{\Theta}}''$ is estimated on the filtered series $\widehat{\boldsymbol{X}}$, which is the recommended procedure (Craigmile et al. 2004). $\widehat{\boldsymbol{T}}$ is a consistent estimator (see Sec. 3.3.5), but it not only contains the deterministic trend component but also stochastic variations on large scales. Therefore, for short time series, $\widehat{\boldsymbol{\Theta}}''$ is likely to be biased in the other direction than $\widehat{\boldsymbol{\Theta}}'$, which is demonstrated in Sec. 3.3.3. Using both estimates for the analysis helps to interpret trend test results. A more reliable parameter set $\widetilde{\boldsymbol{\Theta}}$ could be obtained by extracting a best trend guess $\widetilde{\boldsymbol{T}}$ out of $\widehat{\boldsymbol{T}}$, that is by fitting a polynomial to the trend estimate. $\widetilde{\boldsymbol{\Theta}}$ is then estimated on the detrended series, where $\widetilde{\boldsymbol{T}}$ is subtracted. Another possibility is to use the magnitude of $\widetilde{\boldsymbol{T}}$ in terms of the standard deviation of the process itself to estimate the bias of $\widehat{\boldsymbol{\Theta}}'$ and $\widehat{\boldsymbol{\Theta}}''$ by a Monte Carlo simulation (see Fig. 3.3).

To evaluate the performance of the models with optimised parameters a goodness-of-fit test (see Sec. 2.6.1) is applied and the Akaike-type criterion HIC is used to select among different models (see Sec. 2.5.1). In Tab. 3.1 the best performing models are listed for the monthly averaged data and the monthly maxima, respectively. The apparently best HIC models coincide with a goodness-of-fit test failure in some cases. This either means that no appropriate model has been found to represent the data reasonably well within the FARIMA(p, δ, q) model class, or that the presence of a fairly weak trend is indicated which biases the parameter estimation on $\boldsymbol{Y}$ as well as on $\widehat{\boldsymbol{X}}$. In most cases a goodness-of-fit test failure occurs for the original data as well as for the filtered data. For the Erms River at Riederich (mean values of the summer data) the best HIC model for the filtered data is rejected by the goodness-of-fit test, while the best HIC model is accepted for the original data. So the trend test result of not detecting a significant trend is supported here. In this way goodness-of-fit results may be utilised to check HIC choices and trend test results.

A central result is that three relatively simple models presented in Sec. 2.1.3 perform best in terms of the HIC in the majority of cases, which is presented in Tab. 3.1. By studying these model fits, conclusions about the short and long-term behaviour of the time series can be drawn. In particular the short-term correlated AR(1) model reproduces the dynamics of the fluctuations of the data best in a lot of cases. However, the long-term parameter δ is more important for the monthly maxima data than for the monthly mean values. By considering monthly maxima as indicator for the characteristics of extreme events. This may suggest that long-term correlation plays an important role when examining extremes.

Regarding the trend test result using the standard method, i.e. creating the estimated distribution of the test statistic p_c from the stochastic model fitted to the filtered data, we find a significant trend in 14 of the 60 series as shown in Tab. 3.1. Mostly the trend is combined with an AR(1) model. In case we do not want to risk to falsely detect a trend by any means, we can consider to only accept a trend to be significant when it is found to be significant for parameter estimations on $\boldsymbol{Y}$ as well as for parameter estimates on $\widehat{\boldsymbol{X}}$. Then we find a definitively significant trend for the Dreisam River at Ebnet (complete

(A) monthly means	best FARIMA(p,δ,q) model (original data/filtered data)		
	Total	Summer	Winter
Josbach/Hölzlebruck	[1,0,0]/[1,0,0]	(1,0,0)/(1,0,0)	[1,0,0]/[1,0,0]
Ammer/Pfäffingen	(2,0,1)/**(1,0,0)**	(1,0,0)/**(1,0,0)**	**(0,δ,1)**/**(1,0,0)**
Erms/Riederich	(1,0,0)/(1,0,0)	(1,0,0)/[1,0,0]	(1,0,0)/(1,0,0)
Glatt/Hopfau	[1,δ,0]/**[1,0,0]**	(0,δ,1)/(1,0,0)	[1,0,0]/[1,0,0]
Wiese/Zell	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)	[1,0,0]/[1,0,0]
Dreisam/Ebnet	**(1,0,0)**/**(1,0,0)**	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)
Rems/Neustadt	(1,δ,0)/(1,δ,0)	(1,0,0)/(1,0,0)	(1,0,1)/(1,0,1)
Fils/Plochingen	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)
Neckar/Horb	(1,δ,0)/(1,0,1)	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)
Neckar/Plochingen	(1,0,1)/(1,0,1)	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)
(B) monthly maxima	best FARIMA(p,δ,q) model (original data/filtered data)		
	Total	Summer	Winter
Josbach/Hölzlebruck	[1,0,0]/[1,0,0]	(1,0,0)/(1,0,0)	(1,0,0)/**(0,0,1)**
Ammer/Pfäffingen	(0,δ,0)/(0,δ,0)	(0,δ,0)/(0,δ,0)	(0,δ,0)/**(1,0,0)**
Erms/Riederich	**(1,0,0)**/**(1,0,0)**	[1,0,0]/**[0,0,1]**	(1,0,0)/(1,0,0)
Glatt/Hopfau	**[1,0,0]**/**[1,0,0]**	(1,0,0)/(1,0,0)	(1,0,0)/(1,0,0)
Wiese/Zell	(0,δ,0)/**(1,0,0)**	**(1,0,0)**/**(1,0,0)**	(0,δ,0)/(1,0,0)
Dreisam/Ebnet	[0,δ,0]/**[1,0,0]**	(0,δ,0)/**(1,0,0)**	(1,0,0)/**(1,0,0)**
Rems/Neustadt	(0,δ,0)/**(1,0,0)**	(1,0,0)/(1,0,0)	(0,δ,0)/**(1,0,0)**
Fils/Plochingen	(1,δ,0)/(1,0,0)	[1,0,0]/[1,0,0]	(1,0,0)/(1,0,0)
Neckar/Horb	(0,δ,0)/**(1,0,0)**	(1,0,0)/(1,0,0)	(0,δ,0)/(1,0,0)
Neckar/Plochingen	(0,δ,0)/(1,0,1)	(1,0,0)/(1,0,0)	(0,δ,0)/(1,0,0)

Table 3.1: FARIMA(p,δ,q) models performing best in terms of the HIC for **(A)** original/filtered monthly averaged data and for **(B)** original/filtered monthly maxima data. By filtering, variations on large scales are eliminated to exclude a bias due to a possible trend. Cases with significant trend component are depicted in red. Models enclosed by round brackets have passed the goodness-of-fit-test on a 95%-level of significance, whereas models in squared brackets have been rejected.

data) and the Ammer River at Pfäffingen (winter data) for the monthly mean values and the Erms River at Riederich (complete data) and the Wiese River at Zell (summer data) for the monthly maxima. Thus, all in all, trends are present in the data, but not for the majority of the basin. In comparison to the complete data for the whole year, there is no increased occurrence of trends detectable for the summer or winter season.

In the cases where a possible trend component is found to be significant only under assumption of the trend estimate, i.e. the parameter values estimated on the filtered data, it cannot fully be excluded that this "trend" may be explained as well by stochastic variability. Such a spurious case is, e.g., the monthly averaged complete data for the Ammer River at Pfäffingen. Part of the variability of the data may be explained by a second AR parameter or by a deterministic trend component. The goodness-of-fit values do not provide a distinction here. Regarding the monthly maxima, a typical constellation

occurs in several cases: Part of the variability in the data may either be explained by a long-term correlated model or by a short-term model in combination with a deterministic trend component (cf. Giraitis et al. 2001). In those cases a refined parameter estimation by using a trend guess $\widetilde{T}$, as described above, or the incorporation of knowledge about influencing processes may lead to further conclusions. In case we rather accept the idea of a trend being present in the data than long-range correlation, we detect a fortified occurrence of trends for monthly maxima, that is 6 series out of 10 possess a trend.

When not only focusing on the best fitting model, but comparing the trend test results of all models which have been fitted to the data, we find that in the majority of cases a significant trend was found only under assumption of an AR(1) model (results not shown). This emphasises the influence of the model choice on the trend test result, which is described in Sec. 3.3.3.

The obtained trend estimates help to assess linear trend behaviour in the data. As shown e.g. in Fig. 3.4, the run-off data of the Erms River at Riederich (monthly maxima) is characterised by an increase. Nevertheless, a general tendency of increase or decrease cannot be found for the investigated river run-off data sets in the Neckar basin.

3.5.2 Comparison with Mann-Kendall trend test (Danube River basin)

The rank-based Mann-Kendall test is a very common technique to assess the significance of trends in hydro-meteorological time series (cf. Mann 1945, Kendall 1975, Survey 1994, Douglas et al. 2000 and Kundzewicz et al. 2004). KLIWA (2003), for example, use this test to assess the long-term behaviour of river run-off in the Danube basin. It accepts or rejects the null hypothesis of randomness against the alternative of a monotonic trend. In doing so, the sum of the signs of differences of value pairs is calculated first:

$$s_M = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sign}(x_i - x_j) \quad , \tag{3.10}$$

with the $\operatorname{sign}(x)$ function defined as

$$\operatorname{sign}(x) = \begin{cases} 1 & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x = 0 \\ -1 & \text{for} \quad x < 0 \end{cases} \quad , \qquad . \tag{3.11}$$

s_M is normal distributed with expectation value 0 and variance $\sigma^2_{s_M} = 1/18(n(n-1)(2n+5))$. The Mann-Kendall test statistic is now derived as

$$p_M = \frac{s_M}{\sigma^2_{s_M}} \quad , \tag{3.12}$$

and thus p_M is a standard normal random variable, $p_M \sim N(0,1)$. Therefore the standard normal distribution can be used to apply a two-sided significance test.

An advantage of the Mann-Kendall test is that it relies only on few assumptions: The potential trend may be either linear or non-linear and no assumptions are made regarding the underlying statistical distribution. Nevertheless, it is well recognised that the Mann-Kendall test is not robust against auto-correlation in the sense that false positive trend

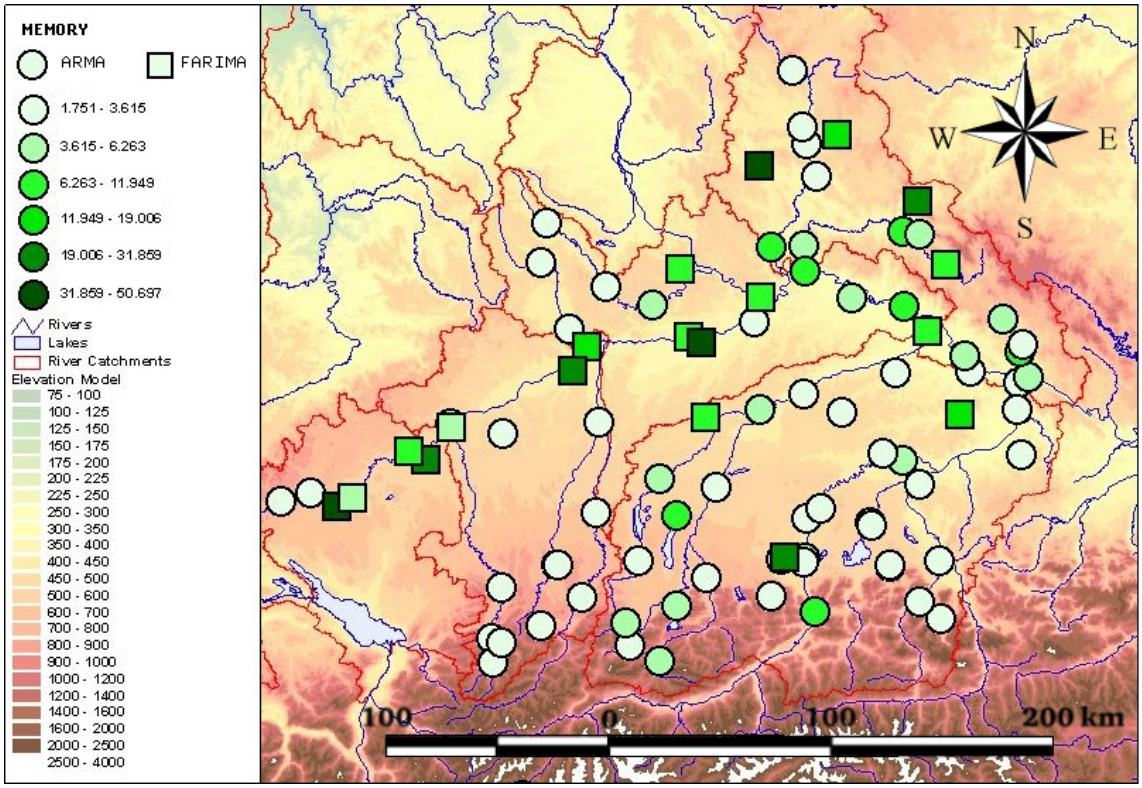


Figure 3.8: Auto-correlation in the Danube basin. Memory of river discharge in the Danube basin (weak (light green) to strong (dark green)). The memory is indicated as a partial sum of the auto-correlation function. In case the wavelet trend estimate of a time series is tested as significant, it is removed from the data before a FARIMA auto-correlation model is fitted. Gauges with long-range correlations are represented with squares whereas circles denote that an ARIMA model seems most suitable.

identifications become more likely (Fleming and Clarke 2002). This effect depends on the sample size as well as on the magnitude of the trend to be identified. Pre-whitening techniques introduced to remove effects induced by auto-correlation may also bias the Mann-Kendall test result and the power of the Mann-Kendall test depends on the underlying distribution (Yue and Wang 2002). Based on these findings, we here propose the CGP trend test which explicitly models the auto-correlations of the data. The soundness of this approach has already been demonstrated. Here a comparison of both methodologies is given for time series in the Danube River basin to assess the severity of the differences for discharge data.

Both trend tests are carried out on monthly mean and maxima data of 97 river run-off records, measured daily at gauges in the Danube basin. The annual cycle in mean and variance is removed from the data as described in Sec. 2.8.1. All available measurements are used. The longest available records cover the time span from 1901 to 2003.

The results of the CGP trend test are obtained with the methodology described in this chapter. This means, the time series is separated into a trend component and a stochastic component, represented by FARIMA models. For most of the analysed data of the Danube River basin dependence is found, which is depicted in Fig. 3.8. Thereby an indicator of the persistence is shown which takes into account the auto-correlation in

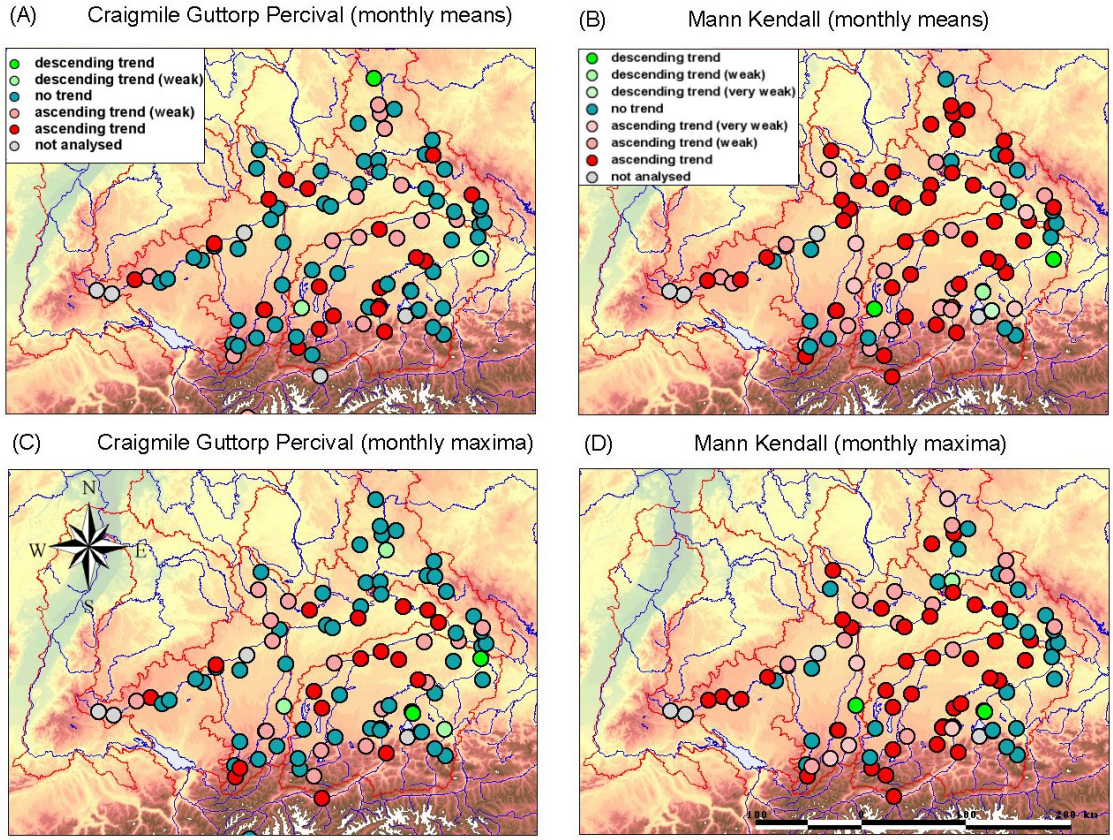


Figure 3.9: Comparison of Mann-Kendall and CGP trend test. Trend test results (green = descending, light green = weak descending, blue = not significant, pink = weak ascending, red = ascending, grey = not examined) for river discharge data in the Danube basin. CGP trend test results **(A)** and **(C)** and Mann-Kendall trend test results **(B)** and **(D)**, for monthly mean and maxima, respectively. The Mann-Kendall results are further investigated according to their significance level (from weak significance with α between 80% and 90% (pink, light green) up to strong significance with α greater than 99% (red, green)).

the data up to a lag of 5 years, that is Memory $= 1 + 2\sum_{k=1}^{5\cdot 365} \rho(k)$. As outlined, auto-correlations are likely to bias trend tests in case they are not considered. River run-off which comprises influences of a lot of processes on different scales, is predestined to reflect some of these forces by a "memory", that is auto-correlations. In Fig. 3.8 an indicator of the strength of the dependence in these series is shown. Some of the gauges even exhibit long-range dependence (quadrates). Apparently, a lot of the gauges along the large Danube River exhibit this long-range dependence. This may be explained by the fact that the memory is mostly stronger for large catchments, where the time series comprises a lot of processes (cf. Mudelsee 2007). Furthermore, gauges close to the Alpes show less memory than stations close to the Danube River. This may be due to fast downhill transportation of the water surplus.

To be able to compare the trend test results, a straight line is fitted to the trend estimate of the CGP trend test. The results are depicted in Fig. 3.9. According to the slope of the trend and the result of the significance test, the gauges are then classified into four categories, namely "not examined" (grey), "descending" (green), "not significant" (blue) and "ascending" (red). Because of the potential bias for the parameter estimates for the

stochastic model (which is described in Sec. 2.4.1), the results of the trend test are furthermore classified into "weak descending" (light green) and "weak ascending" (pink), in case a significant trend was only found for the stochastic model being fitted to the fluctuations on small scales $\widehat{X}$ and not for the stochastic model being fitted to the time series Y itself. In Sec. 3.3.5, references to the quality of the wavelet trend estimate in terms of consistency and asymptotic normality are given. Therefore, it would be sufficient to just use the three categories in case the time series are long enough. However, in case the error of falsely detecting a trend has to be reduced to a minimum, the light green and pink marked trend test results would have to be investigated further.

The Mann-Kendall trend test results are classified into five categories, according to a common scheme (KLIWA 2003). We have

$\alpha < 80\%$	not significant	(blue)
$80\% < \alpha < 90\%$	weakly significant	(descending: light green, ascending: light pink)
$90\% < \alpha < 99\%$	probably significant	(descending: green, ascending: pink)
$\alpha \geq 99\%$	very likely significant	(descending: dark green, ascending: red)

where α is the significance level.

In Fig. 3.9 (A) and (B) the trend test results are given for monthly mean values and in Fig. 3.9 (C) and (D) for monthly maxima, respectively. Both tests mostly find increasing trends, if significant. Nevertheless, the Mann-Kendall trend test often detects a significant trend where the CGP trend test does not.

The CGP and the Mann-Kendall trend test results differ even qualitatively, which is depicted in Fig. 3.9. The differences in the test results may be explained by the auto-correlation structure of the data, which is considered when applying the CGP trend test and which is likely to bias the Mann-Kendall trend test. In this case, the CGP trend test offers more reliable results. Altogether the CGP trend test considers two-thirds of the monthly mean trends and half of the monthly maxima trends found by the Mann-Kendall trend test as not significant. This is an important difference. Thus, for series which evidently exhibit auto-correlations it is more adequate to use the CGP trend test.

Most of the significant trends are increasing. Nevertheless, no directly interpretable spatial pattern for the catchments near the Danube River is detectable. Other site-related influences may have to be considered, e.g., catchment size, soil composition near the station or altitude. This holds for monthly mean values and monthly maxima. Analysing monthly maxima is an attempt to get an impression about the trends of more extreme river run-off observations. This attempt is refined by applying standard extreme value analysis methods as pointed out in Chap. 4.

CHAPTER 4

TRENDS IN EXTREMES

4.1 TREND DETECTION APPROACHES FOR EXTREMES

Climate change is associated with an increasing global mean temperature. The interactions between this temperature rise and the hydrological cycle is still a matter of debate. But what impact on the extremes of hydrological data, that is floods and droughts, do we expect from global warming? As outlined in Chap. 1, an increase in frequency and magnitude of floods and droughts because of climate change is anticipated.

There exist approaches to directly conclude from changing properties of mean values of a series to the change in the extremes. However, this goes along with a specification of the relationship of mean values and extremes. In general, the function between a trend in the extremes and a trend in the mean values is non-linear and does not show an obvious relation (Nogaj et al. 2006). The frequency of extreme events might change dramatically as a result of even a small change in the mean of the sample (Wigley 1985). We rather try to directly consider trends in extremes, despite of the scarcity of available data. In the last decades several methods have been established to do so. On the one hand, there exist semi-parametric methods. Here the distribution is described using parameters, but the trend is estimated data driven, i.e. the trend shape is estimated from the empirical time series itself. Davison and Ramesh (2000) and Ramesh and Davison (2002) use local polynomial fitting (kernels) and a generalized extreme value (GEV) distribution, which is fitted to block maxima, to model non-stationary extremes. Gaetan and Grigoletto (2004) combine this GEV distribution with a latent time process. Mudelsee et al. (2004) rather use peaks over a threshold (POT) and kernels whereas Hall and Tajvidi (2000) combine kernel smoothing with the generalized Pareto distribution (GPD). Splines and point processes are joined by Chavez-Demoulin and Embrechts (2004) to model extremes. On the other hand, non-stationary extremes are modelled using fully parametric methods. Here the trend form of the parameters has to be assumed before the analysis and is described using given functions, such as polynomials, et cetera. Smith (2003), Strupczewski et al. (2001), Coles (2001), Willems (2005), Bárdossy and Filiz (2005) and others examine hydro-meteorological data using this technique. The parametric approach can as well be combined with a variety of distributions, such as the GEV or GPD. Both distributions can be integrated in a class of point processes to model extremes. Smith (1989), for example, fits point processes to ozone data. Fully parametric approaches provide the possibility to extrapolate the estimated trend in the future and to use it, for example, for prediction issues. The sketched methods to assess non-stationary extremes provide a sound basis for extreme value analysis. However, they exhibit advantages and disadvantages and a universally applicable framework does not exist. There is still potential to improve the methodology and to apply it to hydro-meteorological data. In the following, we set up a point process framework to examine possibly non-stationary extremes.

4.2 TREND DEFINITION

Trends in sample extremes are important in many contexts. An event is classified as *extreme* by society if the losses caused by this event are intolerable and future events of this kind have to be prevented. From a procedural point of view a *trend* in extreme values of hydrological data is given by a change of the *severity of flood events*. This is described by a change in the magnitude of flood events of a fixed return period, or by a change in the *frequency of flood or drought events*, which is expressed by a change of the return period (i.e. frequency of occurrence) of a flood event of a fixed magnitude. We choose a point process approach to model extreme events and therefore are able to capture both phenomena (see Sec. 4.5).

In Sec. 3.2 we provided a definition of trend which holds also for the evaluation of extreme values: We look at trends as a slowly varying change which manifests on large scales. In extreme value analysis we use a fully parametric approach. This means we have to assume a certain trend shape before the analysis. The shapes chosen, i.e. slowly varying polynomials or exponentials, reflect this view of trends.

Here, the presented framework to assess non-stationary extreme values is capable to deal with a broader range of trends, which even may incorporate abrupt jumps. This is especially useful in case covariates have to be taken into account. By this means classificatory data, such as indices or series of weather regimes can be used as explanatory variable.

4.2.1 Trend shape and trend detection methods

The necessity to incorporate non-stationarity into extremal models has been met in different ways. All authors referenced here introduce non-stationarity by varying the parameters of distributions, which model the extreme events, with time. There exist fully parametric approaches where a trend shape has to be assumed before the assessment, and semi-parametric methods where the trend shape is obtained in a data-driven way. The aspects of both approaches which have been discussed in Sec. 3.2.3 hold also here.

A fully parametric approach is, e.g., introduced by Smith (2003), Coles et al. (2002), Strupczewski et al. (2001), and Bárdossy and Filiz (2005). In these frameworks only the location and scale parameter are non-stationary. A semi-parametric procedure is used by Chavez-Demoulin (1999) and Davison and Ramesh (2000). Thereby the trend shape of the parameters is derived directly from the data by using kernels or splines. In most cases the GEV or GPD distribution is selected to represent the data. Naturally other extreme value distributions can also be extended to non-stationary models (see, for example, Strupczewski et al. 2001 and Bárdossy and Filiz 2005).

In the framework presented here, a point process is used to model magnitude and frequency of extreme events. Thereby we combine an inhomogeneous Poisson process with the GPD, as outlined in Secs. 2.3.2 and 2.3.4. We therefore derive the likelihood given in Eq. (2.54) for the point process. In this fully parametric approach, non-stationarity is introduced by a time varying rate λ_t and scale parameter ψ_t of the GPD. By this means, we can benefit from several advantages. First of all the representation of the extremes is improved by using threshold excesses (see Sec. 2.3.3). Furthermore, we are capable

to model magnitude and frequency of occurrence of the extremes. Finally, we want to separately maximise the likelihood of the two Poisson processes which model frequency and magnitude of the extreme events. Therefore we use the notation of the point process in terms of the GPD parameters.

Trends in parameters of extreme value models and relation to extremes

As outlined in Sec. 2.3.4 non-stationarity is introduced via non-stationary parameters. These parameters are directly linked to the moments of this distribution. The connection is exploited in the *method of (weighted) moments*, where estimators of the moments are used to estimate the parameters of the GEV (see, e.g., Maidment 1993). Let $\{M_t\}$ be a sequence of iid Gumbel distributed maxima, then

$$\begin{aligned} E(M) &= \mu + \sigma\gamma \\ \gamma &\approx 0.57722 \text{ (Eulers constant)} \\ \mathrm{VAR}(M) &= \sigma^2 \frac{\pi^2}{6} \approx 1.645\sigma^2 \quad . \end{aligned} \tag{4.1}$$

For $\{M_t\}$ having a GEV distribution with shape parameter $\xi > 0$ or $\xi < 0$, we get

$$\begin{aligned} E(M) &= \mu + \frac{\sigma}{\xi}[1 - \Gamma(1+\xi)] \\ \text{and for } \xi > -0.5 & \\ \mathrm{VAR}(M) &= (\frac{\sigma}{\xi})^2\{\Gamma(1+2\xi) - [\Gamma(1+\xi)]^2\} \quad . \end{aligned} \tag{4.2}$$

This shows that a trend in the location or in the scale parameter of the GEV distribution causes as well time dependence of mean and variance of the extremes.

Using Eq. (2.29) and Eq. (2.34) we transfer these relationships to a connection between the parameters $\mathbf{\Theta} = (\lambda, \psi, \xi)$ of a point process and the moments of the distribution of the extremes. Let $\{M_t\}$ be a sequence of iid Gumbel distributed maxima and $Y = X - u$ the corresponding excesses over a threshold u which are distributed as $GPD(y) = 1 - \exp(-y/\psi)$ for $y > 0$. Then we get

$$\begin{aligned} E(Y) &= u + \psi\gamma \\ \mathrm{VAR}(Y) &= \psi^2 \frac{\pi^2}{6} \approx 1.645\psi^2 \quad . \end{aligned} \tag{4.3}$$

For $\{Y_t\}$ having a GPD distribution with shape parameter $\xi > 0$ or $\xi < 0$ we get

$$\begin{aligned} E(Y) &= u + \frac{\psi}{\xi}[1 - \lambda^\xi\Gamma(1+\xi)] \\ \text{and for } \xi > -0.5 & \\ \mathrm{VAR}(Y) &= (\frac{\lambda^\xi\psi}{\xi})^2\{\Gamma(1+2\xi) - [\Gamma(1+\xi)]^2\} \quad . \end{aligned} \tag{4.4}$$

Furthermore, important assessment measures, e.g. return levels, are computed via the point process parameters (cf. Eq. (2.45)). From there it follows that the non-stationarity of the rate and scale parameter, respectively, influence the evolution of the return level. We assume for all our analyses a fixed shape parameter ξ (cf. Sec. 2.3.4). In this case the return level z_p changes linearly with the scale parameter ψ and the rate λ times ξ.

4.3 CHOICE OF EXTREMES

To be able to estimate a trend in extreme values some assumptions have to be fulfilled which are outlined here.

4.3.1 Choice of season

In the setting proposed here, seasonality has to be handled as a deterministic part of hydro-meteorological data. As outlined in Sec. 2.8.1 this kind of dependence can either be modelled explicitly or has to be eliminated before the analysis. Here we decided to remove seasonality by looking at subsamples of the data, i.e. classifying the time series into seasons. Thereby we assume that no intra-seasonal effects remain. For all our extreme value analysis, we examine the winter season, that is December to February. Within this period we find reasonably low inter-seasonal variability and, moreover, in this season we expect variation because of changes in the precipitation pattern (see Sec. 2.7). Further details about the treatment of the annual cycle are given in Sec. 2.8.1.

4.3.2 Threshold selection

We obtain the extreme values by selecting a threshold u and by defining all data values which exceed this threshold as extremes. To select the threshold we use the non-parametric mean residual life (MRL) plot and a parametric approach. The MRL plot depicts the sample mean of the threshold excesses for different u. In case the extreme values are distributed according to a GPD the sample mean is an estimator for the expected value of the extremes. We select the lowest u_0 such that the MRL plot changes linearly for $u > u_0$. In the parametric approach a GPD distribution is fitted to the extreme values for a range of thresholds. Then the lowest u_0 is selected, such that parameter estimates for $u > u_0$ stay approximately constant. Further details are given in Sec. 2.3.2.

For all our studies on empirical data we proceed as follows: For each individual river gauge we compare the results of a MRL plot and a GPD fit over a range of thresholds to cross-check and to derive reliable results. We had to proceed station by station and we did not automatise this step, because an "objective" procedure in which u_0 is derived automatically from the results of the MRL method and the fit of the GPD over a range of thresholds gave unplausible results in too many cases. For each time series, we were able to determine a reliable threshold u by using the methods proposed in Sec. 2.3.2. The choice is verified by analysing the corresponding threshold excesses (see Sec. 4.7).

4.3.3 Declustering

Subsequently, we take care of the auto-correlation structure in the data. The Fisher-Tippett theorem holds for iid or dependent but stationary data (cf. Secs. 2.3.1 and 2.3.2). Thus, to be able to introduce non-stationarity, we remove the auto-correlation structure in the data before the analysis. By using threshold excesses we are able to group the data into clusters and only keep the maxima of these clusters, i.e. we decluster the time series

as described in Sec. 2.8.2. Regarding the empirical data of the Danube River basin a cluster size of 6 to 12 days is sufficient to remove the auto-correlation structure in the data (cf. Tab. A-4).

Analysing the relation between cluster size and threshold we find for some gauges that it is more effective to choose a higher threshold than the threshold selection criteria have suggested. In these cases at first a lot of extreme values were available because of the low threshold. But then we always had to choose a large cluster size when declustering which in turn reduced the number of extreme values available. In case a higher threshold was selected, nearly the same number of extreme values was left after declustering, but the remaining extreme values were larger, i.e. this set is better representing the extreme events. Regarding the adequacy of the cluster size chosen, we find for two thirds of the empirical series that the remaining excesses are compatible with being independent at a 5% level and one third at a 10% level, which we consider as satisfactory. In Tab. A-4 the threshold choice and cluster size are listed for each station.

4.4 TREND ESTIMATE

Having selected the extremes to analyse, we select a point process with a possibly time dependent parameters to model them (see Secs. 2.3.2 and 2.3.4 for further details). We assume trend forms for the rate parameter λ and scale parameter ψ. As outlined in Sec. 2.3.4, we let the shape parameter be constant for all our analyses.

Regarding the rate parameter λ_t, a constant $\lambda_t \equiv \lambda = a_0$ is included and exponential trend shapes:

$$\lambda_t = \exp\Big\{ \sum_{i=0}^{k} a_i t^i \Big\} \qquad k = 1, \ldots, 4 \quad , \tag{4.5}$$

with the a_i being constants. These shapes assure a slowly varying λ_t. Furthermore, the exponential relation is regarded as the "most natural" simple relationship for rates of occurrence (cf. Cox and Lewis 1966). It is chosen by various authors (see MacLean 1974 and Mudelsee et al. 2004).

With respect to the scale parameter we consider the following shapes:

$$\begin{aligned} \text{(i)} \quad \psi_t &= \sum_{i=0}^{k} b_i t^i \qquad k = 0, \ldots, 4 \\ \text{(ii)} \quad \psi_t &= \exp\Big\{ \sum_{i=0}^{k} b_i t^i \Big\} \qquad k = 1, \ldots, 4 \quad , \end{aligned} \tag{4.6}$$

with the b_i being constants. Polynomials are rather flexible and therefore fit well to a variety of shapes. We test this by simulation studies: even maxima with an S-curve shaped trend are well approximated by using models with polynomial shaped trend guesses. Furthermore, we utilise also the exponential trend guesses described in Eq. (4.6) (ii). This trend form does not allow for a shape parameter $\psi_t < 0$. This is a necessary condition, because the scale parameter is related to the variance of the extreme values.

Our extreme value model class is derived from all possible combinations of Eqs. (4.5) and (4.6). By fitting the models to the empirical data we obtain values a_i and b_i for each

model and together with the respective trend guesses this results in the estimates $\hat{\lambda}_t$ and $\hat{\psi}_t$.

Considering consistency and efficiency we use the maximum likelihood approach to obtain the estimates for λ_t of the Poisson process and the parameters ψ_t and ξ of the GPD. Therefore, regarding the characteristics of the trend estimate, all statements made in Sec. 2.4.3 about consistency and efficiency of maximum likelihood estimators hold for $\xi > -0.5$ (cf. Chavez-Demoulin 1999). For almost all empirical records we found ξ being larger than -0.5. We proceeded with those gauges to benefit from the asymptotic properties of the ML estimator.

4.4.1 Trend extrapolation in the future

As outlined beforehand a fully parametric framework makes it possible to extrapolate an estimated trend in the future. On the one hand, this feature is desirable, because especially water management issues often require assumptions about future flood evolution. On the other hand, caution is advised, because implicitly all assumptions about the system which have been made for the analysis are expected to stay true in the future as well. Several ways to use extrapolated trends for the calculation of return levels are given in Sec. 2.3.4. A cautious and common procedure is to estimate and consider a trend for the observed time series, but not to extrapolate it in the future (see, e.g., Bárdossy and Pakosch 2005). We proceed also in this manner.

4.5 POINT PROCESS APPROACH

Having selected the extremes which incorporates the choice of a threshold u and a cluster size $2r$, a class of point process models is fitted to the extreme values by using the likelihood given in Eq. (2.54). That is, a stationary point process model and a variety of non-stationary point process models with non-stationary parameters λ_t and ψ_t, as given in Eqs. (4.5) and (4.6) are used as a model library. All models are fitted to the time series and the best one is chosen via some selection criterion (see Sec. 4.6). This model is then used to calculate assessment measures for the represented extreme values, such as the return levels.

To obtain a visual impression about the performance of non-stationary point process models we compare them to a sliding window approach. In Fig. 4.1 the evolution of the parameter estimates $\hat{\boldsymbol{\Theta}} = (\hat{\lambda}_t, \hat{\psi}_t, \hat{\xi})$ of the best non-stationary point process for the Inn River at Rosenheim is shown in red. Good agreement is achieved with the sliding window approach, which are depicted in black. We have to choose rather short intervals for the sliding windows to track changes, which is a common disadvantage of the sliding window methodology. Therefore the parameter estimates within these windows are afflicted with great uncertainty and for some windows no reliable estimation is possible, as apparent in Fig. 4.1 for the scale and shape parameters around 1975. The confidence intervals[1] of one standard deviation are nearly never left by the parameter estimates of the

[1] These confidence intervals just give an indication. On the one hand, they are too broad, because only information of 20 years of measurement is used. On the other hand, they are too narrow, because the asymptotics they assume is not reached yet. By Monte Carlo simulation more accurate confidence intervals could be obtained.

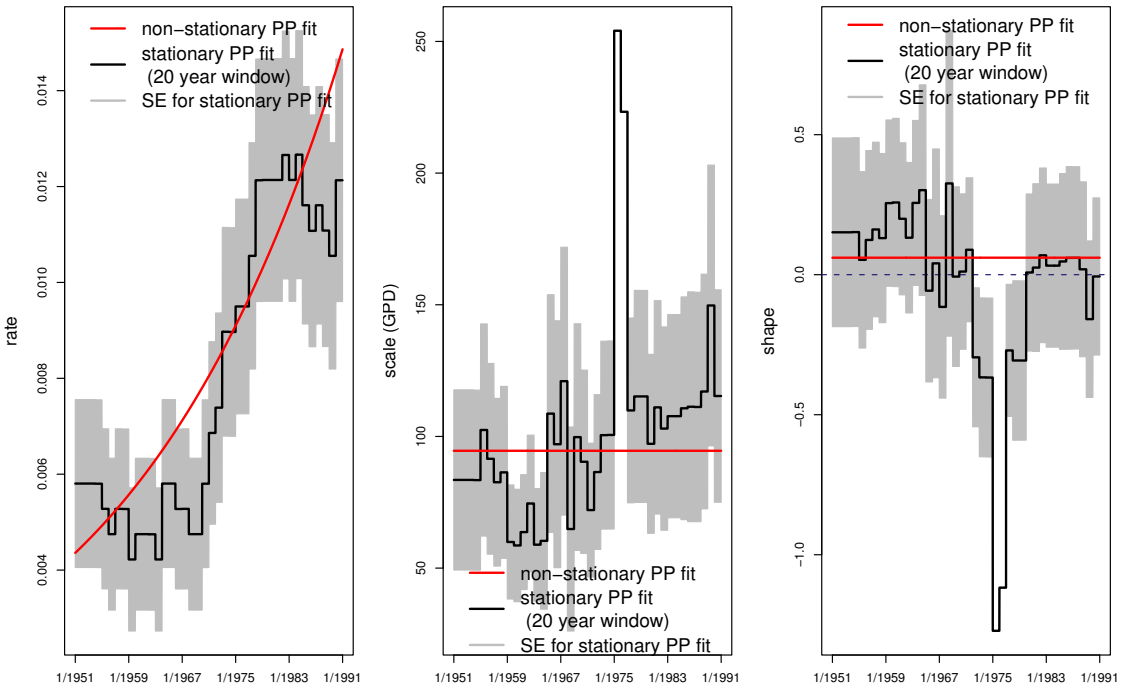


Figure 4.1: Validation of non-stationary point process. The evolution of parameter estimates of the best non-stationary point process representing discharge of the Inn River at Rosenheim is shown in red (**left**: rate λ, **middle**: scale ψ, **right**: shape ξ). For comparison, parameter estimates of a sliding window approach are depicted in black. The estimates are obtained from a stationary point process, which is fitted to the data in windows of a length of 20 years. The windows are moved over the time series with annual interspace, i.e. they overlap. The corresponding confidence intervals of one standard deviation, obtained with the delta method are shown in grey.

non-stationary point process. It is also visible that the sliding window approach suggests a time varying shape parameter which even varies across zero. We obtain this outcome for several river run-off series in the Danube River basin. Nevertheless, we keep the shape parameter ξ constant according to the reasons sketched in Sec. 2.3.4. Altogether, this visual validation of the non-stationary point process approach shows good results for most of the time series analysed.

Regarding the choice between a point process notation in terms of $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$ or in terms of $\boldsymbol{\Theta} = (\mu, \sigma, \xi)$, i.e. choosing between the likelihood given in Eqs. (2.54) and (2.55), we conduct some simulation studies to evaluate the performance of both approaches. First of all we assess the quality of the fit of the point process using both notations as described in App. A-2. The point process with notation in terms of the GEV parameters produces acceptable results, but is outperformed in most of the cases by the GPD. To assess the performance of the point process using $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$ the uncertainty of estimating λ has to be added to the GPD performance. However, this uncertainty is small compared to the uncertainty induced by the estimation of the other parameters. The reason is that λ is estimated on the whole time series of length n. Therefore we get satisfactory results with low uncertainty when estimating λ[2]. Furthermore, the capability of both approaches to generate realisations of point processes is tested (see App. A-3).

[2]To estimate λ, we use the standard ***R*** routine *glm()*.

Artificially generated data are useful for uncertainty assessment and the simulation of distributions of test statistics in case no theoretical distributions can be assumed. Using the point process with notation in terms of $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$ we systematically obtain more adequately simulated series. This holds, because here we are able to use the same smooth exponential functions for λ_t when fitting and when generating point processes. When generating data from the point process with notation in terms of the GEV parameters, we have to calculate λ from $\boldsymbol{\Theta} = (\mu, \sigma, \xi)$. Therefore we use the point process with notation in terms of $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$ for all our studies.

4.6 TEST FOR SIGNIFICANCE OF THE TREND IN EXTREMES

In the proposed framework the trend and distribution of the data are jointly estimated. Thus the test for significance of a trend in the extreme values is transferred to a problem of model selection: A class of point process models with non-stationary rate and shape parameter and a stationary point process model is fitted to the data. Then the deviance statistic is used to select the best model. The deviance statistic is based on maximum likelihood estimation and designed to compare models of different complexity (Sec. 2.5.2). By using the point process expressed as function of $\boldsymbol{\Theta} = (\lambda, \sigma, \xi)$ we are able to split the test for trend into two steps: On the one hand, the deviance statistic is used to select between the stationary Poisson process and inhomogeneous Poisson processes with an exponential rate which model the frequency of extreme events. On the other hand, we use the deviance statistic to choose between a stationary GPD model and GPD models with time-varying scale parameters which model the magnitude of the extremes. Advantageously in each case only the non-stationarity of one parameter is tested. Furthermore, in our methodology the extreme value distribution is explicitly considered when testing for trends. Such trend detection frameworks are superior to, e.g., ordinary least squares regression or methods based on Kendall's tau (cf. Zhang et al. (2004)).

4.6.1 Power of the trend test

The deviance statistic is asymptotically ($n \to \infty$) χ^2 distributed. We study the power of the trend test (based on the deviance statistic) by means of a simulation study. Thereby the length of the time series and the magnitude of trends resemble the situation given in our empirical analyses.

Concerning the modelling of the occurrence rate, artificial data are generated from a Poisson process with a linear trend in the exponential function of λ_t, that is $\lambda_t = \exp(\log(\lambda_0) + s \cdot t)$ for $t = 1, \ldots, n$ where s is a constant slope. The slope s of this trend component is varied between zero and 0.0005. Regarding the GPD distribution, artificial data are simulated from a GPD model with a linear trend in ψ_t with slope ranging from zero to 0.05. To mimic empirical conditions, the parameters for both simulation procedures are taken from a stationary PP fit to the Danube River at Beuron, i.e. $u = 65\text{m}^3/\text{s}$, $\psi_0 = 43.9$, $\xi = 0.076$ and $\lambda_0 = 0.01$. We found occurrence rates ranging from 0.01 to 0.04 in our empirical series and thus we use these two rates as λ_0 in our simulation study. We generate time series with 5 415 (18 050) data points, which equals 60 (200) years of daily measurements and evaluation at a season of three months.

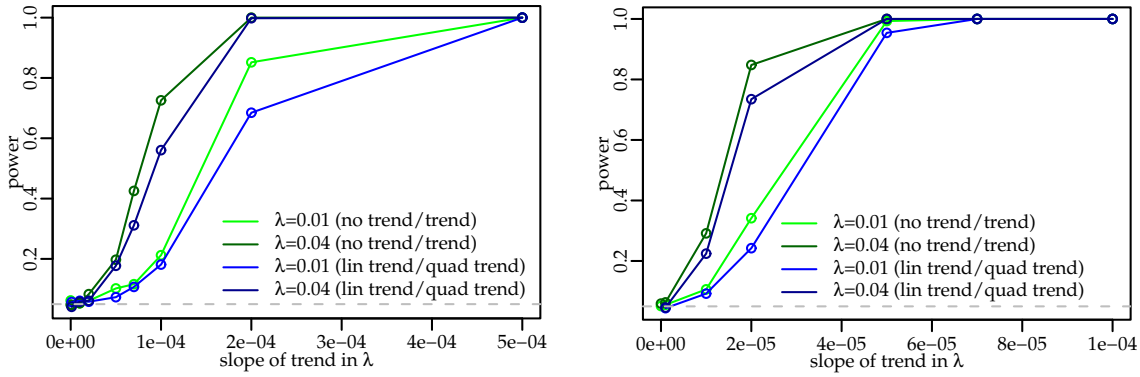


Figure 4.2: Power of the trend test (based on the deviance statistic) for the Poisson process: Evaluated with 1 000 simulation runs for each trend slope. A non-stationary Poisson process with an exponential trend in λ_t is generated (its slope is varied along the x-axis) and a stationary Poisson process and a non-stationary Poisson process with exponential trend assumption in λ_t are fitted to the artificial series. At a significance level of 5% the deviance statistic is then used to select the best model. **Left**: Power of the trend test according to the deviance statistic for data with length n =5 415. **Right**: Power of the trend test for data with length n =18 050.

Then two models, namely a stationary Poisson process and one assuming an exponential trend in λ (the "right" model), are fitted to the artificial Poisson process data. Respectively, the stationary GPD model and one assuming a linear trend in ψ are fitted to the generated GPD series. In both cases the better model is chosen according to the deviance statistic with a significance level of 5%. For each value for the slope this study is repeated 1 000 times. In case no trend is present in the artificial data, we therefore expect to falsely detect a trend in 50 cases (error of Type I). Even if a weak trend is present in the data, we would like the test to detect it. That is the test should detect a trend in all cases which is equivalent to a power of one.

In Fig. 4.2 the power of the trend test (based on the deviance statistic) for the choice between the Poisson processes is depicted. For the length of 5 415 data points (left hand side of the figure) and a rate of $\lambda_0 = 0.01$, i.e. about 60 extreme events at hand, a trend with slope $s = 0.0002$ is detected in 85% of the cases. That implies for the observation period of 60 years the rate of occurrence of extreme events changes from once per year to 2.7 times a year. The corresponding power curve of the trend test is depicted in green. For a higher rate $\lambda_0 = 0.04$, that is about 240 extremes available, the power of the trend test is improved (dark green line). Here we detect 72% of the trends at a slope $s = 0.0001$, which indicates a change of 0.9 extreme events per year to 1.5 extreme events per year within 60 years. The length of the data plays an important role as well when evaluating the power of the trend test: In the right hand side of Fig. 4.2 the results for series with a length n =18 050 data points, representing 200 years of measurement, are depicted. Here the performance of the trend test is much better (for better visibility only the slopes from 0 to 0.0001 are depicted). A trend is detected at a slope of 0.00005 in nearly 100% of the cases, thereby such a trend indicates a change from one extreme event per year to 1.2 extreme events per year within 60 years.

In either case we expect to falsely detect a trend for 5% of our ensemble (this level is indicated by the dotted grey line in Fig. 4.2) when no trend is present in the data, i.e. $s = 0$. This expectation is met, the rate of falsely detecting a trend in stationary series

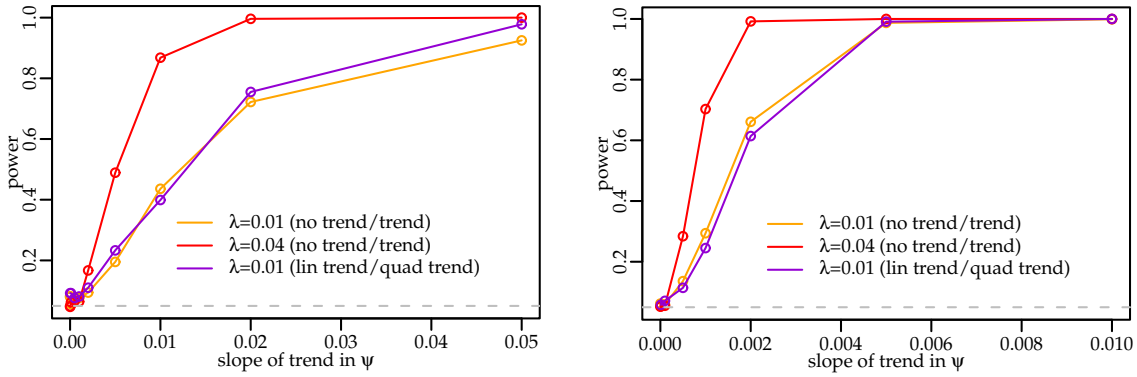


Figure 4.3: Power of the trend test (based on the deviance statistic) for the GPD model: Evaluated with 1 000 simulation runs for each trend slope. A non-stationary GPD process with a linear trend in ψ_t is generated (the slope is varied along the x-axis) and a stationary GPD model and a non-stationary GPD model with linear trend assumption in ψ_t are fitted to the artificial series. At a significance level of 5% the deviance statistic is then used to select the best model. **Left**: Power of the trend test for data with length n =5 415. **Right**: Power of the trend test for data with length n =18 050.

ranges from $4.4 - 6.5\%$ for all our simulation settings. Thus, by using this method, we surmount the risk of a systematically false detection of trends.

We evaluate also the performance of the deviance statistic to differentiate between types of trends, which is represented in Fig. 4.2 by blue lines. To do so, we generate series from inhomogeneous Poisson processes with an exponential rate and a quadratic trend, i.e. $\lambda_t = \exp(\log(\lambda_0) + s \cdot t^2)$ for $t = 1, \ldots, n$. Then we fit an inhomogeneous Poisson process with the assumption of a linear trend in the exponential and an inhomogeneous Poisson process with the assumption of a quadratic trend in the exponential (the "right" model) to the data. Finally, we use the deviance statistic to decide which model to prefer. The corresponding power curves for the trend distinction (based on the deviance statistic) are depicted in blue (λ_0=0.01) and dark blue ($\lambda_0 = 0.04$). We obtain similar results as for the discrimination between no trend and linear trend. Therefore we conclude that no extra procedure is necessary to decide, whether there is a trend present in the data at all (such strategies are proposed by, e.g., Cox and Lewis 1966). We are able to fit the whole class of the stationary model and non-stationary models to the data and then select the best one using the deviance statistic.

Regarding the performance of the deviance statistic to discriminate between GPD models, the results are depicted in Fig. 4.3. Here a linear trend is added to ψ_t. Thereby the time unit of the slope is days, i.e. a slope $s = 0.02$ denotes a $\Delta\psi$ of 1.8 within 90 days or one year. On the left hand side the results for series with a length of 5 415 data points are depicted. A rate $\lambda = 0.01$ indicates that there are about 60 extremes available. Two GPD models are fitted to them, namely a stationary GPD model and a GPD model with linear trend assumption for ψ_t. The corresponding power curve of the trend test to discriminate between these two models is depicted in orange. Here a trend with slope $s = 0.02$ is detected in 80% of the cases. Such a trend indicates a $\Delta\psi = 110$ within 60 years. This variation can be expressed in terms of a change of the variance and the expected value of the excesses (see Eq. (4.4)). For ξ remaining constant, we obtain $\Delta E(Y) = c_1 \Delta\psi$ and $\Delta\mathrm{VAR}(Y) = c_2(\Delta\psi)^2$ with c_1 and c_2 being constants. In our simulation study $c_1 = 0.4$

and $c_2 = 1.4$ is given, that is a change of the mean of the extremes of $44\text{m}^3/\text{s}$ or a change in the standard deviation of $\Delta SD = 130\text{m}^3/\text{s}$ within 60 years. This seems to be quite a lot of change which is only detected in 80% of the cases (orange curve). So here the power of the trend test is not good. For comparison, we examine the change in the standard deviation which is present in the empirical series we analyse (see Sec. 4.7 for further details on extreme value analysis of the Danube River basin), and there we get a range from $\Delta SD = 0\text{m}^3/\text{s}$ to $\Delta SD = 64\text{m}^3/\text{s}$.

Naturally improved results are obtained in case a higher rate of extreme events is available (red curves in Fig. 4.3) or in case the time series are longer (right hand side of Fig. 4.3). For a series of 18 050 data points and a rate $\lambda = 0.04$, i.e. about 800 extreme events at hand, the power of the trend test becomes remarkably better. In 100% of the cases a trend with slope $s = 0.002$ is detected, which indicates a $\Delta E(Y) = 4.4\text{m}^3/\text{s}$ or a ΔSD of $13\text{m}^3/\text{s}$ within 60 years.

In case no trend is present in the data, a trend is detected too often for 5 415 data points, i.e. in 8% of the cases. This effect vanishes in the other simulation settings, that is we do not too often falsely detect a trend.

As for the Poisson model it is the aim to clarify for the GPD model whether it is necessary to proceed the trend test in two steps. That is, first to decide whether a trend is present in the data at all and then, in case we assume a trend in the data, to use the deviance statistic to decide which of the non-stationary models to choose. Therefore we generate non-stationary GPD data with a quadratic trend in ψ_t and fit two non-stationary GPD models, namely one with a linear trend assumption and one with a quadratic trend assumption in ψ_t (the "right" model) to the artificial data. The resulting power curves for the trend test are depicted in violet in Fig. 4.3. The power curves are comparable to those which indicate the power of the deviance statistic to discriminate between trend and no trend. Therefore we here conclude as well that it is not necessary to introduce an extra step to decide whether a trend is present in the data at all.

As outlined in Sec. 2.5.2 it is theoretically well founded to use the deviance statistic to discriminate between the models presented here, that is to discriminate between Poisson processes with non-stationary rate λ_t and GPD models with non-stationary scale parameter ψ_t, respectively. The results shown here evaluate the power of the deviance statistic in case it is applied to time series with few data points. They indicate the potential and limits of this trend test. These limits depend on factors such as the parameters $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$, the length of time series and the shape of the trend which has to be detected.

4.6.2 Simulating the distribution of the deviance statistic

As revealed by simulation studies in the previous section the power of the deviance statistic to detect a trend in the scale parameter ψ_t of the GPD is weak for time series which resemble the empirical records we analyse.

Therefore we investigate, whether the power of the deviance statistic increases in case we simulate its distribution. Simulating the distribution of a test statistic is a feasible method in case asymptotic results are not usable because of insufficient data length. By simulating the distribution we do not have to consider, whether the models compared are

nested[3] or not. Furthermore, we use that a poorly fitting model generates data which do not have the same statistical properties as the original data set. The approach we follow is introduced in a similar way by Hinde (1992).

Let D^{emp} be the deviance statistic, derived from the comparison of fits of models M_1 and M_2 to empirical data with parameter vectors $\hat{\mathbf{\Theta}}_\mathbf{1}(t)$ and $\hat{\mathbf{\Theta}}_\mathbf{2}(t)$. We now generate a distribution of D under the condition that the empirical time series is a realisation of model M_1 by

1. generating an ensemble of R datasets from model M_1 using $\hat{\mathbf{\Theta}}_\mathbf{1}(t)$ as parameter vector.
2. fitting model M_1 and M_2 to the data and calculating $D(1)^i_{(1)(2)}$ for $i = 1, \ldots, R$.

$D(1)_{(1)(2)}$ denotes the deviance statistic comparing models M_1 and M_2 using data which is generated from model M_1.

We proceed in the same way to get a distribution of D under the condition that model M_2 is the "true" model. Then we compare D^{emp} with the two distributions to decide, whether the observed time series has been generated from one of the models or not. Thereby a significance level α has to be defined, which we set to 5%.

We evaluate the performance of the deviance statistic using a simulated distribution by simulation studies. Here we use similar settings as for the simulation studies in which the performance of the deviance statistic using its theoretical distribution is tested (see Sec. 4.6.1) and we set $R = 1\,000$. A Poisson process time series with a linear trend in the exponential of λ_t and a GPD distributed time series with a linear trend in ψ_t are generated and three types of models are fitted to them:

(1) a stationary model (too simple),

(2) a non-stationary model with a linear trend assumption (appropriate model),

(3) a non-stationary model with a trend assumption with linear and quadratic terms (over-fitted model).

To compute the "empirical" deviance statistic the maximum likelihood for these fits (1)-(3) is calculated and the deviance statistic is retrieved according to Eq. (2.61). $D^{emp}_{(1)(2)}$ compares models (1) and (2), $D^{emp}_{(1)(3)}$ and $D^{emp}_{(2)(3)}$ compare the other models, respectively. The distributions of these deviance statistics under the condition that the data are a realisation of one of the processes is now generated by simulation, that is the $D(a)^i_{(a)(b)}$ and $D(b)^i_{(a)(b)}$ are calculated with $a, b = 1, 2, 3$ and $i = 1, \ldots, R$ with R being the size of our simulation ensemble.

To decide whether to choose model (1) or (2) to represent the artifically generated Poisson process (or GPD distributed time series) we compare $D^{emp}_{(1)(2)}$ with the two distributions which are composed of $D(1)^i_{(1)(2)}$ and $D(2)^i_{(1)(2)}$ for $i = 1, \ldots, R$. In the same way $D^{emp}_{(1)(3)}$ and $D^{emp}_{(2)(3)}$ are compared with the corresponding simulated distributions of D.

[3]Model M_1 is nested in model M_2 when it is a special case of M_2, i.e. its parameter vector is a subset of the parameter vector of M_2.

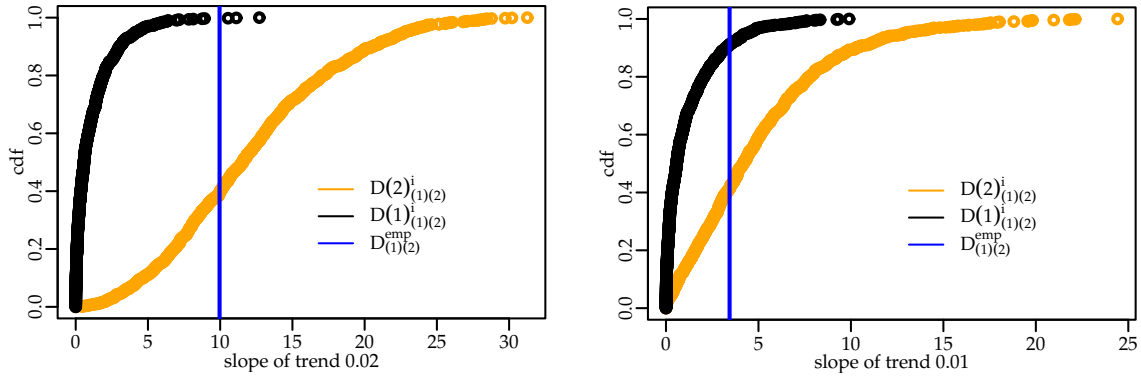


Figure 4.4: Power of the trend test based on the deviance statistic with simulated distribution. **Left**: Simulated cumulative distribution function for the deviance statistic from a stationary GPD model (model (1), black) and from a GPD model with linear trend assumption in ψ_t (model (2), orange). The artificial data is actually a realisation of model (2) and has a length of 5 415 data points with about 60 extreme events, i.e. the occurrence rate λ is 0.01. The trend present in the data has a slope $s = 0.02$. It is tested whether this artificial data are a realisation of one of the two models by comparing its deviance statistic (blue line) with the two simulated distributions. **Right**: The same situation as for the left hand side, but the trend present in the data has a slope of $s = 0.01$.

According to our setting, we expect $D^{emp}_{(1)(2)}$ to lie within the $D(2)^i_{(1)(2)}$ and not within the $D(1)^i_{(1)(2)}$, i.e. the data are generated from model (2). We expect also $D^{emp}_{(2)(3)}$ to lie within the $D(2)^i_{(2)(3)}$ and not within the $D(3)^i_{(2)(3)}$, because model (2) is the "true" model. We furthermore expect $D^{emp}_{(1)(3)}$ rather to lie within the $D(3)^i_{(1)(3)}$, because this is the over-fitted model, than within the $D(1)^i_{(1)(3)}$.

In Fig. 4.4 the results for the GPD distributed time series for the comparison of model (1) and model (2) are depicted: The "empirical" series is generated from model (2), i.e. it is generated from a non-stationary GPD model with linear trend in ψ_t and has a length of 5 415 data points with extreme events occurring with a rate $\lambda = 0.01$ (the GPD parameters are the same as in Sec. 4.6.1). On the left hand side, results for a time series with a trend slope $s = 0.02$ are depicted. Its "empirical" deviance statistic is shown as a blue line. Here the deviance statistic with a simulated distribution is able to discriminate between both models: At a level of 5% we would accept the assumption that the data are a realisation of model (2) (orange cdf) and we would reject that it is a realisation of model (1) (black cdf). In Fig. 4.4 correspondingly, less than 50 black circles cross the blue line. On the right hand side, results for a time series with a trend slope of $s = 0.01$ are depicted. In this situation, the deviance statistic cannot discriminate any more between both models at a level of 5%. That is, the deviance statistic used with a simulated distribution does not produce better results than the deviance statistic used with its theoretical distribution (see Sec. 4.6.1).

We obtain the same findings for all other cases presented in Sec. 4.6.1, also regarding the discrimination between Poisson processes: the usage of a simulated distribution does not enhance the power of the deviance statistic. We therefore apply the deviance statistic with its theoretical distribution in all our analyses of empirical data.

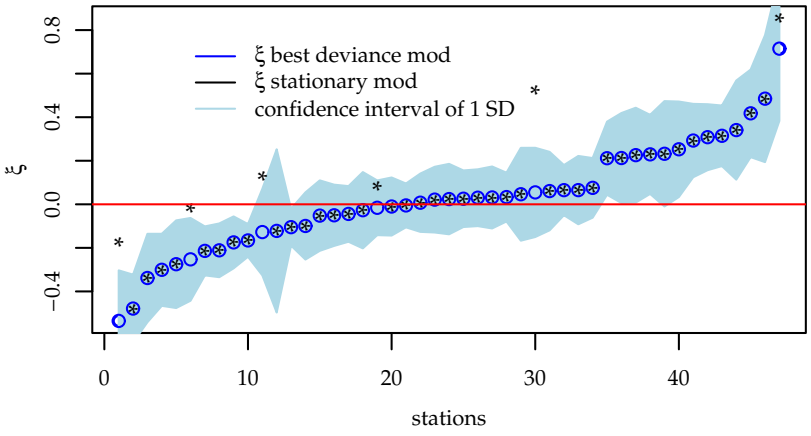


Figure 4.5: Distribution of the shape parameter ξ. Sorted estimate of the shape parameter $\hat{\xi}$ for gauges in the Danube River basin according to the best point process model (blue circles) and stationary point process model (black stars). Confidence intervals of 1 SD are depicted in light blue.

4.7 APPLICATION, RESULTS AND DISCUSSION

For this analysis, daily discharge observations of about 50 gauges in the Danube River basin in Southern Germany are investigated. The examined time period is restricted to a jointly covered period of 60 years, i.e. from 01.01.1941 to 31.12.2000. For each time series, a threshold u is chosen and the data are declustered as described in the previous sections. We investigate data of the winter season. Within this season trends are anticipated because of a changing precipitation pattern. About 5 000 data points are available for analysis purposes, which results in approximately 30 to 120 extrema per time series. As described before, a stationary and a range of non-stationary point processes are fitted to these excesses. The processes are formulated in terms of a rate λ and the shape and scale parameters ψ and ξ of the GPD. From this class of models the best one is chosen using the deviance statistic and assessment measures, such as the return level, are derived from this model. We carry out an extreme value analysis for a station only in case more than 30 extreme values are left after declustering and in case the best model passes all goodness-of-fit tests.

In Fig. 4.5 the distribution of the shape parameter estimates $\hat{\xi}$ of the best fitting point process models across the Danube River basin are depicted in sorted order (blue circles). The corresponding $\hat{\xi}$ values of the stationary model are added as black stars. The shape parameter determines the "thickness" of the upper tail of the flood frequency distribution and thus helps to characterise the properties of extreme floods. In most of the cases a stationary GPD model represents the data best, i.e. both values are identical. In the other 6 cases, the shape parameter of the stationary model is always larger than the one of the non-stationary, best fitting, model. This implies that in case only stationary models are fitted to the data, falsely a heavy tailed distribution may be assumed and therefore the probability of occurrence of severe floods may be overestimated. Furthermore, the estimated shape parameters in the region cover a wide range of values, that is from $\hat{\xi}$ in $[-0.49, 0.72]$. In about 25% of the cases the extremes exhibit an upper end point, i.e. they are Weibull distributed. About 50% of records possess shape parameter estimates and corresponding confidence intervals wich suggest compatibility with the Gumbel distribution, that is, $\xi = 0$ (red line). For about 25% of the cases we detect excesses with a heavy tailed Frechet distribution. In comparison to Gumbel distributed extremes here a fortified occurrence of extreme floods is expected.

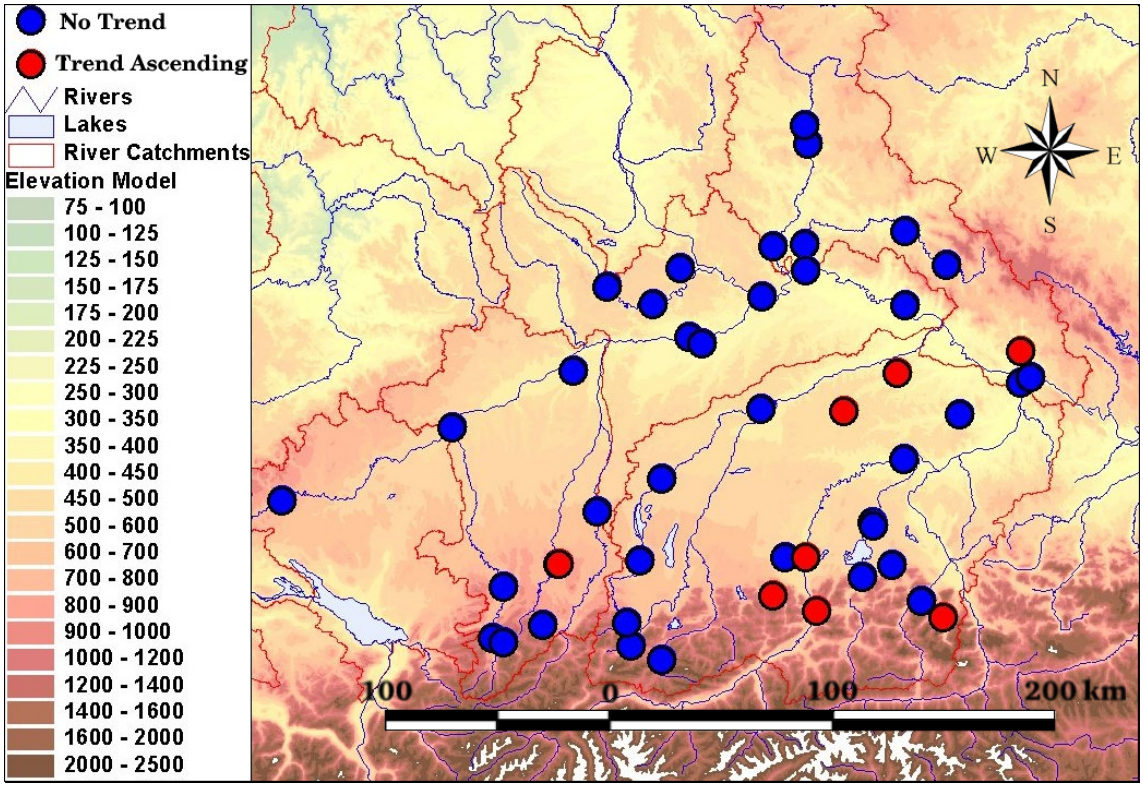


Figure 4.6: Trend in frequency of occurrence of extreme events. Trend tendency of the possibly non-stationary rate λ of the best point process model fitted to extremes of discharge data at 47 gauges of the Danube River catchment. A model with a constant rate λ is marked in blue, whereas red indicates an increasing frequency.

Using the point process approach with a possibly non-stationary rate λ, we are able to measure the change in the frequency of occurrence of extreme events. In Fig. 4.6 the trend tendency of the rate λ is depicted. In most of the cases a model with a constant λ represents the data best. These gauges are marked with blue circles. In 8 cases the frequency of occurrence increases. These gauges are the Geltnach River at Hörmannshofen, the Vils River at Rottersdorf, the Große Vils River at Vilsbiburg, the Ramsauer Ache River at Ilsank, the Inn River at Oberaudorf and Rosenheim, the Leitzach River at Stauden, and the Ilz River at Kalteneck. The findings are in line with the observed increase in frequency of heavy winter precipitation events in Bavaria and Baden-Württemberg (cf. Dietzer et al. 2001 and KLIWA 2004). They are anticipated to become even more frequent in the future (Schröter et al. 2005). More frequent winter floods are expected in Northern Europe as well (IPCC 2007).

In Fig. 4.7 the trend tendencies of the estimated mean and variance of the excesses are depicted. They are calculated using the parameter estimates of the best point process model. As apparent in Eq. (4.4), the rate λ is needed for the calculation. Therefore we here find trends at the same sites where a non-stationary rate λ is indicated in Fig. 4.6. Furthermore, in Fig. 4.7 the influence of a non-stationary GPD model, i.e. a time varying scale parameter ψ, becomes apparent. We did not find a station where the excesses are modelled best by a non-stationary point process, which is composed from an inhomogeneous Poisson process and a non-stationary GPD. Throughout, either only the Poisson process or only the GPD is non-stationary. To cross-check our procedure we model also

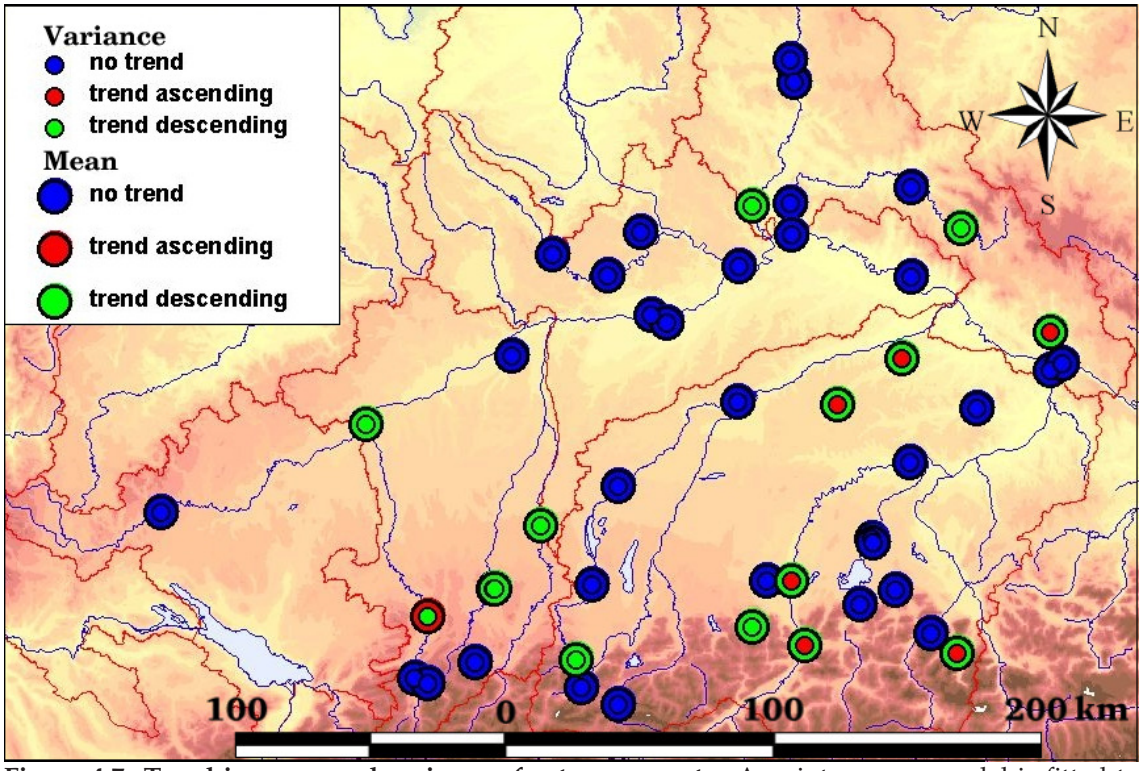


Figure 4.7: Trend in mean and variance of extreme events. A point process model is fitted to excesses of discharge data of 47 gauges in the Danube River basin. Mean (small circles) and variance (big circles) are calculated from the parameter estimates. Stationary mean and variance are depicted in blue, whereas green indicates a decreasing trend tendency and red an increasing trend tendency.

the block maxima of the data using the GEV. In the cases where both approaches suggest a stationary model as best one, the values of the parameter estimates are in the same range. However, for several stations the GEV approach suggests a non-stationary model where the point process approach does not, and vice versa.

In most of the cases, we find stationary extreme events (blue circles in Fig. 4.7). However, for about one third of the gauges a non-stationary model is necessary, which results in time dependent mean and variance of the extremes. The tendency of their trend is determined by evaluating the sign of the slope of a straight line, which is fitted to the mean estimate and the variance estimate. In some cases mean and variance have the same trend tendencies (an increasing tendency is marked in red, a decreasing one in green), then we always observe a decreasing trend tendency. In the cases where both trends are of opposite direction, we mostly observe an increasing variance and a decreasing mean. An explanation for this phenomenon may be an increase in the frequency of occurrence of extreme events, but not in magnitude. Then more outliers occur, but also much more extreme events near the threshold. This causes the estimate of the mean to be lower. The heterogeneous spatial pattern of the trends in extremes in Fig. 4.7 is not directly interpretable. We observe all three sorts of trend tendencies and no spatial accumulation becomes apparent. Maximum daily water levels during winter show spatial patterns related to topography (Pfister et al. 2004). We therefore also evaluate the elevation model (depicted at the back of Fig. 4.7), but this does not suffice to interpret the trend

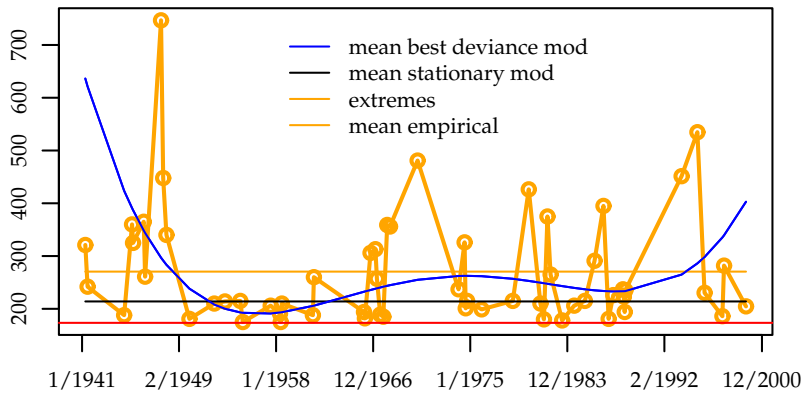


Figure 4.8: Estimates of the mean of the extremes of the Naab River at Heitzenhofen. The excesses over $u = 173\text{m}^3/\text{s}$ (red) and their mean estimate are depicted in orange. The estimates of the mean using the best fitting point process model and the stationary model are shown in blue and black, respectively.

results. So further influencing factors have to be considered. Precipitation is the main influencing factor for floods. As outlined beforehand, the observed frequency of heavy precipitation is increasing for the whole Danube River basin (KLIWA 2004). Land use change also has a strong impact on changes of the rainfall-runoff relationship and land use has undergone significant changes over the last few centuries in all European countries (cf. Pfister et al. 2004). So we would have to incorporate, e.g., information about changes of the forested area, urbanisation patterns and agricultural land use. Furthermore river regulation measures such as construction of levees or dykes or the straightening of rivers would have to be investigated. de Wit et al. (2001), for example, analyse the Meuse River and show that drainage and irrigation practices in agricultural lands lead to an increase in winter discharge. Moreover, potential discharge capacity is strongly dependent on the soil condition near the gauges and the hydraulic conditions within the river system. The attribution of signals cannot be derived unambiguously from studying time series only, so here hydrological or hydraulic models may help to assess the various contributions of environmental conditions or changes in theses conditions to changes in river regime and floods (cf. Bronstert et al. 2005). This more detailed attribution of the trends we detected in the Danube River basin is beyond the scope of this work.

Further details of our findings are exemplarily presented for the Naab River at Heitzenhofen. We conducted the following studies for all gauges, but do not depict the results here. The shape parameter estimate $\hat{\xi}$ for the Naab River at Heitzenhofen is $\hat{\xi} = -0.12$ when using the best model, which is a point process with stationary rate λ and non-stationary scale parameter ψ. For the stationary model we get $\hat{\xi} = 0.12$, both fits are compatible with a Gumble distribution when taking the confidence intervals into account. The auto-correlation function for the 59 extremes indicates compatibility with an uncorrelated series, so the declustering worked well. In Fig. 4.8 the excesses over the threshold $u = 173\text{m}^3/\text{s}$ and their mean estimate are shown in orange. For better visibility only the extrema are depicted, that is the time axis is compressed. The mean is also estimated as given in Eq. (4.4) using the parameter estimates of the stationary model (black) and the best model (blue). Apparently the estimates of the non-stationary model are influenced by the high magnitude of the excesses before 1949, which results in a decreasing trend tendency for the estimated scale parameter ψ and the estimates of the mean (depicted in Fig. 4.8) and the variance. This trend tendency is determined by the sign of the slope of a

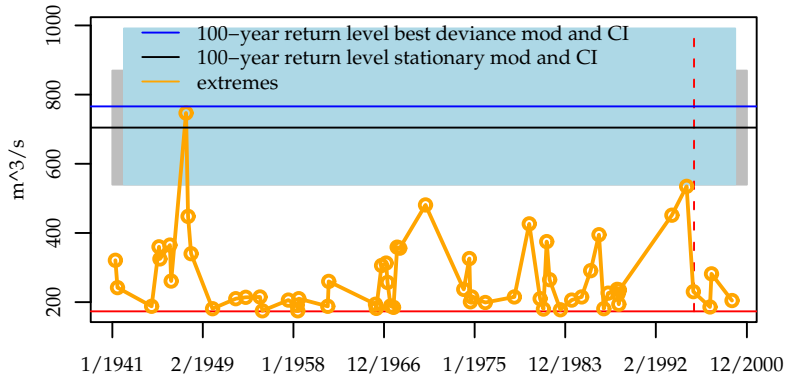


Figure 4.9: 100-year return level for the Naab River at Heitzenhofen. The excesses over $u = 173\text{m}^3/\text{s}$ (red line) are depicted in orange. The corresponding 100-year return levels of the best point process model and the stationary model are shown in blue and black, respectively. 68% confidence intervals are calculated using the delta method, see light blue and grey areas.

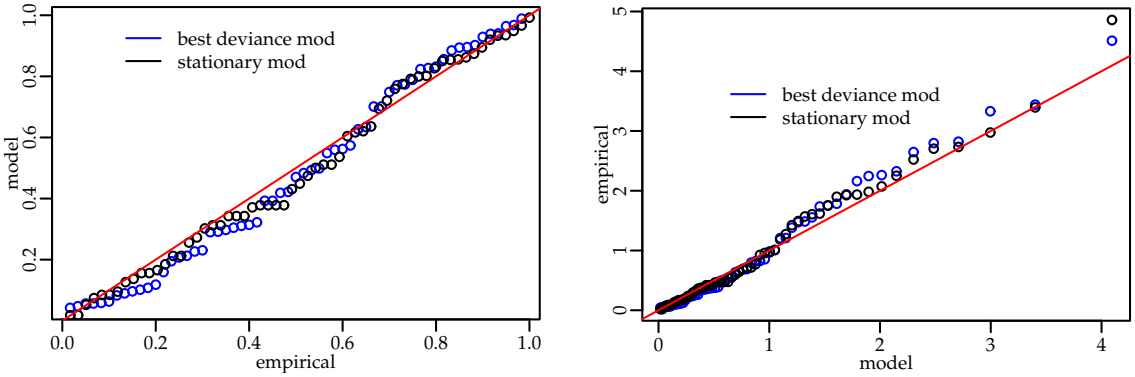


Figure 4.10: pp-plot and qq-plot for the Naab River at Heitzenhofen. Left: Residual probability plot for the best fitting point process model (blue) and the stationary point process model (black). **Right**: Residual quantile plot (exponential scale) for the best fitting point process model (blue) and the stationary point process model (black).

straight line, which is fitted to the estimates.

The parameter estimates of the non-stationary GPD model seem to become less credible at the edges. Therefore we calculate all return levels at $t_0 = 01.01.1996$, that is five years before the end of the analysed period. For the Naab River at Heitzenhofen we get an increase in the exceedance probability of 85%, that is the 100-year return level of the stationary model is expected to be exceeded nearly twice every 100 years (this is indicated by a red circle in Fig. 4.11). In Fig. 4.9 the return level for the stationary model (black) and the best model (blue) are shown for the Naab river at Heitzenhofen. With $765.87\text{m}^3/\text{s}$ the return level of the best model is about $60\text{m}^3/\text{s}$ higher than the return level which would be assumed in the stationary case. The corresponding confidence intervals are calculated using the delta method (see Sec. 2.3.4).

To evaluate the validity of the best model, pp-plots and qq-plots (see Sec. 2.6.4) are created for each station. In Fig. 4.10 the results for the Naab River at Heitzenhofen are exemplarily depicted. The stationary model (black) shows good results, though it is outperformed by the best model. This becomes especially visible in the qq-plot. Besides this visual tool, two goodness-of-fit tests are used to validate the quality of the best model.

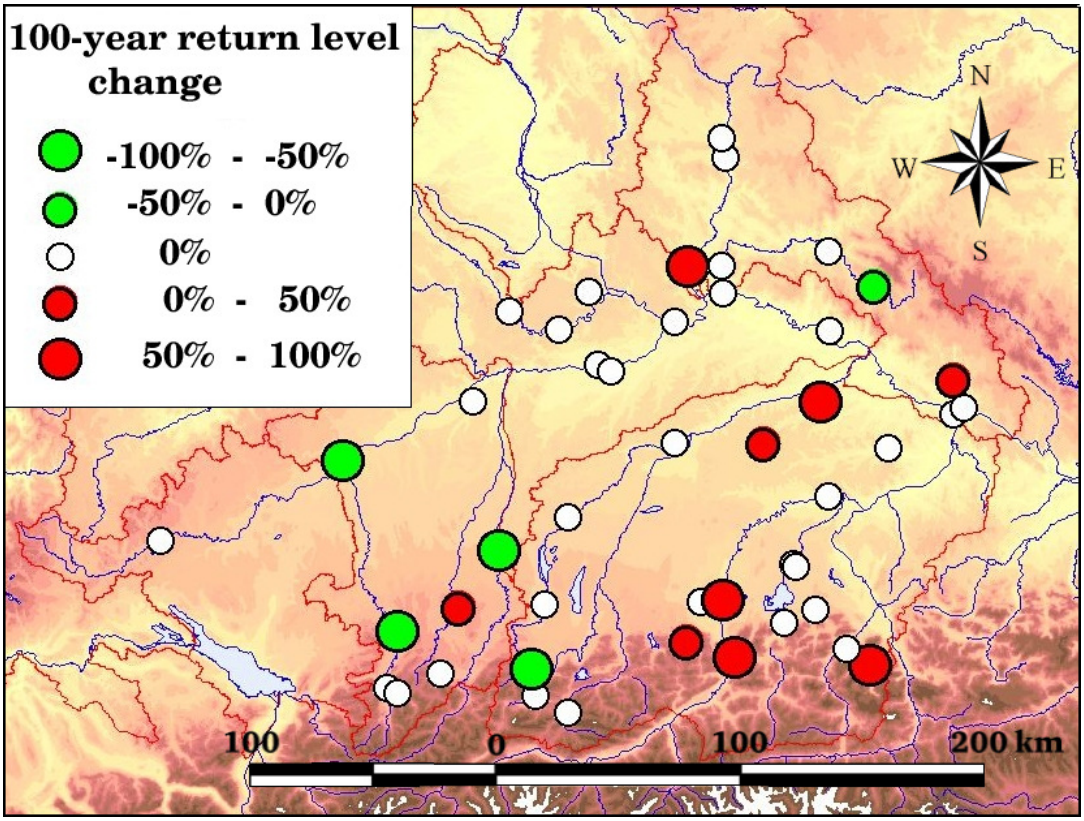


Figure 4.11: Change of probability of excessing a 100-year return level. The 100-year return level, is calculated according to a stationary point process model which is fitted to the extremes. Then we calculate p^*, the probability of exceeding this 100-year return level, using the parameter estimates of the best fitting, possibly non-stationary point process model at time point $t_0 = 01.01.1996$. The corresponding changes of probability are depicted. White indicates no change, i.e. the best fitting model is a stationary one. A decreasing probability is marked in green and an increasing one in red.

For the GPD model a Kolmogorov-Smirnov test is used (cf. Sec. 2.6.3). To evaluate the inhomogeneous Poisson process, properties of the Poisson process are tested (cf. Sec. 2.6.2). Regarding the Poisson process, the best model passes the goodness-of-fit test for all stations examined. The GPD does not fit to the excesses of three stations, namely the Danube River at Donauwörth, the Isar River at Plattling and the Weiße Traun River at Siegsdorf. Therefore these stations are excluded from our analysis. Interestingly for these stations, not only the best GPD model does not pass the Kolmogorov-Smirnov test, but the stationary GPD models also fail. Summing up, our models are capable to represent nearly all empirical data we look at, which is a quite good result.

Finally, to assess the degree of change which is induced when incorporating non-stationary models in the extreme value analysis, we examine the influence on the return level estimation. For this purpose the 100-year return level, which is crossed once a year with probability $p = 0.01$, is calculated for each station using a stationary point process model. Then we calculate the probability of exceedance p^* for the corresponding return level z_p for the best, possibly non-stationary, model. In Fig 4.11 the difference between p and p^* is depicted for stations in the Danube River basin. White circles indicate no change implying that the stationary model fits best. A larger probability of exceedance is marked

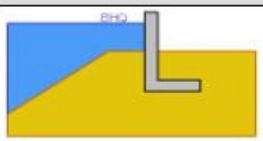
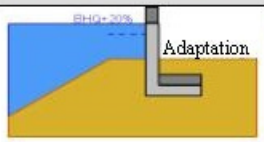
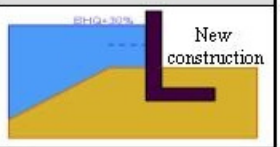
Scenario	Design flood value = HQ_{100}	DFV +20% = +0,29 m	DFV +30% = +0,44 m
Protection strategy: **Flood wall**			
Costs in case of reactive adaptation	**100 %**	**155 %**	**257 %**
Costs in case of proactive adaptation	**100 %**	**110 %**	**113 %**

Figure 4.12: Flood changes and flood protection strategies. Costs of proactive and reactive adaptation to a change of the 100-year return level (*source: KLIWA (2004), pages 184-185*).

with red circles and $p^* < p$ is indicated by green circles. In this context a change of 100% denotes $p^* = 0.02$, that is z_p is expected to be exceeded twice every 100 years when using the best suiting model. As outlined in Sec. 2.3.4, a return level has to be calculated for a certain time point t_0 in case parameter estimates of a non-stationary model are used. We here choose $t_0 = 01.01.1996$.

As shown in Fig 4.11, we observe severe changes in the probability of exceedance in case non-stationarity in the extremes is considered. The occurrence of non-stationary river discharge extremes in the Danube River basin and the corresponding change of flood risk is already recognised (KLIWA 2004). Nevertheless, up to now the common methodological canon of water management authorities does not incorporate the systematic evaluation of this kind of change. We find that the probability of exceedance gets lower in some cases, most of them are situated in the South-Eastern part of the Danube River basin. The stations with a decreasing probability of exceedance are the Iller River at Kempten and Wiblingen, the Lech River at Landsberg, the Schwarzer Regen River at Teisnach and the Ammer River at Oberammergau. For 9 of the 47 assessed stations the probability of exceedance gets higher, in some cases more than 50%. Those gauges are the Geltnach River at Hörmannshofen, the Naab River at Heitzenhofen, the Vils River at Rottersdorf, the Große Vils River at Vilsbiburg, the Ramsauer Ache River at Ilsank, the Inn River at Oberaudorf and Rosenheim, the Leitzach River at Stauden, and the Ilz River at Kalteneck. In Fig. 4.12 the adaptation costs which correspond to an increase in flood risk are depicted exemplarily for flood control walls and dykes in Baden-Württemberg (cf. KLIWA 2004). These costs do not change linearly with an increase in flood risk. Thus, in case the 100-year return level (HQ100) rises 20%, additional adaptation costs of 55% are implied. In case the rise of the return level is already considered in the planning phase, additional costs of 10% become due. The figures get even worse in case the 100-year return level increases 30%. Thus, a sound extreme value analysis framework which allows for non-stationarity is a necessary basis for an adequate water management cost calculation.

CHAPTER 5

CONCLUSION AND PERSPECTIVES

We motivated our thesis by the search for trends in mean values and extreme values of hydro-meteorological data. Furthermore we wanted to assess the impact of these non-stationarities on common assessment measures.

To track these questions we assessed trends within the Neckar River and Danube River basin in Southern Germany. We detected significant trends for up to one half of the mean value records and one third of the extreme value records. Furthermore, we found interdependences between trends and auto-correlation in the data. We showed that these phenomena have an impact on the calculation of assessment measures commonly used by water management authorities such as trend test statistics and return levels. The results presented here therefore particularly highlight the necessity of a sound trend and extreme value assessment framework for hydro-meteorological time series, which considers non-stationarity and the auto-correlation structure (memory) in the data. By simulation studies we investigated the potentials and limits of the approaches provided here regarding the properties of empirical data, and therefore demonstrated their value for a broad range of applications.

The results of this thesis furthermore suggest that we have to proceed gauge by gauge. We found specific, site related structures for all statistical characteristics assessed. We did not detect immediately interpretable spatial trend patterns, neither for mean nor for extreme values. We often observed neighbouring gauges to behave distinctively. These results show that the impact of climate change is not directly detectable in river discharge yet, but is blurred by other factors. Evidence for trends in precipitation, for example, is globally stronger than for changes in flooding (Robson 2002 and Kundzewicz et al. 2004). Possible factors, which are not necessarily in tune with gross climate-related drivers, may be land use, soil conditions, weather regimes, elevation, moisture or catchment size. The extreme value and trend assessment framework presented here is a necessary basis for a physical interpretation of the trends. Besides, the statistical characteristics of river discharge obtained in this thesis may be used to validate the output of, e.g., hydrological models.

5.1 TRENDS IN MEAN VALUES

One part of this thesis is devoted to the analysis of trends in mean river discharge. We analysed monthly mean values and monthly maxima of daily river discharge. Most of these records cover the time period between 1941 and 2000. Regarding the Neckar

River basin, we furthermore examined the summer and winter seasons. We found non-stationarity and auto-correlation being common phenomena in river run-off. Both statistical characteristics are interrelated, which calls for the consideration of the memory in the data when assessing trends. We achieved this by applying and enhancing the Craigmile-Guttorp-Percival (CGP) trend test (Craigmile et al. 2004). Thereby the time series is modelled by an additive combination of a stochastic component and potentially a deterministic trend. We extended the test by allowing for short- and long-term correlation in the stochastic part by using FARIMA models. This enabled us to apply the test to a broad range of empirical time series. We regard trends as fluctuations on large scales. Our results suggest that river discharge data often possesses rather complicated trend shapes with large oscillations. Therefore trend tests which only check linear trend assumptions seem to be inappropriate for this kind of time series. We found a variety of different trend shapes, and typical shapes could not be extracted. A significant trend was detected for about 10% of the analysed 10 gauges in the Neckar River basin and about 45% of the 90 gauges within the Danube River basin. Most of the significant trends are increasing. However, the spatially heterogeneous patterns of increasing and decreasing trends could not be interpreted directly. This is in line with findings of other authors, e.g. Kundzewicz and Schellnhuber (2004), who investigate river discharge worldwide. We compared the results obtained using the CGP trend test with the outcome of the Mann-Kendall trend test (Mann 1945), which is very common in hydrology and does not explicitly consider auto-correlations. For many stations in the Danube River catchment the results differ: The Mann-Kendall trend test does detect a significant trend where the CGP trend test does not. We conclude that this is partly due to the memory being present in the data, and that therefore the usage of the CGP trend test provides the more reliable results.

The natural variability in the data has been modelled using stochastic FARIMA(p, δ, q) processes up to an order of FARIMA(4,δ,3). We found short- and long-term correlation being present in hydro-meteorological data. Both types of correlation have a distinct impact on the uncertainty of statistical estimates such as trend slopes, which underlines the necessity to model them. River discharge possesses strong auto-correlations for periods of one day to two weeks, for example because of weather or soil conditions. Long-range correlation may occur because river networks aggregate precipitation and convert it into long-memory run-off (see Mudelsee 2007). FARIMA(p,δ,q) processes only reflect the linear auto-correlation structure being present in the data. Further enhancements are possible, e.g. GARCH models, which consider also dependence in the variance of the data. However, by the use of goodness-of-fit evaluation procedures, we found that our stochastic models adequately represent most of the empirical data we examined. We detected a spatially heterogeneous auto-correlation structure in the assessed region. Regarding the Danube river basin, gauges close to the Alpes show less memory than stations close to the Danube River. The latter possess especially often long-memory. This may be due to the fact that persistence is stronger for large catchments, where the time series comprises a lot of processes. Although we did not find a predominant correlation pattern in the region, in most cases the short-term correlated AR(1) process is the most suitable model. The long-term correlated FD(δ) process often suffices to represent the series as well. Both models are the most simple representatives of their classes. We therefore have strong evidence that simple models seem to represent the dynamics of the data quite well. The distinction between trend and auto-correlation structure was not possible in some cases, as we have demonstrated for a few records in the Neckar River basin. Due to the shortness of the time series two alternative models represent the data well, e.g. a long-range

correlated model and a short-term correlated model with a deterministic trend component. In these cases, further physically driven information is needed to accurately decide which model to prefer.

Regarding methodological aspects, we chose to estimate the trend by using wavelets, that is, data-driven. Advantageously, we thus do not need a priori assumptions about the trend shape. A potential drawback is that the selection of the wavelet and the degree of smoothing to obtain the trend estimate are not fully objective choices. The idea of trends as fluctuations on large scales allows to address diverse questions related to trend detection. Significant trends, for example, might be examined for oscillations they all share. This serves to assess the relation between discharge and processes acting on decadal scales, such as atmospheric patterns influenced by El Niño. The wavelet trend estimate is consistent, as outlined in this work, and the methodology allows for the assessment of uncertainty by providing confidence intervals for the trend estimate. Furthermore, trend and auto-correlation structure are estimated separately in our approach. On the one hand, this might be a drawback, because the trend is estimated without considering the auto-correlation structure. This might bias the results. On the other hand, a potential bias is avoided which may occur due to the mutual influence of trend and auto-correlation parameters during the estimation process. Moreover, the approach presented here allows for more complicated structures of trend and memory than in case trend and auto-correlation are jointly estimated.

The trend test is conducted by comparing the variance of the time series with and without the estimated trend. On the basis of simulation studies we showed that this trend test possesses sufficient power for various settings. In case a linear trend is added to the data, the power of the CGP trend test is comparable to the one of a standard linear regression test. Two of the key findings of this thesis are that the width of the confidence intervals for the trend estimate as well as the acceptance or rejection of a trend depends on the chosen stochastic model to represent the natural variability. The intercomparison of trend test results for a specific time series under the assumption of different stochastic models shows that a significant trend is less often detected under the assumption of long-range correlated data than under the assumption of short-range correlated or independent data. The reason for this is that long-term correlated data causes long excursions from the mean as well as local trends (Beran 1994 and Kallache et al. 2005). These findings imply that trend tests which do not account for auto-correlation in the data are likely to give unreliable results for auto-correlated data (cf. Yue and Wang 2002).

We conclude that our approach to assess trends in mean values, i.e. the modified CGP trend test, is well applicable for empirical hydro-meteorological records of the length of 60 to 100 years of daily measurement. Trends in mean values are a common phenomenon in river discharge records of Southern Germany. The proposed consideration of the auto-correlation structure in the trend test procedure is crucial to obtain reliable results. Furthermore, a basis for uncertainty analysis is given by the provision of confidence intervals for the trend estimate.

5.2 TRENDS IN EXTREME VALUES

In the second part of this thesis, we developed a methodological framework to assess extremes of hydro-meteorological data which accounts for non-stationarity and auto-

correlation. An unjustified assumption of stationarity may lead to considerable underestimation of the probability of a disastrous extreme event (see, e.g., Coles et al. 2002). In our approach we take the skewed distribution of the extremes into account and apply the Fisher-Tippett theorem. By doing so, we are able to examine independent observations with almost arbitrary distribution: The convergence of their extreme values towards the generalized Pareto distribution (GPD) is theoretically well founded. The extremes are derived as excesses over a threshold and are modelled using a point process (cf. Coles 2001). This allows for a sound modelling of the frequency of occurrence and the magnitude of the extreme values. The auto-correlation structure in the extremes is eliminated by thinning out the data, and non-stationarity is incorporated by allowing for non-stationary model parameters. The most suitable model is then selected out of a class of models which comprises the stationary model and models with a variety of polynomial and exponential trend assumptions. Thus, trend detection is transfered to a model selection problem.

Regarding methodological aspects of our extreme value assessment framework, we evaluate its adequacy and its power to detect trends by simulation studies. The point process representing the extreme events is a combination of two components: An inhomogeneous Poisson process with possibly non-stationary rate of occurrence to model the frequency of occurrence of the extreme events, and the GPD with a possibly non-stationary scale parameter to derive the magnitude of these extreme events. From our empirical data we obtain sets of extreme values with a size ranging from 30 to 120. We find that this is the minimum size to get reliable results. Very good results are obtained for estimating and simulating the frequency of occurrence of extreme events and testing for a trend in the rate of occurrence. We also reliably detect the strong trends in the magnitude of the extremes. However, to be able to detect even weak trends in the magnitude as well, there should be rather 200 extrema at hand. We therefore assume that we did not find every weak trend being present in the Danube River basin. To cross-check our procedure, we compared it to results obtained by using the generalized extreme value (GEV) distribution. Here the extremes are derived by drawing maxima from intervals of equal length and we chose the standard interval length of 365 days. In those cases where both approaches suggest a stationary model as best suiting one, we find good agreement between both approaches. However, they differ regarding the significance of trends: Often the GEV approach suggests a non-stationary model where the point process approach does not and vice versa. We consider the point process framework to be more reliable, because we found that the set of threshold excesses contains the more extreme events of our empirical data and therefore better represents the extremes.

We analysed extremes of daily discharge measurements of about 50 stations within the Danube River basin. Thereby we focused on the winter season and analysed the jointly covered time period between 1941 and 2000. Basic assumptions of standard extreme value theory such as independent and identically distributed extremes are often not fulfilled by hydrological data. Discharge is influenced by complex dynamical processes (precipitation, snow accumulation and snowmelt, evapotranspiration, et cetera), therefore a univariate distribution with two or three parameters is not necessarily a good approximation of its multifaced nature (Engeland et al. 2004). Indeed, we had to remove auto-correlations in all discharge records of the Danube River basin and one third of the stations exhibit also non-stationary extremes. Importantly, the estimates of the extreme value distribution parameters may be biased in case a stationary model is used to represent non-stationary extremes. The magnitude of this bias depends on the trend strength.

By using the framework proposed in this thesis, we are able to augment the number of empirical records which can be adequately analysed.

Concerning the characteristics of the extreme values of the Danube River basin, we give particular attention to the estimated shape parameter of the GPD. The shape parameter determines the "thickness" of the upper tail of the flood frequency distribution and thus helps to characterise the properties of extreme floods. Dependent on the magnitude of this shape parameter, the extreme value distribution can be distinguished into three types. We found all three of them being present in the Danube River catchment. About 25% of the gauges possess a Weibull distribution, i.e., a finite endpoint can be determined. About 50% of the records are compatible with being Gumbel distributed, and the region exhibits 25% discharge records with heavy-tailed extrema. This Frechet distribution deserves special interest, because arbitrarily high and more frequent outliers than in the Gumble case have to be expected here. We found a heterogeneous spatial distribution of the different extreme value distributions. The estimate of the shape parameter was always lower using the most suitable, non-stationary point process model than using the stationary model. This supports the theoretical finding that a mixture of Gumbel distributions (that is the shape parameter is compatible with being zero) is capable to mimic a heavy tailed distribution (here the shape parameter is larger than zero). We conclude that in case a stationary model is solely fitted to the data, heavy-tailed distributions may falsely be detected. This may result in the estimation of too high return levels and therefore too high construction costs of flood protection buildings.

Regarding the trend tendencies, the frequency of occurrence of extreme events always increases when it is detected to be non-stationary. In case a change in the magnitude of the extreme events is found, we observe decreasing and increasing tendencies. Altogether about one third of the 47 analysed records of the Danube River basin exhibit non-stationarity. To demonstrate the relevance of the changes which arise when allowing for non-stationary extreme value models, we exemplarily chose return levels. They are an important assessment measure for water management authorities. We identified changes up to 100% for the probability of exceedance of the 100-year return level. This implies a potential doubling of the damage costs, because the 100-year return level is then actually expected to be exceeded twice within 100 years. The spatial pattern of the trend in extremes is not immediately interpretable. Nevertheless, most of the detected trend tendencies are increasing, which might justify the activity of the Bavarian water management authorities, who anticipate an increase in flood magnitude due to climate change and therefore introduced a climate change factor of 1.15 in 2006, that is, every design value is increased by 15% (BlfW 2006) .

Hydrological design which takes non-stationary conditions into account is a direct consequence of accepting the idea of environmental change. As the approach we provide is fully parametric, in principle it provides the possibility to extrapolate the trends into the future. However, in some cases the estimated trends were found to deviate substantially at the beginning and the end of the time series. Therefore, we decided to proceed cautiously. We considered trends within the measurement period, but we did not extrapolate them. This is in line with other authors (cf. Bárdossy and Filiz 2005).

Our methodology provides several ways to supply uncertainty intervals along with each parameter estimate and assessment measure, namely the delta method, the profile likelihood or bootstrapping. The resulting confidence intervals show that the examina-

tion of uncertainty is a crucial prerequisite for the evaluation and interpretation of the results of extreme value analysis.

We conclude that the application of an extended extreme value analysis framework, as presented in this thesis, is necessary to adequately assess non-stationary extreme values. We demonstrated the usability, the potential and the limits of our framework. Trends in frequency and magnitude of extremes of river discharge are anticipated because of climate change and we already find a noteworthy fraction of the empirical records analysed exhibiting non-stationary extremes.

5.3 FINAL REMARKS AND OUTLOOK

No universal trend tendency could be found in discharge records of river basins in Southern Germany. This does not imply that climate change has no impact on the hydrological cycle. Increases in rainfall and rainfall extremes seem possible because of climate change, and the link between rainfall and flooding suggests an impact on floods and droughts. We conclude that the possibility of more frequent or heavier floods calls for the use of approaches which allow for non-stationarity and assess the uncertainty related to these events. By doing so, common practice of water management authorities can be improved and costs for flood protection buildings can be calculated with higher accuracy. Thus, e.g., construction costs are reduced in case return levels estimates are overestimated and possible future damage costs are dampened in case it is shown that those estimates are too low. The utility of the frameworks provided here is appreciated by, e.g., Bavarian water management authorities. This thesis has been written in close cooperation with them and the incorporation of our approaches in their methodological canon is on the way.

Regarding possible continuations of our work, the physical or process-oriented interpretation of the results obtained in this thesis is one way to go. Thereby supplementary data about soil conditions, land use or vegetation is needed and information about the interrelation of those factors and river discharge. In this context, we can think of several ways to extend our methodology. The extreme value assessment framework can be easily extended to incorporate influential variables in the form of covariates (cf. Coles 2001). It would then be possible to incorporate the influence of globally acting factors other than climate change such as land use changes or soil degradation. The incorporation of measures of weather regimes and the El Niño phenomenon or other large-scale atmosphere-ocean circulation patterns is of high interest as well (see, e.g., Mudelsee et al. 2004 and Katz et al. 2002).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *2nd International Symposium on Information Theory*, Budapest, pp. 267–281. Akademie Kiade.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika 26*, 237–242.

Allen, M. R. and W. J. Ingram (2002). Constraints on future changes in climate and the hydrological cycle. *Nature 419*, 224–232.

Apel, H., A. H. Thieken, B. Merz, and G. Blöschl (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences 4*, 295–308.

Baettig, M. B., M. Wild, and D. M. Imboden (2007). A climate change index: Where climate change may be most prominent in the 21st century. *Geophysical Research Letters 34*, L01705.

Bárdossy, A. and F. Filiz (2005). Identification of flood producing atmospheric circulation patterns. *Journal of Hydrology 313*, 48–57.

Bárdossy, A. and S. Pakosch (2005). Wahrscheinlichkeiten extremer Hochwasser unter sich ändernden Klimaverhältnissen. *Wasserwirtschaft 7-8*, 58–62.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley & Sons.

Beran, J. (1992). A goodness-of-fit test for time series with long range dependence. *Journal of the Royal Statistical Society B 54*, 749–760.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall.

Beran, J., R. J. Bhansali, and D. Ocker (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika 85*(4), 921–934.

Beran, J. and S. Ghosh (1998). Root-n-consistent estimation in partial linear models with long-memory errors. *Scandinavian Journal of Statistics 25*, 345–357.

Bhattacharya, R. N., V. K. Gupta, and E. Waymire (1983). The Hurst effect under trends. *Journal of Applied Probability 20*, 649–662.

Bîrsan, M.-V., P. Molnar, and P. Burlando (2003). Streamflow trends in Switzerland. In V. R. Thorndycraft, G. Benito, M. Barriendos, and M. C. Llasat (Eds.), *Paleofloods, Historical Floods and Climatic Variability: Applications in Flood Risk*, Barcelona. PHEFRA.

Bisaglia, L. (2002). Model selection for long-memory models. *Quaderni di Statistica 4*, 33–49.

Black, A. R. (1995). Flooding and increased flood frequency in Scotland since 1988. *Physics and Chemistry of the Earth 20*, 463–468.

Blatter, C. (1998). *Wavelets – eine Einführung*. Braunschweig/Wiesbaden: Viehweg Verlag.

BlfW (2006). Bayrisches Landesamt für Wasserwirtschaft: Klimaänderungsfaktoren bei Planungen für den Hochwasserschutz. Water Management Directive.

Bloomfield, P. (1992). Trends in global temperature. *Climatic Change 21*, 1–16.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. New Jersey: Prentice Hall.

Brázdil, R., R. Glaser, C. Pfister, P. Dobrovolný, J.-M. Antoine, M. Barriendos, D. Camuffo, M. Deutsch, S. Enzi, E. Guidoboni, O. Kotyza, and F. S. Rodrigo (1999). Flood events of selected European rivers in the sixteenth century. *Climatic Change 43*, 239–285.

Brázdil, R., C. Pfister, H. Wanner, H. von Storch, and J. Luterbacher (2005). Historical climatology in Europe – the state of the art. *Climatic Change 70*, 363–430.

Bremicker, M. (2000). *Das Wasserhaushaltsmodell LARSIM – Modellgrundlagen und Anwendungsbeispiele*, Volume 11 of *Freiburger Schriften zur Hydrologie*. Institut für Hydrologie der Universität Freiburg.

Bronstert, A., J. Carrera, P. Kabat, and S. Lütkemeier (Eds.) (2005). *Coupled Models for the Hydrologcial Cycle*. Berlin: Springer.

Bunde, A., S. Havlin, E. Koscielny-Bunde, and H.-J. Schellnhuber (2002). *Atmospheric Persistence Analysis: Novel Approaches and Applications*. Berlin: Springer.

Burlando, P. and R. Rosso (2002). Effects of transient climate change on basin hydrology. 2. Impacts on runoff variability in the Arno river, central Italy. *Hydrological Processes 16*, 1177–1199.

Casella, G. and R. L. Berger (2002). *Statistical Inference*. Pacific Grove: Duxbury Advanced Series.

Caspary, H. J. and A. Bárdossy (1995). Markieren die Winterhochwasser 1990 und 1993 das Ende der Stationarität in der Hochwasserhydrologie infolge von Klimaänderungen? *Wasser & Boden 47*(3), 18–24.

Chan, N. H. and W. Palma (1998). State space modeling of long-memory processes. *Annals of Statistics 26*, 719–740.

Chavez-Demoulin, V. (1999). *Two problems in environmental statistics: Capture-recapture analysis and smooth extremal models.* Ph. D. thesis, Ecole Polytechnique de Lausanne.

Chavez-Demoulin, V. and P. Embrechts (2004). Smooth extremal models in finance and insurance. *The Journal of Risk and Insurance 71*(2), 183–199.

Claussen, M. (1996). Variability of global biome patterns as a function of initial and boundary conditions in a climate model. *Climate Dynamics 12*, 371–379.

Cohn, T. A. and H. F. Lins (2005). Nature's style: Naturally trendy. *Geophysical Research Letters 32*(23), L23402.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Berlin: Springer.

Coles, S. (2002). The use and misuse of extreme value models in practice. Technical report, Department of Mathematics, University of Bristol, BS8 1TW, U.K., `http://www.stats.bris.ac.uk/~masgc/papers/semstat.ps`.

Coles, S., L. R. Pericchi, and S. Sisson (2002). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology 273*, 35–50.

Coles, S. G. and M. J. Dixon (1999). Likelihood-based inference for extreme value models. *Extremes 2*, 5–23.

Cox, D. and D. Hinkley (1994). *Theoretical Statistics*. London: Chapman & Hall.

Cox, D. R., V. S. Isham, and P. J. Northrop (2002). Floods: Some probabilistic and statistical approaches. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences 360*, 1389–1408.

Cox, D. R. and P. A. W. Lewis (1966). *The Statistical Analysis of Time Series Events*. London: Methuen.

Craigmile, P. F., P. Guttorp, and D. B. Percival (2004). Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics 15*, 313–335.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics 17*, 1749–1766.

Daley, D. J. and R. Vesilo (1997). Long range dependence of point processes, with queueing examples. *Stochastic Processes and their Applications 70*, 265–282.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

Davison, A. C. (2003). *Statistical Models*. Cambridge: Cambridge University Press.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.

Davison, A. C. and N. I. Ramesh (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society B 62*(1), 191–208.

de Wit, M., P. Warmerdam, P. Torfs, R. Uijlenhoet, E. Roulin, A. Cheymol, W. van Deursen, P. van Walsum, M. Ververs, J. Kwadijk, and H. Buiteveld (2001). Effect of climate change on the hydrology of the river Meuse. Technical Report 104, Wageningen University Environmental Sciences.

DeGroot, M. H. (1984). *Probability and Statistics*. Reading: Addison Wesley.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Berlin: Springer.

DFO (2004). Dartmouth Flood Observatory. Technical report. Data available at: `http://www.dartmouth.edu/~floods/`.

Diebold, F. X. and A. Inoue (2001). Long memory and structural change. *Journal of Econometrics 105*, 131–159.

Dietzer, B., T. Günther, A. Klämt, H. Matthäus, and T. Reich (2001). Langzeitverhalten hydrometeorologischer Größen. Technical report, DWD. Klimastatusbericht.

Douglas, E. M., R. M. Vogel, and C. N. Kroll (2000). Trends in floods and low floods in the United States: Impact of spatial correlation. *Journal of Hydrology 240*, 90–105.

Dutilleul, P. and P. Legendre (1992). Lack of robustness in two tests of normality against autocorrelation in sample data. *Journal of Statistical Computation and Simulation 42*, 79–91.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events*. Berlin: Springer.

Engeland, K., H. Hisdal, and A. Frigessi (2004). Practical extreme value modelling of hydrological floods and droughts: A case study. *Extremes 7*, 5–30.

Ferro, C. A. T. (2003). *Statistical methods for clusters of extreme values*. Ph. D. thesis, University of Lancaster, Great Britain.

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 Phil. Soc. 24*, 180–190.

Fleming, S. W. and G. K. C. Clarke (2002). Autoregressive noise, deserialization, and trend detection and quantification in annual river discharge time series. *Canadian Water Resources Journal 27*(3), 335–354.

FOE (1997). Friends of the Earth International: What's happening with the weather? Climate change and weather extremes. `http://www.foe.co.uk/pubsinfo/briefings/html/19971211173223.html`.

Frei, C., H. C. Davies, J. Gurtz, and C. Schär (2000). Climate dynamics and extreme precipitation and flood events in Central Europe. *Integrated Assessment 1*, 281–299.

Fuentes, M., P. Guttorp, and P. Challenor (2003). Statistical assessment of numerical models. *International Statistical Review 71*, 201–221.

Gaetan, C. and M. Grigoletto (2004). Smoothing sample extremes with dynamic models. *Extremes 7*, 221–236.

Giraitis, L., P. Kokoszka, and R. Leipus (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability 38*, 1033–1054.

Giraitis, L. and P. M. Robinson (2001). Parameter estimation under long range dependence. Discussion paper, London School of Economics and Political Science.

Glaser, R. and H. Stangl (2003). Historical floods in the Dutch Rhine Delta. *Natural Hazards and Earth System Sciences 3*, 1–9.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics 44*, 423–453.

Goldstein, M. and J. C. Rougier (2006). Reified Bayesian modelling and inference for physical systems. Journal of Statistical Planning and Inference.

Goupillaud, P., A. Grossman, and J. Morlet (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration 23*, 85–102.

Gradshteyn, I. S. and I. M. Ryzhik (1965). *Tables of Integrals, Series and Products*. London: Academic Press.

Granger, C. W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differences. *Journal of Time Series Analysis 1*(1), 15–29.

Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. London: Chapmann and Hall/CRC.

Haase, D. and H. Nuissl (2007). Does urban sprawl drive changes in the water balance and policy? The case of Leipzig (Germany) 1870-2003. *Landscape and Urban Planning 80*, 1–13.

Hall, P. and N. Tajvidi (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science 15*(2), 153–167.

Hamed, K. H. and A. R. Rao (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology 204*, 182–196.

Hansen, J. and S. Lebedeff (1987). Global trends of measured air surface temperature. *Journal of Geophysical Research 92*, 13345–13372.

Hansen, J. and S. Lebedeff (1988). Global surface air temperatures: Update through 1987. *Geophysical Research Letters 15*, 323–326.

Harvey, A. C. (1993). *Time Series Models*. New York: Harvester Wheadsheaf.

Hassler, U. and J. Wolters (1994). On the power of unit root tests against fractional alternatives. *Economics Letters 45 45*(1), 1–5.

Hinde, J. (1992). *Lecture Notes in Statistics*, Volume 78, Chapter Choosing between non-nested models: A simulation approach, pp. 119–124. New York: Springer.

Hipel, K. W. and A. I. McLeod (1994). *Time Series Modeling of Water Resources and Environmental Systems*. Amsterdam: Elsevier Science.

Hsing, T., J. Hüsler, and M. R. Leadbetter (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields 78*, 97–112.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers 116*, 770–799.

Hurst, H. E., R. P. Black, and Y. M. Simaika (1965). *Long-Term Storage: An Experimental Study*. London: Constable.

IPCC (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

IPCC (2007). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. London [in press]: Camebridge University Press.

Jacobeit, J., R. Glaser, J. Luterbacher, and H. Wanner (2003). Links between flood events in central Europe since AD 1500 and large-scale atmospheric circulation modes. *Geophysical Research Letters 30*(4), 1172.

Jones, R. N., F. H. S. Chiew, W. C. Boughton, and L. Zhang (2006). Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models. *Advances in Water Resources 29*(10), 1419–1429.

Kaiser, G. (1994). *A Friendly Guide to Wavelets*. Boston: Birkhäuser.

Kallache, M., H. W. Rust, and J. Kropp (2005). Trend assessment: Applications for hydrology and climate. *Nonlinear Processes in Geophysics 12*, 201–210.

Katz, R. W., G. S. Brush, and M. B. Parlange (2005). Statistics of extremes: Modeling ecological disturbances. *Ecology 86*(5), 1124–1134.

Katz, R. W., M. B. Parlange, and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources 25*, 1287–1304.

Kendall, M. G. (1975). *Rank Correlation Methods*. London: Charles Griffin.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses and models to advance the science of hydrology. *Water Resources Research 42*, W03S04.

KLIWA (2003). Langzeitverhalten der Hochwasserabflüsse in Baden-Württemberg und Bayern. Technical report, München.

KLIWA (2004). Klimaveränderung und Konsequenzen für die Wasserwirtschaft. 2. Symposium. Technical report, Arbeitskreis KLIWA, München.

Koutsoyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal 48*(1), 3–24.

Kropp, J. P. and H.-J. Schellnhuber (2006). Weiterentwicklung multivariater statistischer Analysemethoden, Anwendung auf meteorologische Parameter zur Beschreibung hydrologischer Vorgänge, Einsatz von N/A-Modellen zur Erzeugung von Basisdaten für die Fluktuationsanalyse. Kurztitel: Skalenanalyse. Technical report, Bundesministerium für Bildung und Forschung (BMBF), Projekt Skalenanalyse (Förder-Nr. 0330271).

Küchenhoff, H. and M. Thamerus (2004). Extreme value analysis of Munich air pollution data. `http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper04.ps`.

Kundzewicz, Z. W., D. Graczyk, I. Przymusinska, T. Maurer, M. Radziejewski, C. Svensson, and M. Szwed (2004). Detection of change in world-wide hydrological time series of maxium annual flow. Technical report, Global Runoff Data Centre (GRCD).

Kundzewicz, Z. W. and H.-J. Schellnhuber (2004). Floods in the IPCC TAR perspective. *Natural Hazards 31*, 111–128.

Kundzewicz, Z. W. and S. P. Simonovic (2002). Non-structural flood protection and sustainability. *Water International 27*, 3–13.

Lange, H. (2006). *Encyclopedia of Life Sciences*, Chapter Time-series Analysis in Ecology, pp. 8, Doi: 10.1038/npg.els.0003276. Wiley InterScience.

Lawrence, A. J. and N. T. Kottegoda (1977). Stochastic modelling of river-flow time series (with discussion). *Journal of the Royal Statistical Society A 140*, 1–47.

Leadbetter, M. R. G. (1991). On a basis for 'peaks over threshold' modeling. *Statistics & Probability Letters 12*, 357–362.

Leadbetter, M. R. G., G. Lindgren, and H. Rootzen (1983). *Extremes and Related Properties of Random Sequences and Series*. New York: Springer.

Lefebvre, M. (2002). Geometric Brownian motion as a model for river flows. *Hydrological Processes 16*, 1373–1381.

Lohre, M., P. Sibbertsen, and T. Könning (2003). Modeling water flow of the Rhine river using seasonal long memory. *Water Resources Research 39*(5), 1132.

Lucio, P. S. (2004). Geostatistical assessment of HadCM3 simulations via NCEP reanalyses over Europe. *Atmospheric Science Letters 5*(6), 118–133.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner (2004). European seasonal and annual temperature variability, trends and extremes since 1500. *Science 303*(5).

MacLean, C. J. (1974). Estimation and testing of exponential polynomial rate function within the nonstationary Poisson process. *Biometrika 61*(1), 81–85.

Maidment, D. R. (1993). *Handbook of Hydrology*. New York: MCGraw-Hill.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence 11*(7), 674–693.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica 13*, 245–259.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature 392*, 779–787.

Mann, M. E., R. S. Bradley, and M. K. Hughes (2004). Corrigendum: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature 430*(105).

Montanari, A., R. Rosso, and M. S. Taqqu (1997). Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation. *Water Resources Research 33*(5), 1035.

Montanari, A., R. Rosso, and M. S. Taqqu (2000). A seasonal fractional ARIMA model applied to the Nile river monthly flows at Aswan. *Water Resources Research 36*, 1249–1259.

Morrison, J. E. and J. A. Smith (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research 38*(12), 1305.

Mudelsee, M. (2007). Long memory of rivers from spatial aggregation. *Water Resources Research 43*, W01202.

Mudelsee, M., M. Börngen, G. Tetzlaff, and U. Grünewald (2004). Extreme floods in central Europe over the past 500 years: Role of cyclone pathway "Zugstrasse Vb". *Journal of Geophysical Research 109*, D23101.

Naveau, P., M. Nogaj, C. Ammann, P. Yiou, D. Cooley, and V. Jornelli (2005). Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience 337*, 1013–1022.

Nelder, J. A. and R. W. M. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society A 135*, 370–384.

Nogaj, M., P. Yiou, S. Parey, F. Malek, and P. Naveau (2006). Amplitude and frequency of temperature extremes over the North Atlantic region. *Geophysical Research Letters 33*, L10801.

Oakley, J. and A. O'Hagan (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika 89*, 769–784.

Ooms, M. and J. Doornik (1999). Inference and forecasting for fractional autoregressive integrated moving average models, with an application to US and UK inflation. Econometric Institute Report 9947/A.

Ooms, M. and P. H. Franses (2001). A seasonal periodic long memory model for monthly river flows. *Environmental Modelling & Software 16*, 559–569.

Osborn, T. J., M. Hulme, P. D. Jones, and T. A. Basnett (2000). Observed trends in the daily intensity of United Kingdom precipitation. *International Journal of Climatology 20*, 347–364.

Pakosch, S. (2004). Statistische Methoden zur stationären und instationären Auswertung von gemessenen Maximalabflüssen mit Hilfe theoretischer Verteilungsfunktionen. Master's thesis, Universität Stuttgart.

Parent, E. and J. Bernier (2003). Bayesian POT modeling for historical data. *Journal of Hydrology 274*, 95–108.

Percival, D. B. (2002). *Encyclopedia of Environmetrics*, Volume 4, Chapter Wavelets, pp. 2338–51. Chichester: John Wiley & Sons.

Percival, D. B., J. E. Overland, and H. O. Mofjeld (2001). Interpretation of North Pacific variability as a short- and long-memory process. *Journal of Climate 14*, 4545–59.

Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

Pfister, C., R. Brázdil, R. Glaser, M. Barriendos, D. Camuffo, M. Deutsch, P. Dobrovolný, S. Enzi, E. Guidoboni, O. Kotyza, S. Militzer, L. Raczii, and F. S. Rodrigo (1999). Documentary evidence on climate in sixteenth-century Europe. *Climate Change 43*, 55–110.

Pfister, L., J. Kwadijk, A. Musy, A. Bronstert, and L. Hoffmann (2004). Climate change, land use change and runoff prediction in the Rhine-Meuse basins. *River Research and Applications 20*, 229–241.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Rahmstorf, S. and H.-J. Schellnhuber (2006). *Der Klimawandel*. München: C.H. Beck Verlag.

Ramesh, N. I. and A. C. Davison (2002). Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology 256*, 106–119.

Rinne, H. and K. Specht (2002). *Zeitreihen. Statistische Modellierung, Schätzung und Prognose*. München: Verlag Franz Vahlen.

Robson, A. J. (2002). Evidence for trends in UK flooding. *Philosophical Transactions of the Royal Society A 360*, 1327–1343.

Robson, A. J., T. K. Reed, D. W. Reed, and A. C. Bayliss (1998). A study of national trend variation in UK floods. *International Journal of Climatology 18*(2), 165–182.

Rust, H. W. (2007). *Detection of Long-Range Dependence – Applications in Climatology and Hydrology*. Ph. D. thesis, Potsdam University, Potsdam.

Rust, H. W., M. Kallache, H.-J. Schellnhuber, and J. Kropp (2007). Confidence intervals for flood return level estimates using a bootstrap approach. Accepted by Advances in Water Resources.

Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006). Long-term persistence in climate and the detection problem. *Geophysical Research Letters 33*, L06718.

Sachs, L. (1991). *Angewandte Statistik*. Berlin: Springer.

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*(22), 332–336.

Schellnhuber, H.-J. (1998). *Earth System Analysis – Integrating Science for Sustainability*, Chapter Earth System Analysis – The Scope of the Challenge, pp. 3–195. Berlin: Springer.

Schellnhuber, H.-J. and J. Kropp (1998). Geocybernetics: Controlling a complex dynamical system under uncertainty. *Naturwissenschaften 85*, 411–425.

Schneider von Deimling, T., H. Held, A. Ganopolski, and S. Rahmsdorf (2006). Climate sensitivity estimated from ensemble simulations of glacial climate. *Climate Dynamics 27*(2-3), 149–163.

Schröter, D., M. Zebisch, and T. Grothmann (2005). Climate change in Germany – vulnerability and adaptation of climate-sensitive sectors. Technical report, DWD. Klimastatusbericht.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–464.

Sibbertsen, P. (1999). *Robuste Parameterschätzung im linearen Regressionsmodell bei Fehlertermen mit langem Gedächtnis*. Berlin: Verlag für Wissenschaft und Forschung.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika 72*, 67–90.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science 4*(4), 376–377.

Smith, R. L. (1993). *Statistics for the Environment*, Chapter Long-range dependence and global warming, pp. 141–161. New York: John Wiley & Sons Ltd.

Smith, R. L. (2001). *Environmental Statistics*. `http://www.stat.unc.edu/postscript/rs/envnotes.pdf`.

Smith, R. L. (2003). *Extreme Values in Finance, Telecommunications and the Environment*, Chapter 1, pp. 1–78. Boca Raton: Chapman and Hall/CRC Press.

Smith, R. L., T. M. L. Wigley, and B. D. Santer (2003). A bivariate time series approach to anthropogenic trend detection in hemispheric mean temperatures. *Journal of Climate 16*, 1228–1240.

Snyder, D. L. (1975). *Random Point Processes*. New York: John Wiley & Sons Ltd.

Stern, N. (2006). Stern review on the economics of climate change. Technical report, UK Government.

Stoev, S., M. S. Taqqu, C. Park, and J. S. Marron (2005). On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of internet traffic. *Computer Networks 48*, 423–445.

Stott, P. A., D. A. Stone, and M. R. Allen (2004). Human contribution to the European heatwave 2003. *Nature 432*, 610–614.

Strupczewski, W. G., V. P. Singh, and W. Feluch (2001). Non-stationary approach to at-site flood frequency modelling. I. Maximum likelihood estimation. *Journal of Hydrology 248*, 123–142.

Strupczewski, W. G., V. P. Singh, and H. T. Mitosek (2001). Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of Polish rivers. *Journal of Hydrology 248*, 152–167.

Survey, I. S. W. (1994). Streamflow conditions, flooding and low flows. Technical report, Illinois State Water Survey.

Taqqu, M. S., V. Teverovsky, and W. Willinger (1995). Estimators for long-range dependence: An empirical study. *Fractals 4*(3), 785–798.

Thodsen, H. (2007). The influence of climate change on stream flow in Danish rivers. *Journal of Hydrology 333*, 226–238.

Torrence, C. and P. Compo (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society 79*(1), 61–78.

Turner, B. L., R. E. Kasperson, P. A. Matson, J. J. McCarthy, R. W. Corell, L. Christensen, N. Eckley, J. X. Kasperson, A. Luers, M. L. Martello, C. Polsky, A. Pulsipher, and A. Schiller (2003). A framework for vulnerability analysis in sustainability science. *PNAS 100*(14), 8074–8079.

Tyndall, J. (1861). The Bakerian lecture: On the absorption and radiation of heat by gases and vapours, and on the physical connexion of radiation, absorption, and conduction. *Philosophical Transactions of the Royal Society of London 151*, 1–36.

Vetterli, M. and J. Kovacevic (1995). *Wavelets and Subband Coding*. New Jersey: Prentice Hall PTR.

Vyushin, D., I. Zhidkov, S. Havlin, A. Bunde, and S. Brenner (2004). Volcanic forcing improves athmosphere-ocean coupled general circulation model scaling performance. *Geophysical Research Letters 31*, L10206.

Wigley, T. M. L. (1985). Climatology: Impact of extreme events. *Nature 316*, 106–107.

Willems, W. (2005). Assessment of design flows in water management: Classical methods, instationary and multidimensional extensions. Talk at conference "Developments in trend and extreme value analysis of hydrometeorological time series", PIK Potsdam.

Woodward, W. A. and H. L. Gray (1993). Global warming and the problem of testing for a trend in time series analysis. *Journal of Climate 6*, 953–962.

Woodward, W. A. and H. L. Gray (1995). Selecting a model for detecting the presence of a trend. *Journal of Climate 8*, 1929–1937.

Yajima, Y. (1985). On estimation of long-memory time series models. *Australian Journal of Statistics 27*, 303–320.

Yue, S. and C. Y. Wang (2002). Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test. *Water Resources Research 38*(6), 1068.

Zhang, X., K. D. Harvey, W. D. Hogg, and T. R. Yuzyk (2001). Trends in Canadian streamflow. *Water Resources Research 37*, 987–998.

Zhang, X., F. W. Zwiers, and G. Li (2004). Monte Carlo experiments on the detection of trends in extreme values. *Journal of Climate 17*(10), 1945–1952.

Symbol Index

α	significance level of hypothesis testing, page 82
B	back-shift operator, page 10
$\boldsymbol{\mathcal{D}}$	time series detail, page 16
δ	long-term correlation parameter, related to Hurst coefficient, page 11
D	deviance statistic, page 38
ϵ_t	error term, noise, page 10
$F(.)$	$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ joint distribution of $X_1, X_2, \ldots, X_n$, page 9
$\gamma(\cdot)$	auto-covariance function, page 97
$\mathcal{G}(\omega), \mathcal{H}(\omega)$	squared gain functions, page 17
$g_{j,l}$	scaling filter, page 14
H	Hurst coefficient, page 12
$h_{j,l}$	wavelet filter, page 13
λ	Poisson process occurrence rate, page 22
$\Lambda(\cdot)$	point process intensity measure, page 24
$\mathcal{L}(\cdot)$	likelihood, page 35
$l(\cdot)$	log likelihood, page 38
μ	location parameter of the GEV distribution, page 20
m	length of interval out of which GEV maxima are drawn (typically $m = 365$ days), page 20
ν_Y^2	wavelet variance, page 19
n	number of observations, page 9
ω	Fourier frequency, page 17
$\mathcal{P}(\omega_j)$	periodogram, page 34
$\phi(z)$	autoregressive polynomial, page 10
ψ	scale parameter of the GPD distribution, page 23
$\psi(z)$	moving average polynomial, page 10
$\rho(\cdot)$	auto-correlation function, page 11
$\boldsymbol{\mathcal{S}}$	time series smooth, page 16
$\mathcal{S}(\boldsymbol{\Theta}; \omega_j)$	spectral density, page 34
Σ	variance-covariance matrix, page 33
σ	scale parameter of the GEV distribution, page 20
σ_ϵ^2	variance of innovations of a stochastic process, page 10
$\boldsymbol{\Theta}$	parameter vector, page 100
τ_j	wavelet scale, page 13

θ extremal index, page 22

$\theta(\cdot)$ phase function, page 18

u threshold, page 22

$V_{j,k}$ scaling coefficients, page 14

$W_{j,k}$ wavelet coefficients, page 14

ξ shape parameter of the GEV and GPD distribution, page 20

$\{X_t, Y_t\}$ stochastic process, page 9

y excesses, $x - u$, page 22

z_p return level , page 31

A-1 SPECTRAL DENSITY AND AUTO-COVARIANCE FUNCTION OF A FARIMA PROCESS

The spectral density of a stationary FARIMA process X_t is determinable from the according ARMA process. Assume $\mu = E(X_t) = 0$, then

$$f(\omega) = |1 - e^{i\omega}|^{-2\delta} f_{\text{ARMA}}(\omega) = |1 - e^{i\omega}|^{-2\delta} \frac{\sigma_\epsilon^2}{2\pi} \frac{|\psi(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2} \quad . \tag{A-1.1}$$

Here σ_ϵ^2 denotes the variance of the innovations and $\phi(z)$ and $\psi(z)$ denote the AR and MA polynomials, respectively (Beran 1994).

It is possible to represent a FARIMA process by an infinite sum of MA terms for $\delta > -0.5$:

$$X_t = \phi^{-1}(B)(1 - B)^{-\delta}\psi(B)\epsilon_t = 1 + \sum_{j=1}^{\infty} \psi_j \epsilon_j \quad . \tag{A-1.2}$$

This can be used to compute the auto-covariance function $\gamma(k) = E(X_t X_{t-k})$ of a stationary FARIMA process. Assume $\mu = E(X_t) = 0$, then the MA representation of the auto-covariance function of a FARIMA process is given by

$$\gamma(k) = \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} \sigma_\epsilon^2 \quad , \tag{A-1.3}$$

with $\psi_0 = 1$ (see Chan and Palma (1998)). A drawback here is that many terms are needed for an accurate approximation, because ψ_j decays hyperbolically. Further techniques to approximate the auto-covariance function of a FARIMA process are listed in Ooms and Doornik (1999).

A relatively simple alternative is to numerically integrate over the power spectrum

$$\gamma(k) = \int_{-\pi}^{\pi} f_X(\omega) e^{i\omega k} d\omega \quad , \tag{A-1.4}$$

(Harvey 1993). The spectrum of the FARIMA process $f_X(\omega)$ is easy to compute, but numerical integration for each k is time consuming.

The covariances of a FD(δ) process follow directly from a formula which is outlined in Gradshteyn and Ryzhik (1965)

$$\gamma(k) = \sigma_\epsilon^2 \frac{(-1)^k \Gamma(1 - 2\delta)}{\Gamma(k - \delta + 1)\Gamma(1 - k - \delta)} \quad . \tag{A-1.5}$$

A-2 SIMULATION STUDY TO EVALUATE THE LIKELIHOOD ESTIMATION OF POINT PROCESSES

To implement the point process approach in terms of the GEV parameters $\mathbf{\Theta} = (\mu, \sigma, \xi)$, the *pp.fit()* routine of the *ismev* package has been changed according to Eq. (2.55). We conduct two simulation studies to evaluate the quality of the point process fit.

In this procedure, several parameters are fitted simultaneously. Therefore, as usual in non-linear optimisation problems, it is possible that the algorithm runs into local minima. To consider this problem and therefore enhance the quality of the fit, it is repeated several (i.e. 20) times. The strength of the trend guess of the parameters is randomly varied, which equals a variation of the initial guesses of the parameters. Then the parameter estimate with the largest likelihood value is taken.

For the first of the two studies, an ensemble of 1 000 log-normal distributed series with 5 415 data points each is created. This equals the length of 60 years of measurement of a 90-day season and is thus a setting we also encounter in empirical records (cf. Sec. 2.7). Then a strong trend which exhibits a fifth of the variance of the whole dataset itself is added to the mean of the data. A non-stationary point process with time-dependent location parameter μ, which should capture the trend added, is fitted to each of the ensemble members. To get a well defined simulation setting with no additional sources of error, the guess for the trend is the true trend. The threshold is chosen such that 1.22 percent of the data (about 65 data points) lie above the threshold. We chose the threshold as such, because this resembles our empirical settings.

To evaluate the simulation study, exemplarily 20 windows of width 10 - equally distributed over the whole data length - are chosen. In the simulation study now the non-stationary point process which is estimated over the whole time series is compared to "local" fits in the windows. The average of many local estimates serves as a reference to evaluate the non-stationary fit. Thus the non-stationary point process fit which is fitted to the whole data set is compared in these windows by

(a) averaging the GPD parameter estimates of the 1 000 non-stationary point process fits for each window and comparing the resulting "average" GPD with the empirical distribution of the maxima of all simulations in this window (i.e. 10 000 data points, about 120 maxima).

(b) comparing the average of the 1 000 non-stationary fits in this window with a stationary point process fit to the maxima of all the 10 000 data points in the window (here the threshold is chosen as 1.22 percent as well).

(c) comparing the return level, which is exceeded on average once every 9 000 observations (i.e. 100 "years") in each window. They are calculated by using the stationary and non-stationary GPD described in (b).

Although we add a trend to the mean of the data, we know that this linear trend will also have an influence on the extremes. Furthermore, in this simulation study we look at the maxima of a whole simulation ensemble and therefore should get an impression of the influence of trends in the mean on extreme values for large n.

The simulation study reveals a good agreement between the averaged theoretical distribution of the non-stationary fits and the empirical data (not shown). The estimates

are worst at the edges of the time series, which is depicted in the left of Fig. A-2.1. The GPD resulting from averaging all parts of non-stationary point process fits covering the first window (time points 1 to 10) lies outside the confidence interval of the stationary point process fit of the first window. The same can be seen for the last window (time points 5405-5410). Furthermore, the proximity of the non-stationary fits suggests that the non-stationary fit underestimates the trend at the edges.

Nevertheless, all parameters are jointly estimated in this maximum likelihood approach. The time evolution of one parameter may balance the movement of another parameter. Therefore the parameter estimates might not be unambiguous. Hence, to get a better picture of the adequacy of the fit, it is better to have a look at assessment measures which involve all of the estimates, such as the return level (see also Naveau et al. 2005). In Fig. A-2.1 we see that the estimated return levels of the non-stationary point process fit lie well within the confidence interval of the return levels obtained by a stationary fit in each window.

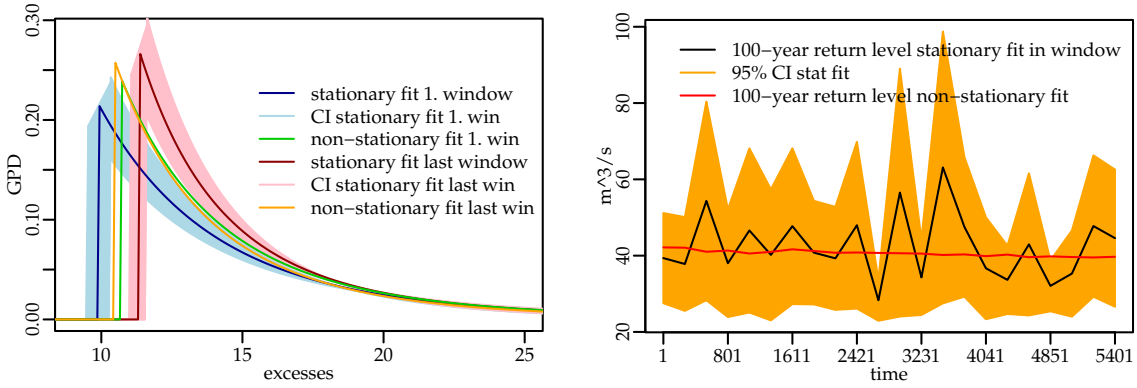


Figure A-2.1: Comparison of non-stationary point process fits and stationary point process fits in small windows. Left: GPD which results in averaging all parameter values of the non-stationary point process fit in the first window (green) and in the last window (orange) in comparison to the GPD of a stationary point process fit over all maxima in the first window (blue) and the last window (red) and related 95% confidence intervals obtained by the delta method. **Right**: Comparison of 100-year return level in 22 windows of length 10. The quantiles are calculated from a stationary point process fit to all maxima in the window (black) and from a GPD resulting from averaging the parameters of all non-stationary point processes in that window (red). The 95% confidence interval (orange) is obtained by the delta method.

As a second study, we generated excesses out of a non-stationary point process with a linear trend in σ (the generating procedure is described in App. A-3). The non-stationary point process is chosen to have a parameter vector $\boldsymbol{\Theta} = (69.63, 43.90, 0.076)$ and a threshold $u = 64.83$, which are values taken from a stationary point process fit to run-off data of the Danube River at Beuron. The slope s of the added trend ranges from 0 to 0.1 and the generated series have a length of 5 415 (18 050) data points, which equals 60 (200) years of daily data, evaluated at a season of three months. Using Eq. (2.55) we fit a non-stationary point process to the series and compare the results to a GPD fit (the *gpd.fit()* routine of the R-*ismev* package). In both cases we use the "true" trend as initial trend guess for the scale parameter σ. The whole procedure is repeated for an ensemble of 1 000 runs. To account for the uncertainty of the simulation study itself, one would have to repeat this study several times and present the averaged results with confidence intervals. This is beyond the scope of this work. Instead, the study is repeated a second time. The results

do not change a lot and qualitative statements do not change at all. So only results of one study are presented here.

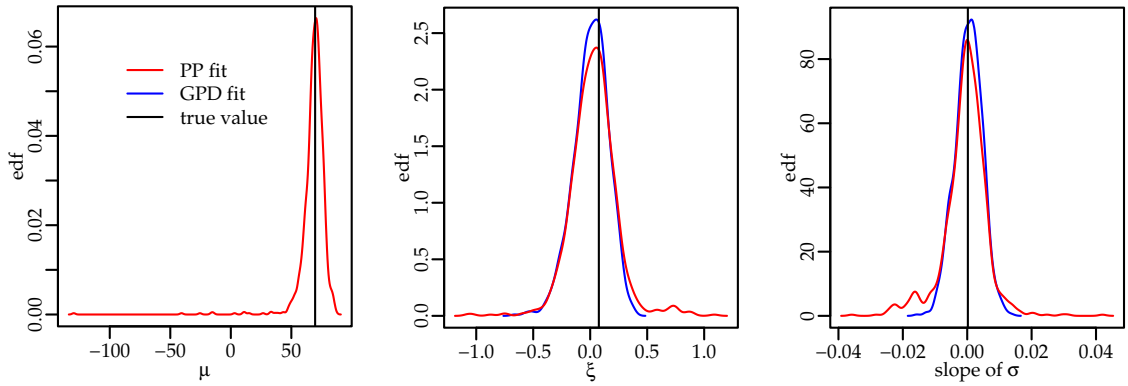


Figure A-2.2: Goodness of point process fit. Left: Distribution of 1 000 estimates of the location parameter μ with a point process model. **Middle**: Distributions of estimates of the shape parameterξ with a point process and a generalized Pareto process model. **Right**: Distribution of estimates of the slope of the time-dependent scale parameter σ obtained from a point process and a generalized Pareto process model.

In Fig. A-2.2 the goodness of the point process fit in comparison to a GPD fit is depicted. In this simulation study only σ is non-stationary, so the distributions of estimates for μ and ξ are directly depicted and, regarding σ, the distribution of the estimated slope of the linear trend in the scale parameter σ is shown. Although the true values lie in the center of the distributions, the distributions obtained with the PP fit using GEV notation exhibit wide-spread tails, that is outliers. The GPD fit does generally not show such far-flung outliers.

In Fig. A-2.3 the simulated cumulative distribution function of the slope of the linear trend in σ of the PP fit using GEV notation in comparison to the GPD fit is depicted for different slopes of the added trend and the length of 5 415 and 18 050 data points. Again it is visible that the GPD fit outperforms the point process fit in most of the cases.

For the length of 5 415 data points which is relevant for our studies of empirical data the point process fit is at least as good as the GPD fit in case no trend is present in the data. For small trends the GPD fit is better than the PP fit, which means that the distribution of the slope estimates generated by simulations is more narrow. This advantage diminishes for strong trends, as can be seen in Figs. A-2.3 (5) and A-2.3 (6). For longer series, the distribution generated out of the PP fit is as narrow as the distribution generated out of the GPD fit, but has wide spread tails.

To quantify these differences, the [16%,84%] and [2.5%,97.5%] quantiles of the distribution of the slope of the scale parameter σ (generated by simulations) are given in table A-2. The depicted intervals enclose 68% or respectively 95% of the distribution generated by simulations. To get an impression of the variability of the estimated slope, we compare it to the variability of σ itself in the stationary case. In our model the variability of the slope s affects σ by $\sigma = \sigma_{stat} + s \cdot t$. The added linear trend is increasing. Therefore the broadest variation of σ by incorporating the variation of the slope is expressed by $\sigma \in [\sigma_{stat} - s_{min} \cdot t_1, \sigma_{stat} + s_{max} \cdot t_n]$. This interval is given for 5 415 data points

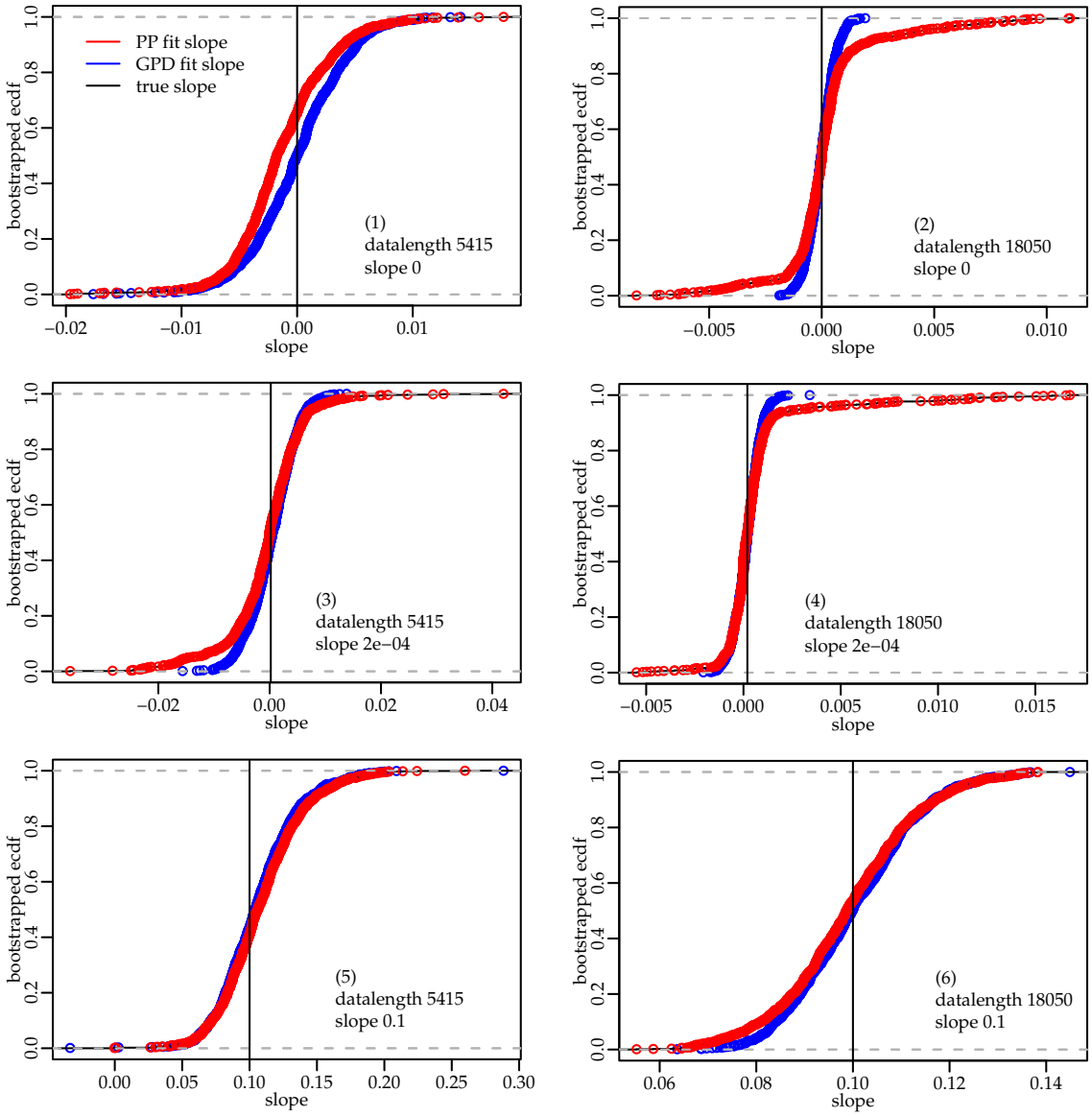


Figure A-2.3: Estimated slope of non-stationary σ. Comparison of point process fit and generalized Pareto fit. **Left**: Data length of 5 415, which equals 60 years of daily data evaluated at a season of three months. No trend added to the data (1), weak trend of 0.0002 (3) and strong trend of 0.1 (5). **Right**: Data length of 18 050, which equals 200 years of daily measurements evaluated for a season of three months. No trend added to the data (2), weak trend of 0.0002 (4) and strong trend of 0.1 (6).

and no trend by $\sigma \in [\sigma_{stat} - 5.036, \sigma_{stat} + 14.78]$. In comparison, one standard deviation for σ in the stationary model fit for the same data is 9.9106. For 18 050 data points $\sigma \in [\sigma_{stat} - 4.332, \sigma_{stat} + 4.8735]$ has to be compared with one standard deviation of 4.7863 in the stationary case. This means that the variability induced by one additional parameter, that is the slope of the trend, lies well within the variability induced by the stationary fit. For the smallest slope $s = 0.0002$ tested, the CI's of one standard deviation change to $\sigma \in [\sigma_{stat} - 7.47, \sigma_{stat} + 8.126]$ for 5 415 data points and to $\sigma \in [\sigma_{stat} - 2.166, \sigma_{stat} + 4.87]$ for 18 050 data points. Evaluating these figures, the variability of the estimated slope using

5 415	68% PP	68% *GPD*	95% PP	95% *GPD*
0	[-0.0031,0.0091]	*[-0.0036,0.0039]*	[-0.0072,0.0178]	*[-0.0082,0.0082]*
0.0002	[-0.0046,0.0050]	*[-0.0044,0.0039]*	[-0.0147,0.0420]	*[-0.0156,0.0068]*
0.001	[-0.0028,0.0060]	*[-0.0030,0.0054]*	[-0.0069,0.0363]	*[-0.0070,0.0115]*
0.1	[0.0688,0.1257]	*[0.0722,0.1251]*	[-0.0002,0.1647]	*[-0.0331,0.1569]*
18 050	68% PP	68% *GPD*	95% PP	95% *GPD*
0	[-0.0008,0.0009]	*[-0.0006,0.0006]*	[-0.0042,0.0069]	*[-0.0011,0.0013]*
0.0002	[-0.0004,0.0009]	*[-0.0005,0.0008]*	[-0.0013,0.0100]	*[-0.0012,0.0014]*
0.001	[0.0000,0.0021]	*[0.0003,0.0018]*	[-0.0050,0.0117]	*[-0.0006,0.0026]*
0.1	[0.0880,0.1142]	*[0.0879,0.1129]*	[0.0748,0.1383]	*[0.0781,0.1284]*

Table A-2.1: Confidence Intervals enclosing 68% and 95% of the distribution of the slope estimates for the non-stationary scale parameter σ. The imposed trends have slope values ranging from 0 (i.e. no trend) to 0.1 of the variability of the process. Length of data: 5 415 and 18 050 data points.

the PP fit seems to be acceptable. Anyhow, the GPD fit outperforms the PP fit with GEV notation in most of the cases. Therefore in our studies we use a PP fit using GPD notation. By doing so, we are able to separately estimate the GPD parameters and the parameter λ to model the frequency of the extreme events. We consider this approach superior to the PP fit using a notation in terms of the GEV parameters.

A-3 GENERATION OF NON-STATIONARY POINT PROCESSES

Artificially generated point process data are useful for uncertainty assessment, e.g. to provide confidence intervals, and to simulate the distribution of model selection criteria. Here the methodology to generate non-stationary point processes which model extreme values is described.

To generate non-stationary PP processes utilising the parameters of the GEV $\mathbf{\Theta} = (\mu, \sigma, \xi)$, we use the likelihood given in Eq. (2.55). The inversion of the integrated rate function and the exponential spacings method, both described in Devroye (1986), are utilised to generate the occurrence times T of the extreme events. We particularly exploit that the waiting time between two extreme events $T_{k+1} - T_k$, given T_k, is exponentially distributed. So, given T_k, we model $T_{k+1} = T_k + \Lambda^{-1}(E + \Lambda(T_k))$, where E is an exponentially distributed random variable with rate 1. Starting with $T_0 = 0$ and proceeding until T_n is crossed, gives the desired times of exceedances[1].

The sizes of the exceedances follow a non-homogeneous Poisson process with intensity

$$\lambda_2(t, y) = -\frac{\partial \Lambda_2([y, \infty])}{\partial y} = \sigma_t^{-1}\left[1 + \xi_t\left(\frac{y - \mu_t}{\sigma_t}\right)\right]_+^{-1/\xi_t - 1} , \qquad \text{(A-3.6)}$$

for $y > u$. So, conditional on n_u exceedances over level u, we obtain the sizes of the excesses by drawing randomly out of a Generalized Pareto distribution with parameter vector $\mathbf{\Theta}_t = (u, \psi_t, \xi_t)$ where $\psi_t = \sigma_t + \xi_t(u - \mu_t)$.

We evaluate the performance of the procedure sketched above to generate non-stationary point processes using a notation in terms of the GEV parameters with simulation studies. To do so, we generate log-normal distributed stationary data of length 5 415, 9 025 and 18 050 (which equals periods of 60, 100 and 200 years of daily observation, evaluated over a season of 3 months). Then for each generated series, a threshold is calculated such that 1.22% of the number of data points lie above it. Although stationary data are modelled, a trend shape has to be specified when using non-stationary models. Here three types of models are fitted to the excesses, namely

(1) a stationary point process,

(2) a point process model with a linear trend assumption in μ, and

(3) a point process with a quadratic trend assumption in μ.

The fits of model (1)-(3) now serve as reference fits, that is we want to reproduce this data using our simulation procedure. To estimate the quality of the reproduced data, we use each of the model fits to generate an ensemble of 1 000 artificial series according to the above described method. Then to each generated time series the same model is fitted from which it has been generated. In this way, a distribution of the GEV parameter estimates is derived for each of the three models. This provides an idea of the uncertainty related to the point process simulation procedure using $\mathbf{\Theta} = (\mu, \sigma, \xi)$ (the uncertainty of the point process fit is assessed in Sec. A-2). In Fig. A-3.1 the evolution in time of the estimates for the parameter μ is depicted for all three models: The reference fit of the

[1] Alternatively, the *rpois()* routine of the *stats* package of ***R*** may be used. This routine is capable of handling a time-dependent λ_t.

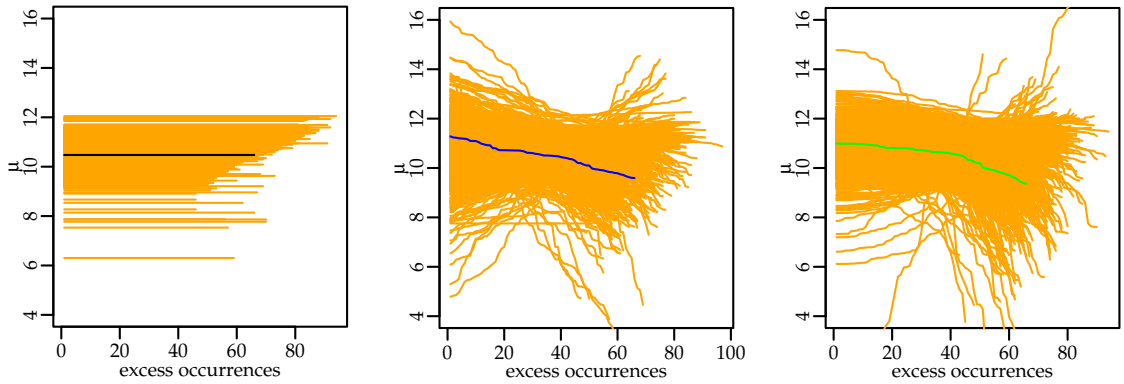


Figure A-3.1: Test of simulation of non-stationary point processes: Evolution of the GEV parameter μ with time, plotted for time points where an excess occurs. Total length of series: 5 415, Monte Carlo ensemble with 1 000 members. **Left**: Stationary PP fit (black) and Monte Carlo ensemble (orange). **Middle**: Non-stationary PP fit with linear trend assumption in μ (blue) and Monte Carlo ensemble (orange). **Right**: Non-stationary PP fit with quadratic trend assumption in μ (green) and Monte Carlo ensemble (orange).

stationary PP fit (black), the non-stationary PP fit with a linear trend assumption in μ (blue) and the non-linear PP fit with a quadratic trend assumption in μ (green) and their respective Monte Carlo ensembles (orange). The variability of $\hat{\mu}$ of the non-stationary PP fits does not seem to be remarkably higher than the one of the stationary PP fit. Nevertheless, the estimates often suggest a trend in μ with, at least for some fits, considerable slope. This may partly be due to the simulation routine and partly due to the fit routine, which is tested in Sec. A-2. The original log-normal distributed data are stationary, i.e. without trend, but the reference fits (blue and green) already show a trend tendency.

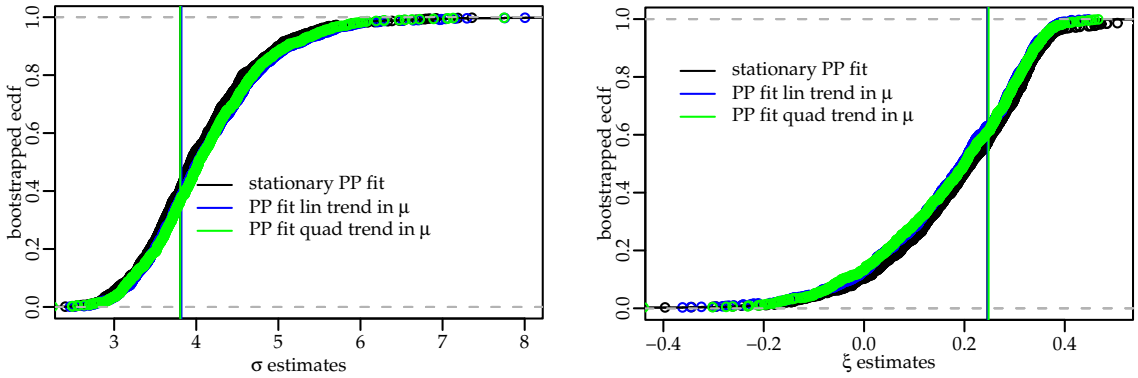


Figure A-3.2: Test of simulation of non-stationary point processes (σ and ξ): Length of series: 5 415. **Left**: cdf generated by simulations (points) and value of original fit (vertical line) for the σ parameter of the stationary PP fit (black), the non-stationary PP fit with a linear trend assumption in μ (blue, hidden under the green points) and the non-stationary PP fit with a quadratic trend assumption in μ (green). **Right**: The same as left for ξ.

The estimates of σ and ξ are depicted as cdf in Fig. A-3.2. The values of the reference fits are added as vertical lines. We see a good agreement between the cdfs of the estimates of the stationary (black) and the two non-stationary fits (blue and green). So for the constant parameters we do not get more uncertainty due to an assumption of

non-stationarity in μ. The cdf for the important shape parameter ξ exhibits a remarkable range from negative to positive values. This effect already has to be considered for stationary model assumptions. In Figs. A-3.3 and A-3.4 the same results are depicted for

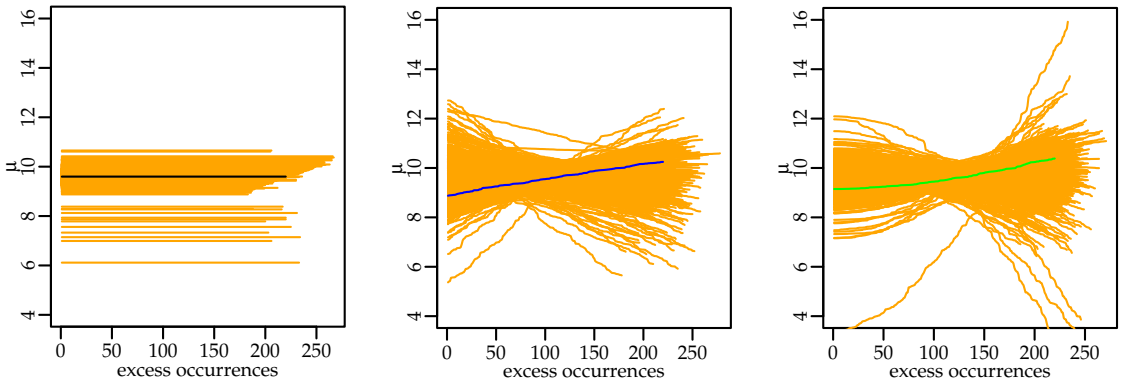


Figure A-3.3: Test of simulation of non-stationary point processes: Evolution of the GEV parameter μ with time, only plotted for time points where an excess occurs. Total length of series: 18 050, Monte Carlo ensemble with 1 000 members. **Left**: Stationary PP fit (black) and Monte Carlo ensemble (orange). **Middle**: Non-stationary PP fit with linear trend assumption in μ (blue) and Monte Carlo ensemble (orange). **Right**: Non-stationary PP fit with quadratic trend assumption in μ (green) and Monte Carlo ensemble (orange).

series of length 18 050 (which equals 200 years of daily observation and evaluation of a season of three months). Qualitatively the same results are obtained. Quantitatively the distributions of all parameter estimates get more narrow, which we have expected.

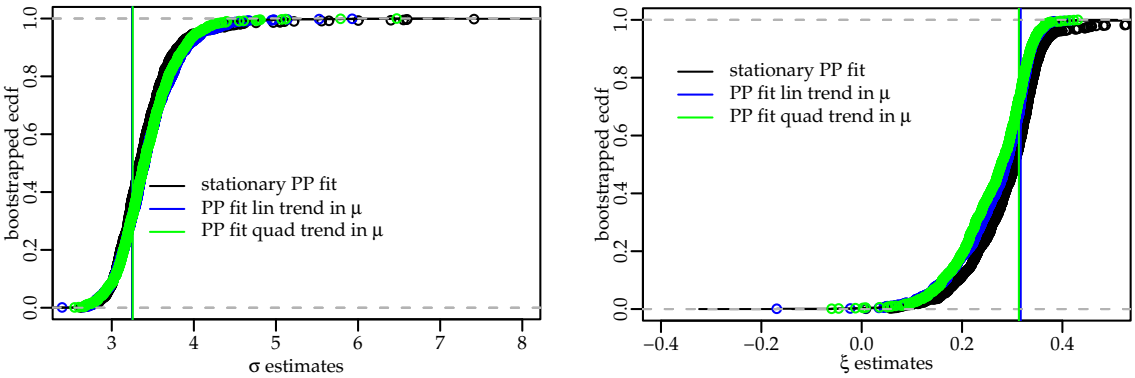


Figure A-3.4: Test of simulation of non-stationary point processes (σ and ξ): Length of series: 18 050. **Left**: cdf of estimates (points) and value of original fit (vertical line) for the σ parameter of the stationary PP fit (black), the non-stationary PP fit with a linear trend assumption in μ (blue, hidden under the green points) and the non-stationary PP fit with a quadratic trend assumption in μ (green). **Right**: The same as left for ξ.

To generate non-stationary PP using the parameters $\mathbf{\Theta} = (\lambda, \psi, \xi)$, we employ the likelihood given in Eq. (2.54). The simulation procedure is exactly the same as the one for generating PP using GEV parameters, which has just been described. But there we have to calculate λ_t out of the GEV parameters, i.e. $\lambda_1(t) = \{1 + \xi_t(u - \mu_t)\sigma_t^{-1})\}_+^{-1/\xi_t}$. Now we are able to use smooth functions to model the time dependence of λ_t, such as the

exponential function. The relationship between both parameter vectors $\boldsymbol{\Theta} = (\mu, \sigma, \xi)$ and $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$, given in Eqs. (2.29) and (2.34), remains, but here the GEV parameters μ and σ would be expressed in terms of $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$, in case they had to be calculated. Apparently, we now use the same functions to describe λ_t when simulating the point processes, as we utilise as trend guesses when fitting non-stationary PP to the data (cf. Sec. (4.4)). We are able to use "nice" shapes for λ_t and therefore we expect the simulation results to get better than when generating point processes using the parameter estimates of the GEV. We checked this for some time series and our expectations were affirmed. Therefore we prefer the notation of the point process in terms of $\boldsymbol{\Theta} = (\lambda, \psi, \xi)$, when generating point processes.

A-4 DANUBE RIVER BASIN: THRESHOLD AND CLUSTER SIZE

River/Gauge	Threshold [m^3/s]	Cluster Size	Number of Extremes (left after declustering)
Donau/Beuron	64.83	6	67
Iller/Sonthofen	35.25	8	56
Iller/Kempten	116.21	6	54
Osterach/Reckenberg	11.57	6	50
Iller mit Kanal/Wiblingen	180.48	6	55
Zusam/Pfaffenhofen	11.83	6	66
Lech/Landsberg	87.69	4	53
Vils/PfrontenRied	8.19	8	100
Wertach/Bissenhofen	31.44	8	100
Geltnach/Hörmannshofen	4.73	8	117
Altmühl/Treuchtlingen	25.80	6	50
Altmühl/Eichstätt	27.52	8	68
Schwarzach/Mettendorf	10.19	8	111
Donau/Ingolstadt	646.59	6	50
Donau/Kelheim	722.58	4	47
Paar/Manching (Bahnbr)	21.80	4	45
Naab/Unterköblitz	86.93	6	60
Naab/Heitzenhofen	173.44	6	59
Haidenaab/Wildenau	29.15	8	66
Regen/Chamerau	79.40	10	84
Regen/Regenstauf	96.29	8	105
Schwarzer Regen/Teisnach	43.44	8	76
Donau/Schwabelweis	818.26	8	87
Donau/Pfelling	840.73	8	78
Isar/Mittenwald	9.40	6	33
Loisach/GarmischudP	11.84	6	51
Amper/Fürstenfeldbruck	37.22	6	41
Amper/Inkofen	71.25	8	60
Ammer/Oberammergau	3.93	8	42
Ammer/Weilheim	28.58	8	65
Vils/Rottersdorf	21.66	8	70
Große Vils/Vilsbiburg	12.09	6	85
Rott/Birnbach	36.69	10	75
Alz/Altenmarkt	68.95	10	34
Tiroler Achen/Staudach	51.40	10	65
Traun/Stein	32.5	10	57
Rote Traun/Wernleiten	9.95	10	74
Ramsauer Ache/Ilsank	6.51	6	52
Saalach/Unterjettenberg	56.51	8	49
Inn/Oberaudorf	201.00	10	57
Inn/Eschelbach	324.49	10	71
Inn/Passau Ingling	899.63	10	53
Leitzach/Stauden	7.13	10	54
Glonn/Bad Aibling	9.18	10	69
Inn/Rosenheim	282.04	6	50
Donau/Achleiten	2450.00	10	38
Ilz/Kalteneck	60.97	10	66

Table A-4.1: Threshold and cluster size for the Danube River basin. For each gauge the threshold is chosen by evaluating the MRL plot and a point process fit over a range of thresholds. The cluster size is then selected such that for the next larger cluster size not significantly more values (i.e. more than 3 values) are eliminated.

A-5 SOFTWARE

Most of the statistical computing for this thesis has been done with the language ***R***. ***R*** is available as free software under the terms of the Free Software Foundation's GNU General Public License and can be obtained at *http://www.r-project.org/*.

For the wavelet analysis, mainly the *R*-packages ***waveslim*** from Brandon Whitcher, ***Rwave*** from Rene Carmona and Brandon Whitcher, and routines from Christopher Torrence and Gilbert P. Compo (originally written for Matlab, ported to ***R*** by myself) have been used (at *http://atoc.colorado.edu/research/wavelets/software.html* the source code can be obtained). The estimation of FARIMA model parameters and simulation with such models has been done utilising the packages ***farisma*** from Henning Rust (the software is available at *http://www.pik-potsdam.de/ hrust/tools.html*) and ***fracdiff*** from Chris Fraley and Fritz Leisch. For the extreme value analysis the packages ***ismev*** from Stuart Coles and Alec Stephenson, ***evd*** from Alec Stephenson and ***extRemes*** from Eric Gilleland, Rick Katz and Greg Young have been the basis.
All these packages can be obtained via *http://www.r-project.org/*.

Additional software produced by myself in ***C***- and ***R***-code is available via email: *Malaak.Kallache@pik-potsdam.de*.

Danksagung

Diese Arbeit wurde am Potsdamer Institut für Klimafolgenforschung in der Abteilung Integrierte Systemanalyse verfasst. Größtenteils wurde sie dankenswerterweise vom Bundesministerium für Bildung und Forschung gefördert (BMBF Projekt "Skalenanalyse" vom 01.03.2002-21.12.2005, Förderkennzeichen 0330271).

Bei meinem Betreuer Prof. Dr. Holger Lange möchte ich mich herzlich für seine intensive Betreuung, Motivation, Unterstützung und Begutachtung bedanken, die er mir trotz großer räumlicher Distanz zuteil werden ließ. Weiterhin danke ich Prof. Dr. H.J. Schellnhuber, welcher meine Arbeit betreut und begutachtet hat. Ganz besonderer Dank gebührt den Kollegen in meiner Arbeitsgruppe, Henning Rust und Jürgen Kropp. Durch die gute Zusammenarbeit und mannigfaltigen Diskussionen haben sich viele Erkenntnisse und Ideen ergeben. Die Mischung aus konzentrierter Arbeit und freundschaftlicher Atmosphäre trug sehr zum Gelingen meiner Dissertation bei, vielen Dank. Auch für den Austausch im Projekt, beispielsweise die DoktorandInnentreffen und insbesondere die gute Zusammenarbeit mit Birgit Thies und Jörg Neumann, bin ich dankbar. Dem im Projekt beteiligten Bayrischen Landesamt für Umwelt danke ich sehr für die Bereitstellung der Daten und die gute Kooperation. Jana Sillmann bin ich für ihre Anregungen im Bereich der Meteorologie dankbar.

Auch meinen Kollegen Brigitte Knopf, Oliver Walkenhorst, Matthias Lüdeke, Klaus Eisenack, Boris Orlowski, Hermann Held, Elmar Kriegler und Douglas Maraun danke ich herzlich für den guten fachlichen Austausch, besonders im Kontext der EURECA und StatCrew Aktivitäten. Ferner bin ich dankbar für die angenehmen Rahmenbedingungen im PIK im Allgemeinen und insbesondere für das Engagement Bärbel Uffrechts bei der Bücherbeschaffung. Beides hat das Arbeiten sehr erleichtert.

Kai Kopperschmidt, Andreas Buneß, Michael Reineke und Martin Güth möchte ich danken für viele fruchtbare Diskussionen und Tipps zu meiner Arbeit. Schließlich möchte ich ganz herzlich meinen Eltern danken. Außerdem danke ich Tanja Eichler für die Korrektur des Englischen und Felix Kupferschmidt für seinen Zuspruch und seine Begleitung.

Berlin, im April 2007
Malaak Kallache

Eidesstattliche Versicherung

Ich versichere an Eides statt, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt habe und alle benutzten Quellen und Hilfsmittel vollständig angegeben habe. Die Zusammenarbeit mit anderen Wissenschaftlern habe ich kenntlich gemacht. Diese Personen haben alle bereits ihr Promotionsverfahren abgeschlossen.

Eine Anmeldung der Promotionsabsicht habe ich an keiner anderen Fakultät oder Hochschule beantragt. Ich habe auch nicht diese oder eine gleichartige Doktorprüfung an einer anderen Hochschule endgültig nicht bestanden.

Berlin, den 07. April 2007